BANK FOR INTERNATIONAL SETTLEMENTS

BIS Papers
No 22

Investigating the relationship between the financial and real economy

Monetary and Economic Department

April 2005

Papers in this volume were prepared for a meeting of senior officials from central banks held at the Bank for International Settlements in October 2003. The views expressed are those of the authors and do not necessarily reflect the views of the BIS or the central banks represented at the meeting. Individual papers (or excerpts thereof) may be reproduced or translated with the authorisation of the authors concerned.

Requests for copies of publications, or for additions/changes to the mailing list, should be sent to:

Bank for International Settlements
Press & Communications
CH-4002 Basel, Switzerland

E-mail: publications@bis.org

Fax: +41 61 280 9100 and +41 61 280 8100

This publication is available on the BIS website (www.bis.org).

ISSN 1609-0381 (print)
ISBN 92-9131-677-6 (print)

ISSN 1682-7651 (online)
ISBN 92-9197-677-6 (online)

Foreword

The papers in this volume were presented and discussed at the Autumn Central Bank Economists' Meeting held at the BIS in Basel on 9-10 October 2003. The purpose of this meeting was to discuss challenges that central banks have faced in the context of monitoring the performance of the financial sector and the interaction between the health of financial institutions and macroeconomic stability. These challenges can be broadly grouped into three distinct but interrelated themes.

The first deals with the influence that financial conditions have on aggregate expenditure and overall economic developments. The second theme reverses the direction and looks at the impact of the macroeconomic environment on the financial health of different economic sectors. Finally, the third theme deals with the evolving nature of the measurement of financial risk both at the micro level of individual economic units and at the macro level of whole sectors or the overall economy.

List of meeting participants

Reserve Bank of Australia	Luci Ellis Economic Analysis Department
Austrian National Bank	Ernest Gnan Economic Analysis Division
	Doris Ritzberger-Gruenwald Foreign Research Division
National Bank of Belgium	Eric Jacobs Research Department
	Raf Wouters Research Department
Central Bank of Brazil	Eduardo Loyo Deputy Governor
Bank of Canada	Céline Gauthier Department of Monetary & Financial Analysis
	Jack Selody Department of Monetary & Financial Analysis
Central Bank of Chile	Alejandro F Jara Financial Analysis Unit
Czech National Bank	Narcisa Kadlcakova Monetary and Statistics Department
European Central Bank	John Fell Directorate Financial Stability and Supervision
	Philippe Moutot Directorate General Economics
	Thomas Wastermann Euro-area Macroeconomic Division, DG Economics
Bank of Finland	Lauri Kajanoja Economics Department
	Esa Jokivuolle Financial Markets Department
Bank of France	Sanvi Avouyi-Dovi Research Center
	Alain Duchateau Banking Supervision, General Directorate
Deutsche Bundesbank	Christian Upper Economics Department
Deutsche Bundesbank and Bank of England	Ulf von Kalckreuth Economic Research Centre

Hong Kong Monetary Authority	Stefan Gerlach Research Department
Bank of Italy	Sebastiano Laviola Vigilanza Creditizia e Finanziaria Aviram Levy Research Department
Bank of Japan	Nobuo Inaba Bank Examination and Surveillance Department Takashi Kozu Bank Examination and Surveillance Department Toshikata Sekine Research and Statistics Department
The Bank of Korea	Jong-Kun Lee Institute for Monetary & Economic Research Ju-Yeol Lee Research Department
Bank of Mexico	Javier Marquez Risk Analysis and Special Projects
Netherlands Bank	Jan Willem van den End Monetary and Economic Policy Department Maarten van Rooij Research Department
Central Bank of Norway	Arild Lund Securities Markets and International Finance Department Ingvild Svendsen Monetary Policy Department
Bank of Spain	Roberto Blanco Monetary and Financial Studies Department Ignacio Hernando Monetary and Financial Studies Department
Sveriges Riksbank	Tor Jacobson Research Department Jesper Lindé Research Department
Swiss National Bank	Caesar Lack Research Group Pierre Monnin Systemic Stability: Research and Policy
Bank of England	Glenn Hoggarth Financial Industry & Regulation Division Jan Vlieghe Monetary Analysis & Statistics Division

Federal Reserve Bank of New York	Arturo Estrella Research and Market Analysis
Board of Governors of the Federal Reserve System	David Reifschneider Division of Research and Statistics
	David Smith Division of International Finance
Bank for International Settlements	Claudio Borio Marco Sorge Nikola Tarashev Kostas Tsatsaronis William White Haibin Zhu

Contents

Investigating the relationship between the financial and real economy

Konstantinos Tsatsaronis

Central banks have always recognised the importance of financial stability for overall macroeconomic performance, but questions related to the health of the financial system have traditionally taken a back seat to those more directly linked to the process of inflation and growth. In recent years, however, financial stability has gained greater prominence on central bankers' agenda. Monitoring the performance of the financial sector and the interaction between the health of financial institutions and macroeconomic stability has increasingly preoccupied central bank economists and decision-makers.

The signs of intensified interest in financial stability are many. Central bank financial stability departments are explicitly mandated to monitor the performance of the financial sector and assess vulnerabilities. An increasing number of regular central bank publications devoted to communicating these assessments now feature prominently alongside other periodic publications more traditionally focused on macroeconomic developments. While these trends are especially pronounced among central banks that do not have direct supervisory responsibilities for financial institutions, they are certainly not confined to them.

The reasons behind this more intense focus on financial stability are linked to the factors that have increased the vulnerability of the macroeconomy to financial system stress. There are both structural and secular factors at work here.

On the structural side, deregulation has transformed the financial system, enabling financial firms to explore profitable opportunities more fully and to expand the scope of their activities. Intensified competition has promoted efficiency and encouraged innovation. As a result, the financial sector has grown rapidly both in size and in terms of its contribution to overall economic activity. At the same time, a deregulated environment is arguably also one more prone to volatility: failure is an integral part of the market adjustment mechanism in a competitive system and provides the natural check on participants' pursuit of profit.

On the secular side, the success of central banks in combating high inflation might also have influenced the nature of the interaction between the real and financial sectors of the economy. Reduced macroeconomic uncertainty has freed resources to transact in other sources of risk. At the same time, this success may also have had the unintended consequence of cultivating a sense of private sector complacency about the potential downside risks. Such an environment might arguably be more permissive of cumulative processes that gradually contribute to the build-up of financial imbalances, which in turn can be the source of macro instability when they unwind.

Beyond these factors, improved risk measurement "technology" has also played a supporting but key role. In particular, advances in the measurement and analysis of financial risk have contributed to a better understanding of the different dimensions of financial risk and vulnerabilities. Advances have been made in developing a greater overall conceptual framework and in more specific measurement tools. At the level of the individual enterprise, this has laid the basis of better risk management. At the macro level, it has spurred more structured and quantitative analysis, not least by improving the availability of information.

The papers in this volume deal with many such issues. They were presented at the annual Central Bank Economists' Meeting hosted by the BIS on 9-10 October 2003. The meeting was organised in three thematic units. The first deals with the influence that financial conditions have on aggregate expenditure and overall economic developments. The second theme reverses the direction and looks at the impact of the macroeconomic environment on the financial health of different economic sectors. Finally, the third theme deals with the evolving nature of the measurement of financial risk both at the micro level of individual economic units and at the macro level of whole sectors or the economy overall.

Impact of financial variables on the macroeconomy

One aspect of the interaction between the real and financial sectors is the influence of financial conditions of firms and households on consumption and investment. One can usefully distinguish the influence that operates through the demand for external funding, on the one hand, and that which operates through its supply, on the other.

On the demand side, production and consumption decisions are critically dependent on the underlying financial condition of economic agents. High levels of debt that are not supported by robust income flows can restrict the absorption capacity of the private sector and become a drag on economic expansion or even result in an economic slump. The risk is particularly acute in the later stages of a strong economic upswing, when the tendency to project the recent past into the future may feed overly optimistic expectations and encourage the build-up of financial imbalances, as balance sheets become overstretched. The vulnerability of these aggregate positions would undermine the validity of the individual expectations on which they are founded.

On the supply side, those same financial conditions are a key factor in determining the terms on which external funding is granted. Asymmetric information between suppliers and demanders of funds generally makes external funding more expensive and less accessible than its internal counterpart, such as retained earnings. It also makes it quite sensitive to the perceived and actual financial strength of economic agents. This is especially so for those that have less of a track record and less security to offer, such as smaller firms, which typically do not have access to capital markets. In addition, the financial condition of suppliers of funds themselves, especially financial intermediaries, can play an important role. A deterioration in their financial health can easily lead to retrenchment. Pressure on capital buffers can restrict the intermediation and risk-taking ability of financial firms, removing in turn a potential source of liquidity that could soften the constraints faced by the non-financial sector. And the fact that markets rely so much on banks for market-making and backstop liquidity services means that their functioning, too, can be impaired by a weakening in the financial vigour of institutions.

Asset prices play a key role in the process, on both the demand and supply sides. Private sector expectations are embedded into the prices of financial and real assets. As such, they reflect the extent of any excessive optimism or pessimism of market participants. In addition, they have a direct impact on the ability of the private sector to obtain financing, not least since a borrower's wealth is a common source of security for lenders. Asset price fluctuations, therefore, can have an important effect on determining macroeconomic outcomes through their impact on balance sheets. For much the same reasons, they can also contain useful leading information about macroeconomic developments.

Recent experience validates the importance of these mechanisms. For example, the euphoria that attracted ample capital into the technology and communications sector in the second half of the 1990s sowed the seeds of the recent slowdown, which was triggered primarily by a collapse in business investment expenditures. Similarly, the increased debt levels that households in a number of countries have recently incurred in order to participate in a soaring residential real estate market may weaken their ability to sustain the pace of consumption growth, so critical to support growth at the current juncture. Ironically, this might be particularly true if interest rates were to rise in view of a pickup in other sectors of the economy. At the same time, by comparison with experience in the early 1990s, the better health of financial intermediaries has helped to cushion the decline in economic activity following the initial slowdown, by limiting the tightening of the supply of external funding.

Impact of the real economy on financial strength of individual sectors

The interaction between prevailing financial conditions and real economic activity also runs in the opposite direction. The state of the business cycle has an important influence on incomes, profits and, by extension, the balance sheets and creditworthiness of various economic players. Understanding these links is no less important, especially if the objective is to gain insight into the feedback mechanisms that determine the overall impact of developments or policy actions on the state of the economy.

Financial conditions in the economy evolve largely in sync with the different phases of the business cycle, ie they are highly procyclical. Periods of expansion boost income and strengthen the balance sheets of households and firms. By contrast, the creditworthiness of borrowers deteriorates during periods of economic slowdown, which are typically associated with thinning income cushions and greater financial strains. In addition, the rise in default rates tends to spread those strains to the economic sectors that are net lenders of funds.

The profitability and balance sheet strength of financial intermediaries is closely linked to such developments. Fee and intermediation margin income has a very strong cyclical component. Similarly, as asset quality follows the business cycle, provisioning costs and outright losses tend to be higher in economic downturns. Moreover, prevailing accounting practices, which lead to a recognition of losses only once negative credit events are clearly identifiable, increase this synchronicity.

Developing a good understanding of the joint dynamics of these processes and their relationship to the business cycle is key to assessing vulnerabilities of financial conditions at any given economic juncture. The greater the common component in the dynamics of credit risk across different economic sectors, the more exposed the economy will be to shocks that can have widespread economic impact. Importantly, the more likely it is that this impact will have longer-lasting effects, owing to the mutually reinforcing interactions between the health of the financial and non-financial sectors.

Financial sector risk measurement in the small and in the large

Parallel to the increased policy interest in the interactions between the real and the financial sectors of the economy, risk measurement methodology has made major progress in recent years. This progress consists of more systematic approaches to data collection, the development of analytical frameworks as well as the modelling and empirical analysis of risk. Importantly, it also takes the form of efforts to embed these approaches into the daily business decisions of financial firms.

The development of a risk measurement and management framework has progressed sequentially across different types of risk and from the micro to the macro levels.

Advances have been most evident at the level of the risks faced by the individual firm. Here the framework for the measurement of market risk is the most advanced, followed by the modelling of credit risk; liquidity risks (market liquidity and funding liquidity risks) have also received considerable attention. Critically, not least in the wake of the autumn 1998 market turbulence, market participants have devoted much effort to understanding the mutually reinforcing interaction of these risks, at least with respect to episodes of market stress, far less so at business cycle frequencies. Typical tools include refinements to value-at-risk methodologies, the extension of similar concepts to the analysis of credit risk, as applied to both portfolios of traded securities and non-traded assets, and the development of stress testing. Focus on articulating a consistent framework for the understanding and measurement of operational risk is of more recent vintage, reflecting partly the absence of data.

Importantly for central banks, many of the basic tools and concepts can be and have been transposed from the micro to the macro level. In this case, the focus shifts from the analysis of the risks incurred by individual firms to those that are faced by the system, whether the “system” is defined in terms of broad sectors, such the banking sector, or the financial sector as a whole. The emphasis here is on the commonalities in risk exposures that signal a heightened probability of joint losses among financial institutions and on the mechanisms that can propagate strains across the financial system.

A key question that arises is how these measures of risk relate to general economic developments. More specifically, what are the lead-lag relationships between measured financial risk and economic activity? In other words, how much advance warning do the measures provide about the materialisation of risk?

The answer to this question largely determines the usefulness of the measures of risk. To the extent that the lead time is sufficiently long, using these indicators can provide useful advance information to both policymakers and market participants, allowing them to take remedial action. If, on the other hand, measures of financial risk tend to move coincidentally with the realisation of strains in firms, their primary function is more descriptive than predictive. In this case, they are less informative about current vulnerabilities as such. That is, they are more like a thermometer, providing an accurate measurement of current temperature, than a barometer, which by measuring current conditions that are imperceptible to our senses can offer insight into impending weather changes.

This is an important distinction. For, to the extent that risk measures are more descriptive than predictive, they can actually contribute to the amplification of business fluctuations. They can do so directly, by influencing funding and risk-taking decisions in a procyclical way. And they can do so indirectly, through the operation of the prudential constraints, as the framework moves away from a reliance on prescriptive rules and regulatory standards to become better aligned with the way financial firms measure, price and manage risks. Thus, during expansions, low levels of measured risk would encourage financial intermediaries to expand their activity, even as imbalances and the associated risk are actually building up. The opposite will be true during slowdowns, when increased levels of measured risk prompt retrenchment, potentially restricting the ability of the financial system to channel funds to their best use. This might seem a prudent course of action when viewed from the perspective of the individual institution in response to exogenous sources of risk. However, it is not necessarily the optimal response from a macro viewpoint, which is more sensitive to the mechanisms that can endogenously amplify the risk to the economy.

This has implications for prudential policy design. Arguably, a prudential policy framework should, to the extent possible, counterbalance the feedback mechanisms that tend to amplify the financial and business cycles. The optimal design of capital requirements, provisioning and reserving rules depends critically on the relative balance between idiosyncratic and systematic movements in the dynamics of asset quality, profitability and cost structures of the financial sector. Prudential norms that help reduce the importance of the systematic component of these movements should lead to more stable outcomes.

This macroprudential perspective is the one that is more naturally associated with central banks. The focus of analysis is on the interaction of different sectors of the economy and the ultimate objective is to ensure that policies are in place to foster macroeconomic stability. In other words, the object of study is financial vulnerabilities that can be the source of macroeconomic costs.

Financial stability research in central banks has sought to develop measures that quantify these vulnerabilities and can shed light on how they can be better understood and identified at an early stage. In this context, macro stress test exercises represent an important tool, as they can help to evaluate the vulnerability of the financial sector to large shocks and are particularly well suited to the assessment of systemic risk. The methodology readily lends itself to the study of the intensity of the mechanisms and interactions of individual responses that can amplify the overall impact of stress. It facilitates the study of the endogenous aspects of financial risk and in this sense adds value compared to the simple aggregation of analyses conducted at the micro level.

Importantly, the benefits of risk assessment exercises from a macroprudential perspective are enhanced when the analysis of financial risk is paired with that of relationships that have traditionally been at the centre of central bankers' attention, namely the links between monetary policy and the behaviour of different sectors of the economy. On the one hand, the reaction function of the monetary authorities is a key ingredient in macro stress tests. On the other hand, a greater understanding of the likely impact of monetary policy actions on financial conditions in the economy, and hence also on the supply of credit, can only lead to better policy decisions. In turn, charting these effects calls for a good understanding of prevailing attitudes towards risk-taking among economic agents, not least financial intermediaries, as critically conditioned by their financial soundness.

Viewed from this angle, financial stability analysis is an integral component of central banks' primary mission, viz the conduct of monetary policy aimed at providing a sound basis for macroeconomic stability and long-term growth. This is so regardless of whether the pursuit of the mission is seen as operating exclusively or just largely through price stability. In a world where the role of the financial sector has become more central in influencing these macroeconomic objectives, central bankers' more intense focus on financial stability is not only natural but also appropriate.

Disinflation and the dynamics of mortgage debt

Luci Ellis[1]
Reserve Bank of Australia

1. Introduction

A permanent reduction in average inflation should be expected to reduce nominal interest rates on deposits and loans by the same amount. Together with financial deregulation, which reduced interest margins and made housing finance more accessible, the reduction in nominal interest rates in Australia since the 1980s has eased the initial repayment burden of a given-sized debt. Households have therefore taken advantage of their increased capacity to borrow, resulting in rapid growth in household debt over the past decade or so, as shown in Graph 1. Consequently, the ratio of household debt to disposable income in Australia has increased from a level well below that in other developed countries, to something close to the upper end of the range of international experience.

In the process of transition to the new equilibrium, household credit should be expected to grow much more quickly than income. This has certainly been the situation in Australia in recent years. However, knowing that such a transition is in progress is not enough when trying to interpret correctly the current expansion in credit. It is also important to understand when the transition will end, and what the new equilibrium debt levels will be - or indeed, whether the process has gone too far and must partly reverse to reach its long-run sustainable path. This paper reports some analysis that tries to provide a sense of the likely magnitude of the change and its determinants, although it does not goes as far as predicting the timing of the end of the transition or the new equilibrium debt/income ratio.

After describing the workings of the key financial product of interest - the standard, variable rate mortgage loan - in the next section, in Section 3 we use a simple mechanical simulation to show the effects of a permanent reduction in nominal interest rates on indebtedness, given various assumptions about demographics and the growth and distribution of income. We then refine this simple framework in Section 4 to incorporate optimising behaviour by households in their choices about housing tenure and consumption of housing services, and the financing of those choices. We use this model to investigate the implications of a permanent disinflation for household sector indebtedness, housing prices and quality, as well as other characteristics of the housing market such as owner-occupation rates. As well as discussing the comparative statics of the long-run equilibria given different average inflation rates, transitions from the high-inflation to low-inflation equilibria receive particular attention. We summarise these conclusions and draw out some of the implications for the Australian economy in Section 5.

We must still take a number of things as given to make the analysis tractable. For example, we do not allow for the possibility that disinflation might induce lenders to alter their lending criteria. The results described here depend crucially on borrowing constraints and disappear if they are not present. The ceiling on the ratio of repayments to income, and in Section 4 the downpayment constraint, serve as the only constraints on intermediaries' willingness to lend. Our assumptions about both lending and borrowing behaviour also imply that the ratio of household debt to income will stabilise in the long run. That is, we exclude the possibility that both sides of the household sector balance sheet might deepen as living standards rise. We also focus on home mortgage debt and ignore consumer credit. We exclude a large number of factors that affect households' housing and financing decisions, including taxation and the possibility of government subsidies for particular kinds of housing arrangements. Finally, we ignore the effects of changing demographics on the debt/income ratio, other than allowing for steady population growth.

---

[1] The author thanks Andrew Stone for helpful comments about the formal model. Responsibility for any remaining errors rests with the author. The views expressed in this paper are those of the author and should not be attributed to the Reserve Bank.

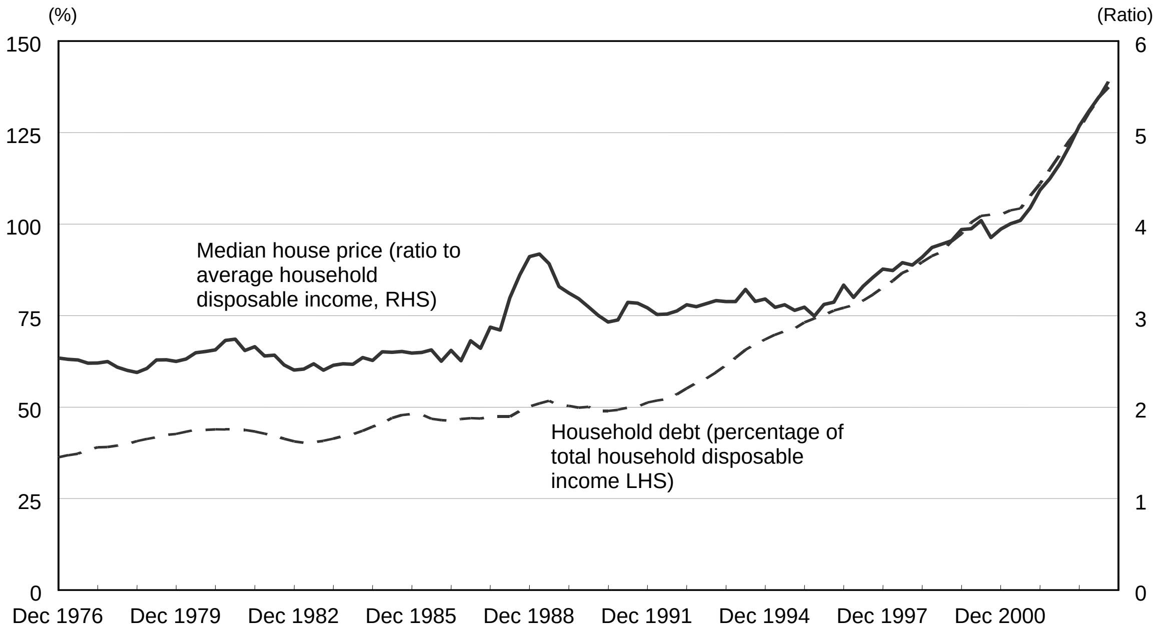
The analysis reported in this paper differs from, but is related to, the considerable literature on the effects of financial deregulation or liberalisation on household balance sheets (Throop (1986), for example). Among other things, this literature finds that borrowing constraints implied by market imperfections and regulation have tended to reduce owner-occupation rates (Zorn (1989), Duca and Rosenthal (1994)), especially for younger households (Haurin et al (1997), Ortalo-Magné and Rady (1999)), as well as constrain housing prices (Meen (1990)). Easing these constraints is therefore likely to increase housing demand and indebtedness, both because existing households can borrow more and because household formation rates rise (Börsch-Supan (1986)), although such mortgage qualification requirements are likely to have some effect even in a deregulated financial system (Linneman and Wachter (1989)). In addition, housing prices are likely to be more sensitive to interest rate shocks when financial sectors are liberalised than when they are regulated (Almeida (2000), Iacoviello and Minetti (2003)), which may be due to the greater responsiveness of asset prices to shocks when leverage is higher (Henley (1999), Lamont and Stein (1999)).[2]

The implications of disinflation and deregulation for household debt and housing are similar, with both events tending to enable greater debt accumulation by home-buying households. In recent decades, many developed countries including Australia have experienced both disinflation and deregulation, so that their long-run effects on debt/income ratios would have tended to compound each other. Both events seem to be necessary in order to generate the substantial deepenings in household balance sheets that have been observed (Ellis and Andrews (2001)). However, there are some subtle differences in outcomes from the two events. They therefore have different implications, particularly in terms of distributions of wealth and debt. This paper should therefore be viewed as complementary to existing literature on borrowing constraints and the effect of financial liberalisation, in effect disentangling the effects of disinflation in the increase in household indebtedness from those arising from financial deregulation.

Graph 1

Household debt and house prices

Relative to household disposable income



Sources: Reserve Bank of Australia; Real Estate Institute of Austrialia; Australian Bureau of Statistics.

---

[2] This is true on the condition that mortgage finance is primarily provided at variable interest rates. In economies such as the United States where long-term fixed interest rate mortgages are the norm, demand and construction activity appear to have become less sensitive to interest rates with deregulation, because the supply of mortgage finance is now less sensitive to variable interest rates (McCarthy and Peach (2002)).

2. The fundamental object: credit-foncier home mortgages

The basic object of analysis in this paper is the standard, variable rate loan of the credit-foncier type. Under this type of loan contract, the borrower must repay the original principal over an agreed maximum term, by making regular repayments. The repayments are a constant nominal amount if interest rates do not change; and if interest rates do change, the repayment is recalculated to ensure the debt is still fully repaid over the original term. (With debt contracts of this type, the borrower may also be permitted to make early repayments without penalty.) The required repayment *rp* per period is a function of the initial amount borrowed (P), the per-period interest rate (i) and the number of repayments (T) to be made over the life of a loan, as shown in equation (1).[3] This is a standard calculation available in spreadsheet packages and handheld calculators. Given this repayment, the remaining principal outstanding falls slowly at first, and then more quickly later in the life of the loan, as shown in the top left-hand panel of Graph 2. Credit-foncier loan contracts have the property that the remaining debt outstanding in any period k part-way into the life of the loan is equal to the loan size that would generate the same per-period repayment *rp* over the shorter loan life $T - k$. Therefore there is an analytical expression for the remaining outstanding debt P_k in period $k \geq 1$, as shown in equation (2). The fixed total repayment therefore changes in composition through time, with a declining fraction being interest and a greater fraction being repayments of principal, as shown in the top right-hand panel of Graph 2. The real burden of this fixed nominal repayment naturally declines at a rate determined by the rate of growth in nominal incomes.

$$rp = P\,[i\,(1+i)^T] \,/\, [(1+i)^T - 1] \tag{1}$$

$$P_k = P\,[(1+i)^T - (1+i)^k] \,/\, [(1+i)^T - 1] \tag{2}$$

A permanent disinflation can result in an increase in the ratio of household debt to disposable income because lower nominal rates allow borrowers to service larger debts with the same repayment. This is particularly true in a country like Australia where most home mortgage finance is provided at variable rates.[4] The bottom left-hand panel of Graph 2 shows how large the initial loan size can be at different interest rates, while maintaining the same nominal total repayment as on a AUD 100,000 loan at 10% interest with a 20-year term repaid monthly. If it is the initial burden that is the binding constraint on households' ability to borrow, a disinflation will clearly allow households to borrow more. This is known as the repayment tilt effect (Howitt (1990)). Although the dollar value of the repayment is unchanged over these different combinations, the implications for debt and repayment ratios to income are very different. The burden of the fixed nominal repayment declines more slowly when nominal income growth is lower, as shown in the bottom right-hand panel of Graph 2 (Stevens (1997)). In addition to the effect on repayment burdens, slower growth in income - taken in Graph 2 to vary by the same amount as nominal interest rates - compounds the effect of the higher initial value of the maximum allowable debt on aggregate debt/income ratios. This occurs because an individual borrower's ratio declines less quickly when incomes grow at a slower rate.

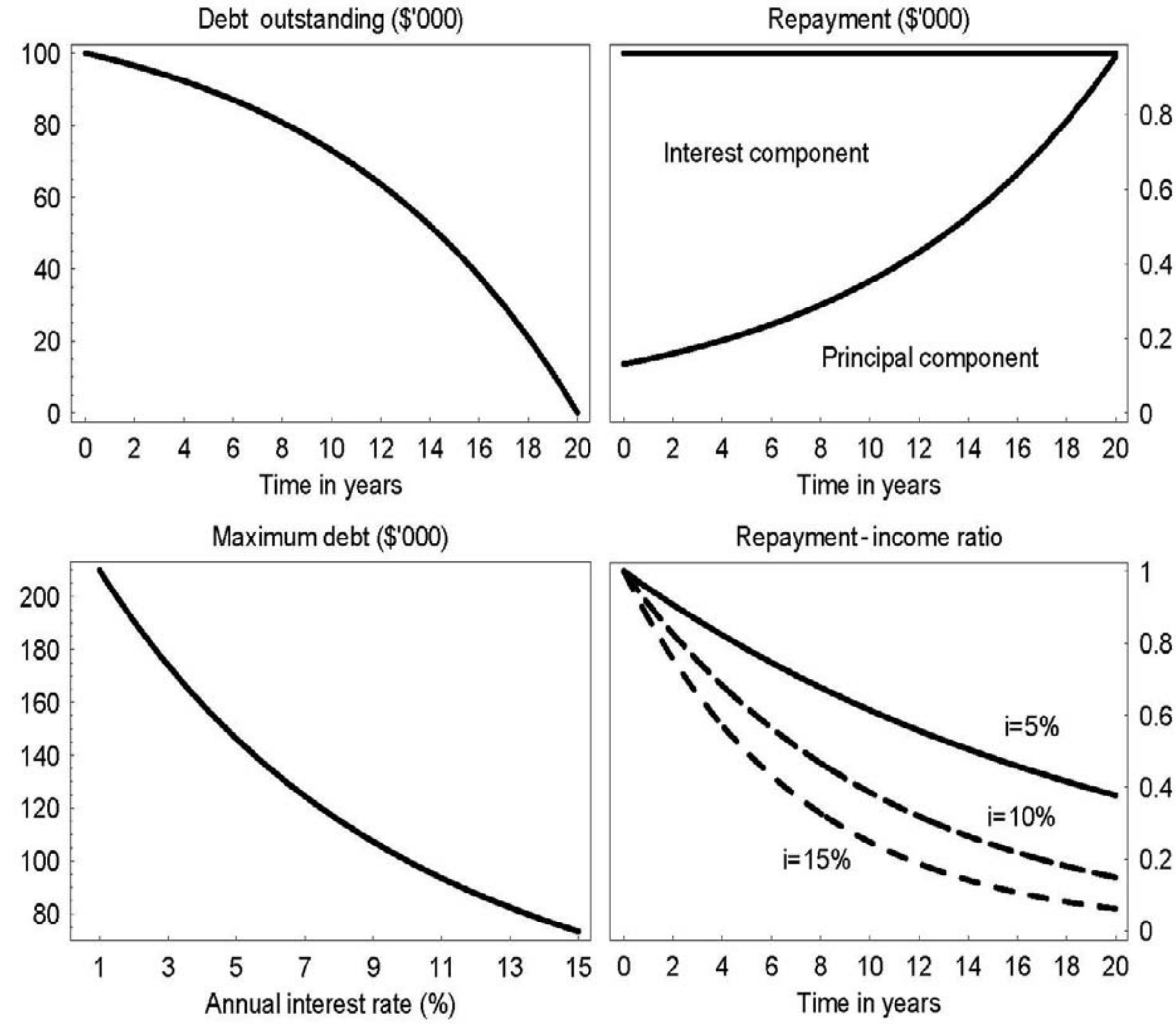
---

3 This formula assumes that repayments are made in arrears, that is, at the end of the period, and that the loan is fully paid off at the end of the term.

4 Although a minority of new borrowers (usually between 10 and 20%) do fix the rate on some or all of their mortgage loan, lenders generally only offer products with a fixed interest rate of one to three years, with five- and 10-year fixed rates being very much the exception. After the fixed period has expired, the loan reverts to the interest rate applying to a standard variable home loan.

Graph 2

Debt and repayment profiles for credit-foncier home mortgages



Source: Reserve Bank of Australia.

3. Disinflation and aggregate mortgage debt

From the preceding discussion, we can see that a permanent, recognised reduction in inflation and nominal interest rates can increase debt/income ratios. This occurs both because initial loan sizes can rise in absolute terms, and because the ratio of debt to income diminishes more slowly through the life of the loan when nominal income growth is slower. In this section, we develop a first pass at quantifying these effects on aggregate household debt and repayment burdens, using a highly stylised model of households incurring and then paying off home mortgage debt. We assume that lenders impose a *repayment ratio test*, lending to the individual household only up to the amount that would generate a prespecified ratio of the total repayment to current nominal income. We assume that households are always willing and able to borrow this amount; effectively, the repayment ratio test is the only constraint on households' decisions to borrow. Initially, we will ignore downpayment requirements - effectively assuming that homebuyers can borrow 100% of valuation - and instead defer this consideration to Section 4. We then show how the mechanics of the credit-foncier loan contract imply that the ratio of aggregate household debt to aggregate income converges on a long-run equilibrium level that depends on the nominal interest rate, the rate of nominal income growth and the distribution of income by age.

The model is extremely simple and mechanical. Households are formed at age 25, and purchase a home using 100% debt funding. To do so, they borrow the maximum amount lenders will extend to them, given their income. This maximum is determined by a repayment ratio test imposing a maximum ratio of repayments to gross income of 30%. We choose this figure because it approximates the kinds of lending conditions actually imposed by Australian banks. The loan repayments are calculated on the basis of monthly repayments for a 25-year term (300 payments) at the prevailing interest rate. Households pay down their mortgage according to the required schedule, and then spend the remainder of their life until age 75 as outright owners of their home. Given the implied path for debts of households of different ages, and an assumed distribution for household income by age, the debt/income ratio for the whole household sector can be calculated by simply aggregating debts and income across cohorts. For a given rate of inflation, nominal interest and nominal income growth, as well as the age/income distribution and ratio imposed by the repayment ratio test, there is a steady state debt/income ratio for the whole household sector.

The top left-hand panel of Graph 3 shows how this aggregate debt/income ratio varies with inflation, given the repayment ratio test of 30% mentioned above and real income growth of 2%, for a range of levels of real interest rates. The income distribution by age used is that implied by the 2001 Household, Income and Labour Dynamics in Australia (HILDA) Survey, smoothed using a non-parametric lowess regression, as shown in the bottom right-hand panel of the graph.[5] The property of the income distribution that matters most for the long-run debt/income ratio is the ratio between (average) household income at the age the loan is borrowed, and the average income of the whole household sector. This is because the repayment ratio test is only applied when the loan is first borrowed. As discussed earlier, for a given level of real interest rates, lower inflation increases the aggregate debt/income ratio in two ways. First, the resulting lower nominal interest rates allow young households to take out larger loans and still meet the repayment ratio test. Therefore every cohort of households have higher nominal debt relative to their income when nominal rates are lower. Second, the implied lower rate of nominal income growth implies a slower decay in the ratio of debt to income. Higher growth in real income naturally results in faster nominal income growth for a given rate of inflation. Therefore higher real income growth results in a lower long-run debt/income ratio, given the rate of inflation, as shown in the centre left-hand panel of Graph 3. Similarly, the longer the loan term, the higher the long-run debt/income ratio, as shown in the bottom left-hand panel of Graph 3. This occurs because a larger proportion of all age cohorts still have debt if the term is longer, and because the path at which the debt is paid down is more gradual.

The implications of a permanent, credible disinflation for the debt/income ratio are therefore unambiguous. Suppose that, at some point $t = 0$, inflation falls credibly and permanently, with nominal interest rates and income growth falling in tandem. The households that had originally borrowed when inflation was higher could then potentially lower their repayments, while newer cohorts could borrow greater amounts as implied by the larger maximum permitted under the repayment ratio test. Once all the borrowers who borrowed when inflation was high have paid off their loans - which by definition occurs within 25 years, the assumed loan term - the system reaches a new steady state. The comparative statics of this change can be read off the schedules shown in the top left-hand panel of Graph 3.[6]

Assuming the older borrowers, who now face unexpectedly lower interest rates, lower their repayments (and presumably consume the difference), the transition path is smooth and concave. The debt/income ratio approaches its new steady state level at a diminishing rate; examples of these transition paths are shown in the top right-hand panel of Graph 3. This follows from the concavity of the path for nominal debt as shown in the top left-hand panel of Graph 2 above. The steady state debt/income ratios shown in the left-hand panels of Graph 3 are based on the presumption that all age

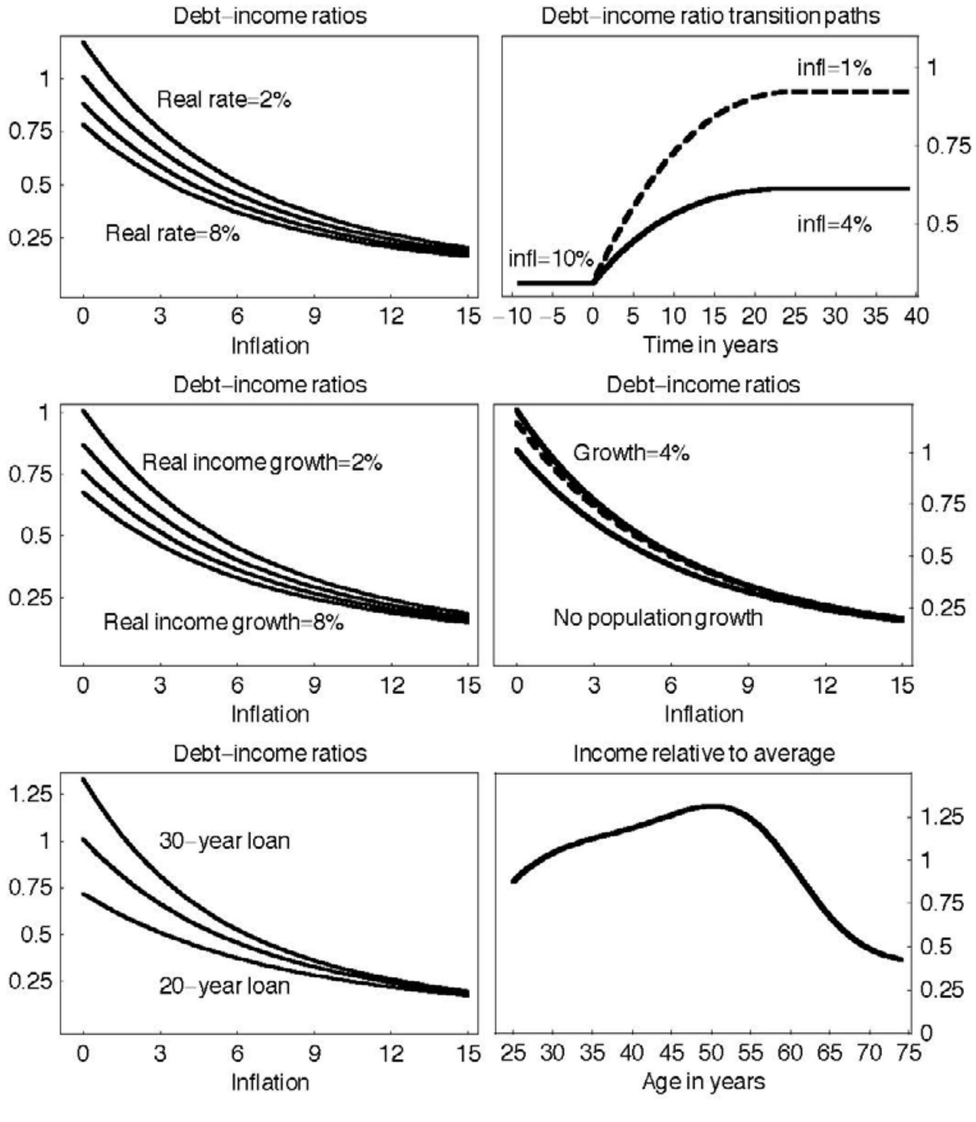
5 For compatibility with the assumptions of the simulations presented here, these income relativities are based on an unweighted average household income rather than one that takes the actual age distribution of Australia's population into account.

6 Similar results were shown in tabular form in Reserve Bank of Australia (2003), without the added complication of population growth or the distribution of income by age.

cohorts are of equal size. That is, there is no population growth and lifespans are identical. It is straightforward to see that population growth will increase the steady state aggregate debt/income ratio. This occurs because, when the population is growing, a greater proportion of households are therefore in their high-debt years. The centre right-hand panel of Graph 3 shows the effects of various constant rates of natural population growth on the steady state debt/income ratio for a range of inflation rates (π), assuming real rates are constant at $r = 4\%$ and real income growth is constant at 2%. Constant population growth affects the assumed long-run level of the debt/income ratio, but it does not alter the factor of proportionality between the ratios implied by different rates of inflation.

Graph 3

Aggregate debt/income rate



These steady state debt/income ratios and transition paths can readily be calculated by simulating the debt profiles of the required number of age cohorts, N. An analytical expression for the ratio can also be obtained, although it is rather cumbersome. Using equations (1) and (2), we define the maximum repayment ratio as being some fraction ψ of the income of the youngest cohort $y_0 Y_0(1 + w)^t(1 + \pi)^t$, where y_0 is the ratio between the youngest cohort's income and average income, $Y_0(1 + w)^t(1 + \pi)^t$. Y_0 is nominal average household income in the initial period 0, π is inflation and w is real income growth. We can therefore write the steady state debt/income ratio $D(\cdot)$ as equation (3).

$$D(y_0, \psi, N, T, i, \theta, \pi, w) = \frac{\psi y_0}{i} \frac{\sum_{j=1}^{N} \left\{ (1+\theta)^{N-j} \left[1-(1+i)^{\min(j-T,0)} \right] (1+\pi)^{-j} (1+w)^{-j} \right\}}{\sum_{j=1}^{N} (1+\theta)^{N-j} y_j} \tag{3}$$

The other parameters in equation (3) are the loan term T, number of annual cohorts N, nominal interest rate i, population growth rate θ, and, as mentioned before, inflation π and real income growth w. The ratio between the income of each cohort j and average income is denoted as y_j. This expression assumes that there is only one repayment per period, that is, that households that take out a loan within a single year can only be treated as a single cohort if they make one repayment per year on their loan. The case of multiple repayments per period can be accounted for using a version of equation (3) where the interest rate and rates of inflation and income growth are suitably redefined. The effect of multiple repayments per cohort, given the same annual interest rate, is to allow a higher initial loan size for young households and therefore a slightly higher long-run debt/income ratio.

Putting all these influences together, it would seem that the disinflation and reduction in margins on home loan interest rates seen in Australia since the late 1980s would be broadly consistent with an approximate doubling of the aggregate household debt/income ratio. Since the ratio has in fact more than doubled, from around 50% in 1991 to more than 125% in 2003, it seems likely that this transition has completed, as well as possibly being reinforced by a relaxation of other lending conditions, resulting from financial deregulation. Moreover, any further increase in this ratio is presumably attributable to other factors, such as the easing of other kinds of borrowing constraints, or an increase in the prevalence of refinancing with equity withdrawal.

3.1 Early repayment

Because interest payments on home mortgage debt are not tax-deductible in Australia, households are effectively repaying their mortgages out of post-tax income. Since interest and other investment earnings are taxed, this creates a strong incentive for homebuyers to repay their existing debt ahead of schedule if they can, rather than invest in some other asset where the return is taxed (Zorn and Lea (1989)). This is one of the features of the market that encourages the prevalence of variable rate debt; lenders then have no maturity mismatch, and thus have no incentive to impose a prepayment penalty.

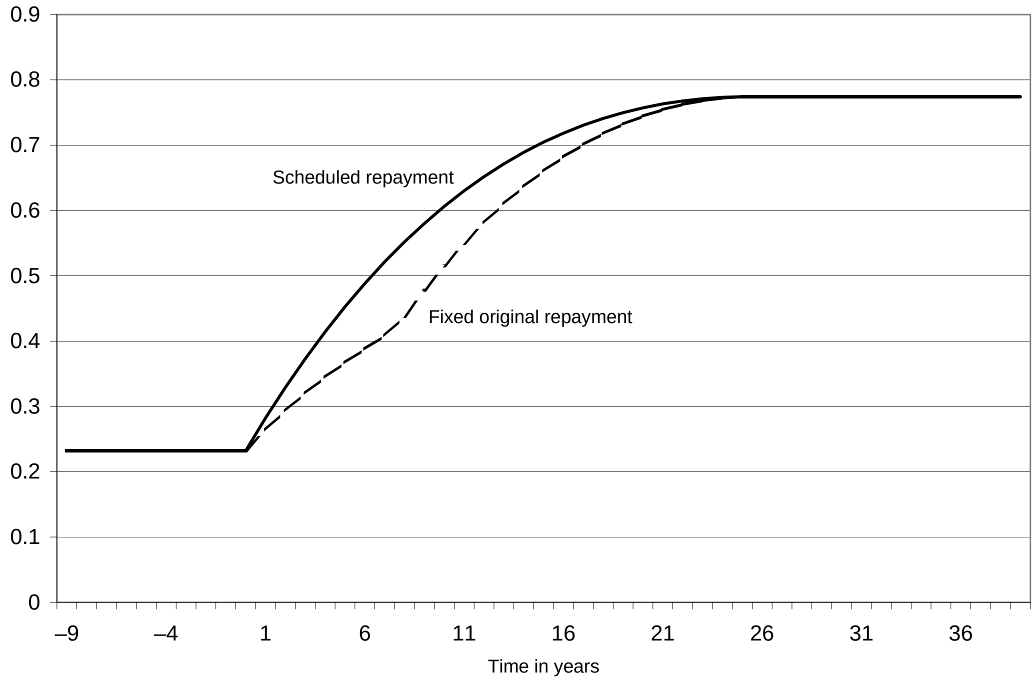
The scope for early repayment is potentially very important for the transition period from a high-inflation state to one with a permanently lower inflation rate. Households that initially borrowed when inflation was high will find that their repayments have fallen below the level they originally expected, although the burden of these lower repayments will also diminish more slowly because income growth is slower in the low-inflation state. These households may choose to maintain their repayments - at least in nominal terms - at or close to the level that would have been implied by the higher nominal interest rates in the high-inflation state. Since rates are actually lower than they had been when the loan was first taken out, this means that the debt is paid down much more quickly than implied by a normal credit-foncier loan contract.

The implications of this response by older cohorts serve to make the transition to the new steady state debt/income ratio more drawn out, although the steady state ratio itself is unaffected. The extent of this effect depends entirely on the difference between the original and the new nominal interest rate. Graph 4 shows the effect of a permanent disinflation from 12% per annum to 2%, reducing nominal rates from 16% to 6%, when existing borrowers reduce their repayments to maintain the original term (scheduled repayments), and when they maintain their original repayments.[7] In the latter case, the increase in the debt-income ratio is substantially slower than if the earlier borrowers reduce their repayments, but begins to catch up again after a decade or so. Although it is difficult to be certain, it is possible that this effect served to make the current transition of Australia's household debt/income ratio to be more drawn out than it otherwise would have been, even though the rapid increase in the ratio did begin almost immediately when inflation and nominal rates came down.

[7] For simplicity, this figure shows the case where households make only one, in-arrears payment on their mortgage per year.

Graph 4

Debt/income ratio transition paths



Source: Reserve Bank of Australia.

4. Adding in the downpayment constraint

In the previous section, we assumed that the lenders' repayment ratio test is the only constraint on households' willingness or ability to borrow. A reduction in nominal interest rates would therefore always induce further borrowing to restore households' repayments to the maximum proportion of income allowed. Even though lower nominal income growth increases the burden of a given mortgage repayment in the later part of the life of the loan, it was assumed that this did not affect borrowers' decisions about the initial loan size they would take on. Neither did we account for the *downpayment constraint*: the fact that, even if the household can service a debt of given size, it still must have accumulated enough savings to fund the downpayment, before it can purchase a home. Moreover, we previously assumed that, in the transition period, households that had originally borrowed at the higher interest rate did not respond to the change in circumstances brought about by the disinflation. Previously they just consumed the unexpected reduction in their mortgage repayments (or, in Section 3.1, maintained a constant repayment), because they neither extracted their windfall equity gains by borrowing more, nor did they respond to the resultant increase in the relative price of housing by downgrading to a smaller home.

In this section, we relax these strong assumptions to get a better sense of the consequences of a permanent disinflation, in particular its implications for the level of household debt. Downpayment requirements have previously been recognised as important constraints on access to home ownership (Stein (1995), Haurin et al (1997)) and are frequently considered to be a summary measure of the extent of financial repression or constraint in home mortgage finance (Iacoviello and Minetti (2003)). The analysis presented here assumes no taxes or subsidies, and a fixed minimum downpayment constraint. However some lenders do not impose a downpayment constraint and instead lend 100% of valuation. This eliminates the effects of the downpayment constraint entirely. Government upfront subsidies to first-home buyers would also serve to ameliorate this constraint.

4.1 A model of tenure and housing choice

As in Section 3, we use an overlapping generations model to capture the life-cycle aspects of home ownership and mortgage finance, with 50 cohorts so that a reasonably realistic annual frequency of decision-making is possible. Previous literature used overlapping generations with fewer cohorts, but this requires some heterogeneity within cohorts (Ortalo-Magné and Rady (1999)). The added complexity of a model with many cohorts permits us to assume that households are identical within a cohort, and still permits reasonably smooth responses to small parameter changes. Young households initially rent a home, and save to accumulate a downpayment. Once they have accumulated a downpayment and can meet the repayment ratio test for a home that satisfies their demand for housing services, they will take out a 25-year mortgage and become owner-occupiers. We assume that externalities in the landlord-tenant relationship result in rents exceeding the housing services provided by rental properties, following Henderson and Ioannides (1983). Therefore households will always prefer owning to renting, all else equal. The households will then pay down the debt according to the required schedule, and own their homes outright for the remainder of their lives.

When the household dies, after 50 years of adult life, the home is sold to the marginal young household that is ready to buy its own home. The proceeds of the sale are distributed as a lump sum transfer equally to all households. This implies that when housing prices and the population size are constant, young households could simply passively receive these inheritances and accumulate the required downpayment over a few years - for example, five years if the downpayment requirement is 10%. To ensure an interior solution where young households save from their own labour income, we would need to calibrate the model so that the externality involved in renting is large enough that young households would rather save more and buy sooner than pay another year of rent. Allowing for population growth will also tend to result in young households actively saving in this model, since the number of new young households will therefore exceed that of old households at the end of their lives. Even for relatively low rates of population growth, a sizeable fraction of these new households will have to purchase newly built homes, rather than the deceased estates homes left by the oldest cohort. Although in principle the young households could simply wait longer to accumulate enough of an inheritance, this would result in an ever increasing age at first-home purchase rather than a steady state equilibrium.

We assume that households choose real consumption of a composite consumption good (c), housing services (h) and leisure (l) to maximise their expected utility over their N-period finite lives (4). Utility is assumed to be additively separable through time, and across goods. The price of purchasing housing, relative to the consumption good, is denoted as p. The consumption good's actual price rises at a constant rate π, so at any period t the (normalised) price level is $P \equiv (1+\pi)^t$ and the price of housing is $p_t(1+\pi)^t$.

$$\max U = E\left\{\sum_{j=1}^{N} \delta^{j-1} u(c_j, h_j, l_j)\right\} \tag{4}$$

In each period, households must also choose their housing tenure (χ), where $\chi = 1$ if the household is an owner-occupier and zero if it rents. Because households always prefer to own rather than rent their home, if they can, they will rent while young and own when older, and never choose to revert to renting once they have bought their first home. Therefore the sequence of tenure states χ_j will be comprised of a sequence of zeros followed by a sequence of ones, with combined length N. We denote the age of first-home purchase (first one in the sequence of χ_js) as z. As well as borrowing mortgage debt d with an initial term T to purchase a quantity of owner-occupied housing h, households can use the same debt to invest in rental properties a, from which they receive a rental return R^*.[8] Even if households borrow additional amounts later in life, it is assumed for simplicity that they must still pay off their entire debt by the end of the original term, so any later borrowings must be paid off over a shorter term. The loan contracts are of the credit-foncier form described in Section 2, so that (1) and (2) hold. The j subscript on R and p refers to time periods experienced by a given cohort at each age j, not age-specific prices and rents.

[8] Although it has not always been the case in Australia that households could borrow for investment property on the same terms as for owner-occupied property, we assume that it is possible here.

Households can also hold a risk-free financial asset b, which is assumed to return a nominal interest rate of $i - m$, where i is the nominal interest rate paid on mortgage debt; households would therefore never borrow to buy the financial asset. Therefore households may receive labour $P\ W_j\,(1 - l_j)$, interest and rental income, as well as receive inheritances from the oldest households.[9] The wage rate may differ across age cohorts j in a given time period. They pay out this income for consumption and rent or mortgage repayments as appropriate, with the remainder going into asset purchases, whether of financial assets, or owner-occupied or rental housing. Putting these different sources and uses of income together, it turns out that the household's problem is to maximise U subject to the asset accumulation condition (5). Because home-owning households can adjust their consumption of housing services through time, this budget constraint allows for endogenous decisions about renovation or upgrading to a better home.

$$\begin{aligned} b_j - d_j &= W_j P(1-l_j) + (1+i-m)b_{j-1} + R_j^* P a_{j-1} + P p_j h_N \left[\sum (1+\theta)^k\right]^{-1} - Pc_j - (1-\chi_{j-1})R_j P h_j \\ &\quad - \chi_{j-1} d_{j-1}\left(1 + i - \frac{i(1+i)^{T+z-j}}{(1+i)^{T+z-j}-1}\right) - (\chi_j - \chi_{j-1})P p_j h_j - P p_j (a_j - a_{j-1}) - \chi_j \chi_{j-1} P p_j (h_j - h_{j-1}) \end{aligned} \quad (5)$$

The repayment/income ratio test and downpayment test on households' mortgage debt are captured as further constraints on their maximisation problem, as shown in (6) and (7). The maximum ratio of repayment to income is denoted as ψ as in Section 3, while the maximum loan/valuation ratio is denoted as $\overline{\omega}$. We assume that lenders treat owner-occupied and investment properties together when calculating the loan/valuation ratio. Households that own both owner-occupied and investment properties are treated as though they have the same gearing on both types of property, although in a model with taxes they may prefer a different arrangement depending on how the two types of housing are taxed.

$$d_j\left(\frac{i(i+1)^{T+z-j}}{(i+1)^{T+z-j}-1}\right) \le \psi W_j P(1-l_j) \qquad \text{repayment constraint} \quad (6)$$

$$d_j \le \overline{\omega} P p_j (h_j \chi_j + a_j) \quad \text{downpayment constraint} \quad (7)$$

The constraints (6) and (7) only apply if $z \le j \ge T + z$, where again z denotes the age at which the household first becomes an owner-occupier. Beyond age $T + z$, it is assumed that debt outstanding $d_j = 0$. In order that all mortgage debt is repaid before the household dies at age N, lenders require the following condition to hold (8). In general, z will be a product of the equilibrium solution, but for some combinations of parameter values, the term of the original mortgage T might also need to be adjusted to satisfy this condition.

$$T + z < N \quad (8)$$

As noted earlier, landlords' required rental returns compensate for an externality in the provision of rental property, so that it costs more to consume a given amount of housing services h as a renter than as a homeowner, and landlords effectively receive less in rental income (R^*) than their tenants actually pay (R) (Henderson and Ioannides (1983)). The difference can be assumed to be lost in maintenance or monitoring costs; we assume that this is a constant wedge ϕ between R and R^*. This implies the following relationship between rents, mortgage interest and the rate of return of financial assets (9).

$$R > i > \{R^*, i - m\} \quad \text{where } R - R^* = \phi \quad (9)$$

The relationship between the rental return and the return on financial assets depends on households' expectations of future capital gains. Arbitrage implies that the (risk-adjusted) total return on rental housing, including expected capital gain, equates to the return on alternative assets. This is captured in a standard relation used throughout the literature (Meen (1990), Bourassa (1995), Meen (2000), for example), although in this case there are no taxes, so the relationship is as shown in (10). In this arbitrage condition, p_a denotes the relative housing price that investors are willing to pay (which may be different from the relative price of housing actually transacted in the period p_t), and ω_a is the

9 These inheritances are assumed here to be evenly distributed across surviving households, although there is some empirical evidence that older households direct their gift-giving towards particular types of households in key home-buying age groups (Mayer and Engelhardt (1996)).

loan/valuation ratio on the rental properties they own. We abstract from the repayment constraint in the case of investing households, since the interest component of the repayment is already included in (10), and the (initial) principal component is small. The downpayment constraint is enforced by requiring that $\omega_a \leq \overline{\omega}$; in this case, ω_a is an overall leverage ratio, including borrowing for both investor and owner-occupied residential property.

$$(R \div p_a) - i\omega_a + \left[(1+\dot{p}^e)(1+\pi)\right] = (i-m)(1-\omega_a) \tag{10}$$

Finally, we have a condition that all the rental properties have to be owned by someone (11). This differs from previous literature, where rental properties were generally assumed to be owned by a separate landlord sector (Ortalo-Magné and Rady (1999), Iacoviello and Minetti (2003)). It is, however, more in keeping with the structure of the rental housing market in Australia to assume that rental households are owned by other households (Yates (1996)).

$$\sum_{j=1}^{N}(1+\theta)^{N-j} a_j = \sum_{j=1}^{N}(1+\theta)^{N-j}(1-\chi_j)h_j \tag{11}$$

There are no taxes in this model, although they could be added as extensions in further work. Although tax policy is widely recognised as a key driver of outcomes for housing tenure (Hendershott and White (2000), Hendershott et al (2002), Yates (2003)), prices (Capozza et al (1996)) and quality (Gobillon and le Blanc (2002)), we ignore the possible implications of differences in tax treatment of different housing tenures in order to focus on those arising from disinflation.

4.1.1 Equilibrium

Equilibrium in this model requires maximising expected utility (4) by choosing sequences of consumption, housing service consumption, leisure, housing tenure, debt and ownership of financial assets and rental housing for each life stage j, $\{c_j, h_j, l_j, \chi_j, d_j, b_j, a_j\}$ ($j = 1\ldots N$), subject to the $j+1$ equality constraints represented by (5) and (11) (Lagrange multipliers λ_1 and λ_2), the $2 \times j$ inequality constraints (6) and (7) (Lagrange multipliers λ_3 and λ_4), and $3 \times j$ non-negativity constraints affecting a_j, b_j and d_j (Lagrange multipliers λ_5 to λ_7). Conditional on the sequence of housing tenure outcomes $\{\chi_j\}$, this can be depicted as a standard optimisation problem with inequality constraints, using Kuhn-Tucker-style first-order conditions.

$$\begin{aligned}
&V_c = u_c - \lambda_1 P = 0\\
&V_h = u_h \lambda_1 P\left[(1-\chi_{j-1})R_j - (\chi_j - \chi_{j-1} + \chi_j\chi_{j-1})p_j + \delta\chi_{j+1}\chi_j(1+\pi)p_{j+1}\right] - \lambda_2(1+\theta)^j(1-\chi_j) + \lambda_4\overline{\omega}Pp_j\chi_j = 0\\
&V_l = u_l - \lambda_1 W_j P - \lambda_2 \psi W_j P = 0\\
&V_a = \lambda_1\delta R^* P(1+\pi) - \lambda_1 P p_j + \lambda_1\delta P(1+\pi)p_{j+1} - \lambda_2(1+\theta)^j + \lambda_4\overline{\omega}Pp_j + \lambda_5 \leq 0 \quad (\text{if } <, a_j = 0)\\
&V_b = -\lambda_1 + \lambda_1\delta(1+i-m) + \lambda_6 \leq 0 \quad (\text{if } <, b_j = 0)\\
&V_d = \lambda_1 - \lambda_1\delta\chi_j\left(1+i-\frac{i(1+i)^{T+z-j}}{(1+i)^{T+z-j}-1}\right) - \lambda_3\left(\frac{i(1+i)^{T+z-j}}{(1+i)^{T+z-j}-1}\right) - \lambda_4 + \lambda_7 \leq 0 \quad (\text{if } <, d_j = 0)\\
&V_{\lambda_3} = \psi W_j P(1-l_j) - d_j\left(\frac{i(1+i)^{T+z-j}}{(1+i)^{T+z-j}-1}\right) \geq 0 \quad (\text{if } >, \lambda_3 = 0)\\
&V_{\lambda_4} = \overline{\omega}Pp_j(h_j\chi_j + a_j - d_j) \geq 0 \quad (\text{if } >, \lambda_4 = 0)
\end{aligned} \tag{12}$$

As noted previously, because of the externality creating a wedge between the cost of renting and the cost of owner-occupation, households would not choose to revert to rental housing after owning a home, unless they received a sufficiently large negative income shock that meant they no longer fulfilled the repayment constraint (6). Therefore, we only consider as candidate equilibria outcomes where the sequence of housing tenure outcomes $\{\chi_j\}$ consists of a sequence of zeros (renting) of length $z-1$ followed by a sequence of ones (owning) of length $T-z+1$. Given this restriction on the possible sequences of housing tenure outcomes, the solution V^* (12) to the households' problem can be solved in two stages: first, solve the problem conditional on some value of z; then, find the value of z which gives the maximum utility of these conditional solutions.

$$V^* = \sup_z V(c, h, l, a, b, d; z) \tag{13}$$

In (13), $V(\cdot; z)$ is the maximised value of utility obtained by choosing $\{c_j, h_j, l_j, d_j, b_j, a_j\}$ ($j = 1\ldots N$), conditional on z. This involves solving the first-order conditions (12), where subscripts of u denote partial derivatives of the utility function with respect to the relevant choice variable, with the *j* subscript dropped for notational simplicity, as well as the first-order conditions with respect to λ_1 and λ_2, (5) and (11). The parameters of the model are ϕ, i, m, θ, ψ, $\overline{\omega}$ and the sequence of wage rates applying through time and (potentially) across age groups $\{W_{t,j}\}$ ($t = 1\ldots\infty$, $j = 1\ldots N$). If inflation remains at a constant rate π and real income growth is also a constant rate of w, then we can simplify the set of wage rates to $\{W_j(1+w)^t(1+\pi)^t\}$ ($j = 1\ldots N$).[10]

To close the model, we must also specify the total supply of housing. In the short run, it is reasonable to suppose that the stock of housing is fixed. In the longer run, however, some supply adjustment is likely to take place. We do not explicitly model the microfoundations of the construction industry here. However, we can note that, for a given rate of population growth θ, the supply of new housing required to maintain a given average quality of housing - denoted here as $q = (\Sigma_j(1+\theta)^{N-j}h_j)/(\Sigma(1+\theta)^{N-j})$ - in equilibrium is equal to the product of that average quality, and the increment to the population occurring in the period, $s_t\,\Sigma(1+\theta)^N/\Sigma(1+\theta)^{N-j}$, where s_t is simply the current population. With some simplification, this implies a stock supply for housing in period t, H_t as shown in (14).

$$H_t = H_{t-1} + q_t s_t \theta(1+\theta)^N/[(1+\theta)^N - 1] \tag{14}$$

Making the usual assumption that the supply curve is upsloping therefore implies a supply price that is increasing in both population growth and average quality.

In the steady state equilibrium of the present model, the home ownership rate is constant. That is, in each period, the oldest cohort still renting ($z-1$) can meet the borrowing constraints and become homeowners at age z. For this to be true, the highest price Pp_{z-1} that the cohort can pay for their preferred level of housing services h_{z-1} must equal or exceed the maximum price Pp_a that older cohorts are willing to pay to add these homes to their portfolio of rental property. If the older households were not also subject to the same borrowing constraints as the younger households, this maximum price for investors would be found by rearranging the relationship equating the returns obtained from leveraged acquisition of rental property, with contributed equity $(1-\omega_a)Pp_a\Delta a$, where $\Delta a = h_{z-1}$, with that from holding an amount of bonds equal to this contributed equity, as shown earlier in (10), to obtain an expression for p_a (15).

$$p_a = R^*/[i - m(1 - m(1-\omega_a) - (\dot{p}^e + \pi + \dot{p}^e\pi)] \tag{15}$$

Since the young households in cohort $z-1$ are bidding against the potential property investors amongst their elders, the maximum price the investors are willing to pay is also the price that the young households end up paying, conditional on them succeeding in entering into home ownership. Thus p_a would then also be the price that enters into the repayment and downpayment constraints on the young households.

All households are subject to the borrowing constraints, however, which puts a limit on the amount of rental housing assets that older households can accumulate in any period. For example, suppose cohort $z+1$ borrowed the maximum allowed by the repayment constraint when they first became homeowners at age z. Then, allowing for nominal income growth $(1+\pi)(1+w) - 1$ and the principal repayment of their original debt, this cohort could borrow an amount equal to $\Delta a_{z+1}Pp$, as shown in (16). Therefore the actual relative price of housing p may be lower than the expression for p_a shown in (15).

$$\Delta a_{z+1} = \frac{W_{z+1}\psi(\pi + w + \pi w)}{Pp}\,\frac{(1+i)^T - (1+i)}{i(1+i)^T} \tag{16}$$

Because of the complexities of the interaction between the borrowing constraints and households' optimising behaviour, an analytic solution for the equilibrium outcome will not be presented here. In the next section, the qualitative effects of a disinflation are discussed, both in steady state and during the transition.

[10] This involves a normalisation of initial average wage rates to unity, with no meaningful implications for the results.

4.2 Effects of a disinflation

If households were not subject to borrowing constraints along the lines of (6) and (7), then a permanent, credible disinflation (fall in π) would have little effect on outcomes. In perfect foresight equilibrium, where the relative price of housing is expected to remain constant ($\dot{p}^e = 0$), the (unconstrained) price the older households are willing to pay for investment property simplifies to (17), using the Fisher identity to relate the nominal interest rate to the real interest rate r and inflation π, $i = (1 + r)(1 + \pi) - 1$. The role of the inflation rate in (17) is clearly of second-order importance, and is frequently ignored in approximated definitions of the nominal interest rate. Nonetheless, provided $r > \dot{p}^e = 0$, a disinflation does slightly increase the relative price that investors are willing to pay for housing assets, at the expense of the amount of housing services consumed by homeowners, at least initially at age z.

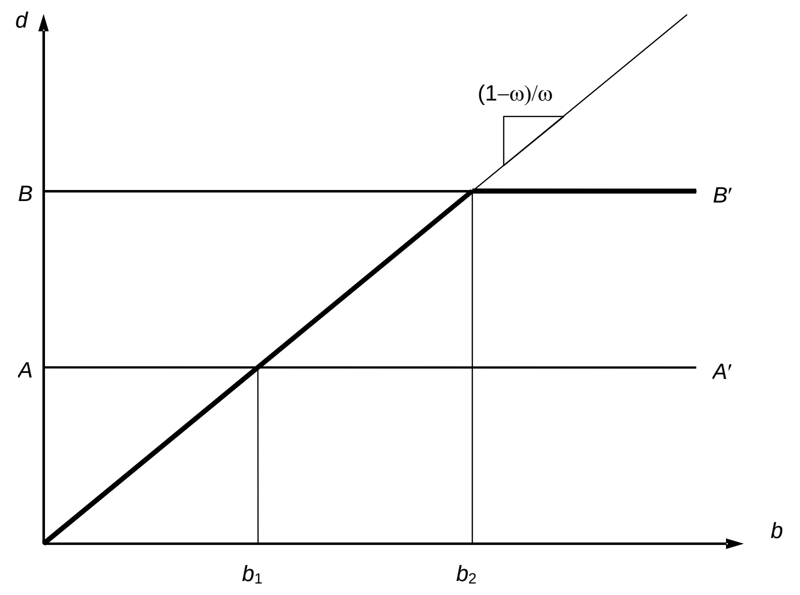
$$p_a = R^*/[(1+r)(1+\pi) - 1 - m(1-\omega_a) - \pi] = R^*/[r + r\pi - m(1-\omega_a)] \tag{17}$$

In the presence of borrowing constraints, however, the effect of a disinflation is potentially much greater. The effect of a disinflation on the two borrowing constraints is shown in Graph 5. The repayment constraint imposes a maximum total amount of debt d, while the downpayment constraint requires that this debt can be no larger than $(1 - \omega)/\omega$ times the deposit, or accumulated financial assets at the time of first-home purchase, b. For low levels of accumulated assets, the downpayment constraint binds but the repayment constraint does not. The combination of the two constraints results in a set of possible combinations of d and b that permit home purchase, represented by the area between the x-axis and the thick piecewise linear frontier between the origin and point B'. Unless the households' rate of time preference is noticeably below the real interest rate, or their preferred level of housing services dramatically different from the constrained level of finance, they will not generally choose to accumulate more financial assets than is necessary to meet the downpayment constraint, conditional on the repayment constraint being binding. Therefore the constrained equilibrium outcome will normally be the corner solution where both constraints are just binding. For example, the debt-asset combination consistent with point b_1 is the likely outcome of a repayment constraint that permits a maximum loan size of A.

A reduction in inflation, and thus nominal interest rates, results in the repayment constraint being consistent with a higher total level of debt. This is represented in Graph 5 by an upward shift in the horizontal part of the constraint frontier, say from AA' to BB'. The downpayment constraint is therefore the binding constraint over a wider range of possible levels of accumulated assets, up to b_2.

Graph 5

Effect of disinflation on borrowing constraints for cohort *z*



With borrowing constraints eased, households will clearly prefer to spend more on housing. If, however, the physical supply of housing is fixed, at least as a first approximation, this tendency will completely manifest in the price in steady state. Even if there is some supply response, prices will still rise to some extent, assuming an upward-sloping supply curve for the flow of new housing, as discussed above. The comparative statics result is therefore for higher p and z, which translates into a lower home ownership rate, but a higher debt/income ratio $\Sigma d_j / \Sigma W_j P$.

In the transition, the downpayment constraint binds by even more, because the young households, who had previously expected a lower price of housing assets, did not accumulate savings sufficient to meet the downpayment constraint given the higher new relative price of housing. In the first period after the disinflation, these households are priced out of the market and must continue to rent. Moreover, older households are not bound by the downpayment constraint to the same extent, because their equity increases disproportionately when p rises. Thus they can both increase their own consumption of housing services h and their holdings of rental properties.[11] This serves to bid up the price of housing assets, but not by as much as would occur if cohort z were not temporarily priced out of the market and were thus adding their demand to the total. The older households therefore expect that, over time once cohort z can return as first-home buyers, the relative price of housing will rise $(\dot{p}^e > 0)$. From (17), this implies that these older households perceive a sufficiently high return from ownership of rental properties that they are willing to hold the extra rental properties in the transition period.

Facing a permanently higher relative price p, the younger households still renting must accumulate sufficient financial assets b to meet the downpayment constraint. If the age at first-home purchase z rises, renting households by definition have longer to accumulate these assets. However, the externality between returns from renting and owning suggests that they will also have some behavioural response in order to minimise the increase in the time spent renting. Depending on the effects of the discontinuity arising from the fact that z must be integer-valued, this adjustment will come from a combination of lower consumption of consumption goods c and (rented) housing services h, as well as lower leisure l.

Even if the stock of housing is fixed only in the short run, and eventually expands to meet the increased demand, the transitory effects will still apply. These effects could be quite persistent; it will take at least z years before the z-aged cohort have experienced only the low-inflation state, and saved accordingly. In addition, the housing stock could take a long time to adjust. In the long run, however, if the housing stock adjusts, the rental rate R and price of housing assets p will return to (approximately) the levels prevailing before the disinflation. Therefore the relative (but not actual) price of a dwelling of constant quality will return to its pre-disinflation level. However, the median transacted price that is commonly used in housing price series will rise because the average quality of dwellings will rise.

5. Conclusion

The results presented in this paper depend entirely on the interaction of the repayment and downpayment, or deposit, constraints in intermediaries' lending decisions. These constraints were assumed to be a result of intermediaries' responses to imperfections in capital markets, particularly information asymmetries affecting the assessment of credit risk. If these imperfections are ameliorated at the same time as inflation falls, the effect on ownership rates could be reduced. Indeed, if there was no downpayment constraint at all, such that intermediaries were willing to lend 100% of valuation, the effect on home ownership rates disappears entirely.

The balance sheets of Australian households have been clearly affected by the consequences of disinflation. Debt/income ratios have risen rapidly since the early 1990s disinflation, with little sign as

---

[11] This kind of reallocation of housing services amongst cohorts assumes that, even though the stock of housing is fixed, it is freely divisible. In reality, households will tend to upgrade to a higher-quality home, or renovate the one they currently reside in; an easing in borrowing constraints should be expected to result in an increase in the average quality of dwellings. However, explicitly tracking the occupation of dwellings of specific quality by different households would require adding another dimension of heterogeneity to the model, making it even more complex than it already is.

yet that this process has completed. This expansion in credit has been associated with strong construction activity, both in the construction of new dwellings and in substantial renovation activity. As would be expected from the model presented in this paper, the average quality of new homes is also rising quite rapidly. Strong growth in ratios of household credit to income has also been observed in other countries with relatively deregulated financial sectors once they experience a sustained disinflation. Disinflation interacts with income-linked constraints on borrowing to increase housing indebtedness. This effect works in the same direction as the effects of financial deregulation in easing borrowing constraints. But the ensuing upward shift in housing prices implies that downpayment-type constraints on borrowing become more binding, not less. Thus although financial deregulation would be thought to increase home ownership rates by making finance more accessible, disinflation may actually reduce home ownership rates for younger age groups, at least in the short run until the housing stock adjusts fully. This effect via deposit constraints is in addition to any effect on ownership rates due to a reduction in the tax advantages of home ownership as inflation falls. Moreover, even if the increase in the relative price of housing is temporary, the transition can take considerable time to work through. This is because the housing stock takes time to adjust and young households take time to accumulate savings sufficient for a larger deposit.

Discerning these effects in Australian data is not easy, however. Ownership rates have certainly fallen for younger age groups, according to Census data, with the overall population ownership rate remaining constant because of population ageing. However, most of this decline occurred through the late 1970s and 1980s, rather than after the early 1990s disinflation in Australia.

The implications of these changes for intergenerational welfare are mixed. In the United States, at least, there is evidence that young households are relying more on gifts from their elders than on their own savings in accumulating the downpayment on their first home (Mayer and Engelhardt (1996)). As with transfers between generations, direct government subsidies to first-home buyers would also offset the increasing importance of the deposit constraint as inflation falls. Intergenerational transfers might seem like an understandable response in the transition period, when the older households bought their homes when inflation was high, and have thus experienced an unexpected windfall gain in housing wealth. But if the current growth in the relative price of housing is simply a transition to a new, higher equilibrium level, then currently young households will not have the same windfall gains to redistribute once they are old. Inheritances will be larger than when housing prices are low, but given population growth, they will account for only a constant fraction of the required downpayment for any given relative price level of housing. This suggests that younger households of subsequent generations might have to save more on their own behalf, relative to their incomes, in order to meet the downpayment constraint. This has obvious implications for the intergenerational distribution of consumption and leisure further into the future.

References

Almeida, H (2000): “Financial constraints and house prices - an international perspective”, *Leonard N Stern School of Business, New York University Finance Department Working Paper Series*, no FIN-00-017.

Börsch-Supan, A (1986): “Household formation, housing prices and public policy impacts”, *Journal of Public Economics*, 30(2), pp 145-64.

Bourassa, S C (1995): “A model of housing tenure choice in Australia”, *Journal of Urban Economics*, 37(2), pp 161-75.

Capozza, D R, R K Green and P H Hendershott (1996): “Taxes, mortgage borrowing and residential land prices”, in Henry J Aaron and William G Gale (eds), *Economic effects of fundamental tax reform*, Brookings Institution, Washington, DC, Chapter 5, pp 171-210.

Duca, J V and S S Rosenthal (1994): “Borrowing constraints and access to owner-occupied housing”, *Regional Science and Urban Economics*, 24(3), pp 301-22.

Ellis, L and D Andrews (2001): “City sizes, housing costs and wealth”, Reserve Bank of Australia, *Research Discussion Paper* 2001-08.

Gobillon, L and D le Blanc (2002): “The impact of borrowing constraints on mobility and tenure choice”, *CREST Working Paper* 2002-28.

Haurin, D R, P H Hendershott and S M Wachter (1997): "Borrowing constraints and the tenure choice of young households", *Journal of Housing Research*, 8(2), pp 137-54.

Hendershott, P H and M White (2000): "Taxing and subsidizing housing investment: the rise and fall of housing's favored status", *NBER Working Papers*, no 7928.

Hendershott, P H, G Pryce and M White (2002): "Household leverage and the deductibility of home mortgage interest: evidence from UK house purchasers", *NBER Working Papers,* no 9207.

Henderson, J V and Y M Ioannides (1983): "A model of housing tenure choice", *American Economic Review*, 73(1), pp 98-113.

Henley, A (1999): "The economics of the crazy British housing market", *Aberystwyth University Economic Research Paper*, no 99/8.

Howitt, P (1990): "Zero inflation as a long-term target for monetary policy", in R G Lipsey (ed), *Zero inflation: the goal of price stability*, C D Howe Institute, Toronto, pp 67-108.

Iacoviello, M and R Minetti (2003): "Financial liberalization and the sensitivity of house prices to monetary policy: theory and evidence", *Manchester School*, 71(1), pp 20-34.

Lamont, O and J C Stein (1999): "Leverage and house-price dynamics in US cities", *RAND Journal of Economics*, 30(3), pp 498-514.

Linneman, P and S Wachter (1989): "The impacts of borrowing constraints on homeownership", *Journal of the American Real Estate and Urban Economics Association*, 17(4), pp 389-402.

Mayer, C J and G V Engelhardt (1996): "Gifts, down payments, and housing affordability", *Journal of Housing Research*, 7(1), pp 59-77.

McCarthy, J and R W Peach (2002): "Monetary policy transmission to residential investment", *Federal Reserve Bank of New York Economic Policy Review*, 8(1), pp 139-58.

Meen, G P (1990): "The removal of mortgage constraints and the implications for econometric modelling of UK house prices", *Oxford Bulletin of Economics and Statistics*, 52(1), pp 1-23.

——— (2000): "Housing cycles and efficiency", *Scottish Journal of Political Economy*, 47(2), pp 114-40.

Ortalo-Magné, F and S Rady (1999): "Boom in, bust out: young households and the housing price cycle", *European Economic Review*, 43(4-6), pp 755-66.

Reserve Bank of Australia (2003): "Household debt: what the data show", *Reserve Bank of Australia Bulletin*, March, pp 1-11.

Stein, J C (1995): "Prices and trading volume in the housing market: a model with downpayment effects", *Quarterly Journal of Economics*, 110(2), pp 379-406.

Stevens, G R (1997): "Some observations on low inflation and household finances", *Reserve Bank of Australia Bulletin*, October, pp 38-47.

Throop, A W (1986): "Financial deregulation, interest rates, and the housing cycle", *Federal Reserve Bank of San Francisco Economic Review*, 3, pp 63-78.

Yates, J (1996): "Towards a reassessment of the private rental market", *Housing Studies*, 11(1), pp 35-50.

——— (2003): "The more things change? An overview of Australia's recent home ownership policies", *European Journal of Housing Policy*, 3(1), pp 1-33.

Zorn, P M (1989): "Mobility-tenure decisions and financial credit: do mortgage qualification requirements constrain homeownership?", *Journal of the American Real Estate and Urban Economics Association*, 17(4), pp 1-16.

Zorn, P M and H J Lea (1989): "Mortgage borrower repayment behaviour: a macroeconomic analysis with Canadian adjustable rate mortgage data", *AREUEA Journal*, 17(1), pp 118-36.

Financial behaviour of Dutch households: analysis of the DNB Household Survey 2003

P J A van Els, W A van den End and M C J van Rooij[1]
De Nederlandsche Bank

1. Introduction

As a result of the economic downturn, the financial position of Dutch households has deteriorated. Disposable incomes are depressed by rising pension and healthcare premiums, wage moderation and increasing unemployment. The stock market crisis has also affected the financial position of households. Net wealth (including pension savings) as a percentage of GDP fell from 208 in 2000 to 145 in 2002. Combined with the uncertain economic prospects, consumer confidence has dropped to the lowest level since 1983. This has not stopped households from borrowing more. These past few years, Dutch households' indebtedness, incurred by mortgage loans in particular, has continuously increased.

The first part of this paper looks at the financial balance sheet of Dutch households from an international perspective, on the basis of macro data. The second part deals with the principal items of the financial balance sheet from the point of view of the households themselves. The focus will be on the rising mortgage debt of households, and on the related risks. The fact is that the household sector has grown more susceptible to developments in asset prices and mortgage interest rates. The survey makes it clear how households have cashed in on the steady fall of mortgage interest rates seen in the past decade or so. This outcome offers several points of departure for an analysis of the financial stability risks that may ensue from the recent rapid rise in interest rates. Furthermore, this paper investigates the effects of expenditures out of mortgage equity withdrawal, while quantifying the macroeconomic consequences thereof. Also, the role of tenants is considered. With house prices rising, tenants eager to buy a house must consider whether they should cut down on their expenditure in favour of savings. In addition, the equity holdings as well as the post-stock market crisis behaviour of Dutch households are highlighted. Special attention is paid to the attitude and expectations of the Dutch public as regards their old age pensions, which are in the spotlight these days owing to the dwindled pension savings and the ageing of the population. On the basis of the survey, the pensions issue is viewed from the households' angle. The paper concludes by looking into two recent phenomena that are relevant to households' saving behaviour: deflation (expectations) and the unfreezing of company saving scheme balances. In a sense, the present survey is a follow-up to the Bank-commissioned surveys conducted by the market research company NIPO in the spring of 2000 and 2002, and reported in the Quarterly Bulletins of June 2000 and June 2002.[2] The Bank intends to report annually on the financial behaviour of Dutch households, drawing on the DNB Household Survey (see Box).

2. The household balance sheet in a macro perspective

2.1 The balance sheet of Dutch households internationally compared

Table 1 presents an overview of the financial balance sheets of households in the Netherlands as compared with those in the euro area, the United Kingdom and the United States, taking the situation in 2001 as gauging point, this being the last year for which comparable data for all countries

[1] We would like to thank G Gajapersad and R B M Vet for excellent statistical support. Views expressed are those of the individual authors and do not necessarily reflect official positions of De Nederlandsche Bank (DNB - the Netherlands Bank).

[2] See DNB (2000, 2002).

are available. For maximum comparability, the balance sheet items are expressed in terms of GDP. Dutch and UK households have accumulated comparatively large pension savings, especially when compared with the euro area. US households, on the other hand, invest substantially more in equities than is the case in Europe. Conversely, compared to their US and euro area counterparts, Dutch and UK households have smaller bond portfolios. On the liability side, Dutch households stand out for having run up relatively high debts. Besides, the rise of 32 percentage points in loans taken out by Dutch households in the period 1995-2001 is considerably higher than in the euro area (+6 percentage points), the United Kingdom (+12 percentage points) and the United States (+10 percentage points). This is related to the soaring increase in the value of Dutch owner-occupied homes, from 123% of GDP in 1995 to 202% of GDP in 2001, a rise that was much stronger than in the other regions under consideration. This development reflects the sharp increase in house prices. Compared to Europe, the value of owner-occupied homes in the United States is considerably lower. Balance sheet data covering 2002 show that due to the stock market crisis, Dutch pension savings have declined to 138% of GDP and equity wealth to less than 30% of GDP.

Box

DNB Household Survey

Surveys constitute a valuable instrument in analysing the financial behaviour and vulnerability of households. For this reason, De Nederlandsche Bank has entered into a sponsoring agreement with CentERdata, a unit of CentER Group, which is closely linked to Tilburg University. Specialising in internet surveys, CentERdata annually questions approximately 1,500 households (over 2,500 persons) about their financial characteristics and behaviour (eg their saving and investment behaviour, their housing wealth, mortgage and other debts, accrued pension rights etc). These data are made available, free of charge, for scientific research. This research contributes significantly to the insight into the financial behaviour of Dutch households and the underlying motives. Questionnaires are extremely flexible as they permit introducing topical issues besides the standard elements, such as old age pensions, deflation and the unfreezing of company savings scheme balances.

Over a number of consecutive weekends, the CentER panel members are asked to complete a variety of questionnaires. A well balanced selection of members ensures that the panel is representative of the Dutch population. It is not a prerequisite that a panel member has a computer or internet access at his or her disposal. Besides, the questionnaires are put out several times over in order to maximise response. This being an annually recurring exercise, it permits monitoring developments over time (the database, initially named VSB Panel and later referred to as CSS (CentER Savings Survey), goes back to 1993. This paper proceeds from the results for 2003, which are preliminary to the extent that they draw on the replies to the first questionnaire. This implies that the analyses are derived from roughly 1,200 households and - depending on the questionnaire’s subject matter - a maximum of 2,000 persons.

2.2 Persistent debt growth

Also during the recent downturn, the debt of Dutch households continued to increase. In addition to the surge in house prices, the financial behaviour of households was a contributory factor in this trend. In the past two years, the sharp interest rate fall influenced household behaviour significantly. The fall in interest rates made it attractive to take out loans to maintain the level of spending. Hence, Dutch households’ debt continued to rise notwithstanding the economic downswing. This development is in contrast with that seen during the downturn in the early 1990s, when the debt level was found to stagnate (Graph 1). In the second quarter of 2003, mortgage debt, which dominates private debt, peaked at 79% of GDP. A persistent increase in debt during a cooling housing market evidences that more is being borrowed than is required to finance owner-occupied homes, among others, by refinancings and second mortgages. With interest rates being low, refinancing is attractive as it helps reduce monthly expenses. By raising the loan amount, households broaden their financial flexibility even more. While constituting a macroeconomic impulse, the persistent increase in debt is not without risks. On the basis of the survey results, both these aspects of mortgage loans are examined further.

Table 1

Composition of household balance sheets in 1995 and 2001 (% GDP)

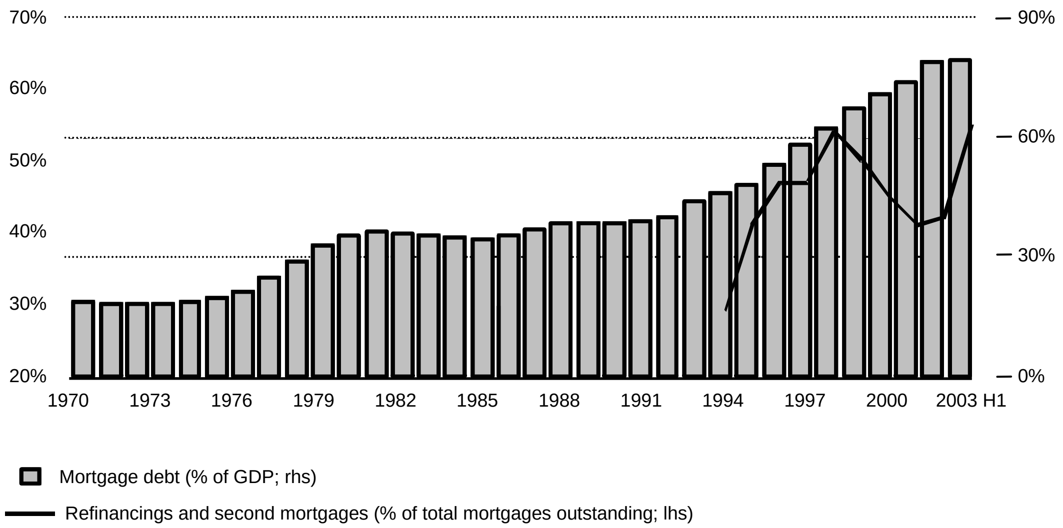
	Netherlands		Euro area[1]		United Kingdom		United States	
	1995	2001	1995	2001	1995	2001	1995	2001
Financial assets								
M3	55	56	66	61	64	71	45	48
Bonds	9	8	24	19	6	5	26	19
Equities	47	51	40	66	49	55	133	149
Pension reserves[2]	128	153	36	50	137	152	84	95
Others	5	6	3	3	9	9	4	4
Total financial assets	243	274	169	199	266	292	292	314
Financial liabilities								
Loans	63	95	45	51	66	78	66	76
Others	0	0	3	5	7	6	3	4
Net financial wealth	181	178	121	143	193	208	223	234
Value of owner-occupied homes	123	202	194	214	146	198	114	130

[1] All euro area countries, with the exception of Greece, Ireland and Luxembourg. [2] Including life insurances and other insurance technical reserves.

Sources: Statistics Netherlands; Eurostat for EU countries; Flow of Funds Accounts for the United States (website Board of Governors of the Federal Reserve System). The value of owner-occupied homes reflects own calculations on the basis of data from the Netherlands Bureau for Economic Policy Analysis and national central banks.

Graph 1

Development of household mortgage debt



Sources: DNB; Statistics Netherlands.

3. Mortgage loans

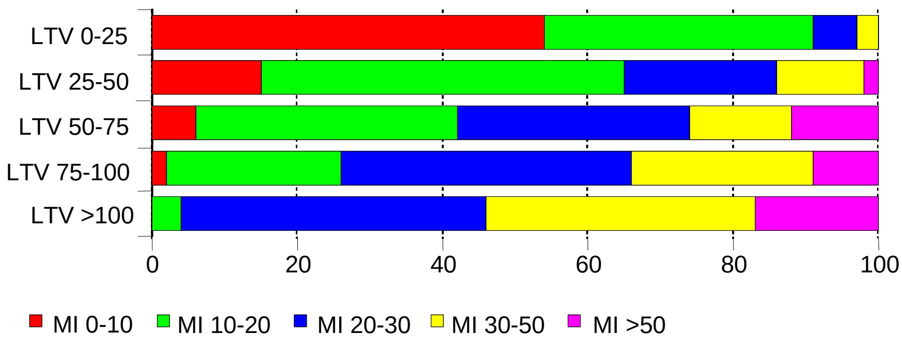
3.1 Risks entailed by mortgage debt

Households can afford higher mortgage debt as the lower mortgage interest rates keep housing costs low. This is corroborated by the survey results. While the share of top-up mortgages (mortgage higher than the property's purchase value) increased from circa 60% in the previous years to over 75% in the period from 2001 onwards, households substantially reduced the interest due on their mortgage loans. Around mid-2003, the average mortgage interest rate was 5.6%, 60 basis points down from the rate that households paid in 2000, according to the survey conducted that year. This development masks that the gross housing costs of some households with top-up mortgages are quite high (Graph 2). One in six households with loan-to-value ratios (LTV, or the ratio between mortgage and the market value of the owner-occupied home) in excess of 100% spends more than half its net disposable income on mortgage debt service (ie a debt service to income ratio >50%). This category of households is vulnerable to financial setbacks. In a scenario of falling house prices, households with top-up mortgages are the first to be confronted with a residual debt in the event they have to move house. Besides, if the debt service to income ratio is high, any loss of income may soon make it impossible to pay the monthly housing charges. The risks that payment problems of the most vulnerable households carry for the financial system as a whole, however, are limited (Van Rooij (2002)). Households with LTVs over 100% and mortgage debt service to income ratios over 50% represent approximately 0.4% of the population. A breakdown of the LTV by age group shows which households are the most vulnerable (see Graph 3). Top-up mortgages are concentrated in the 25-34 age bracket. People in that category, usually being starters on the housing market, are compelled to go to the limits of their finance potential to be able to buy a home.

Graph 2

Mortgage debt service to income ratio (MI) per LTV category

MI and LTV in percentages



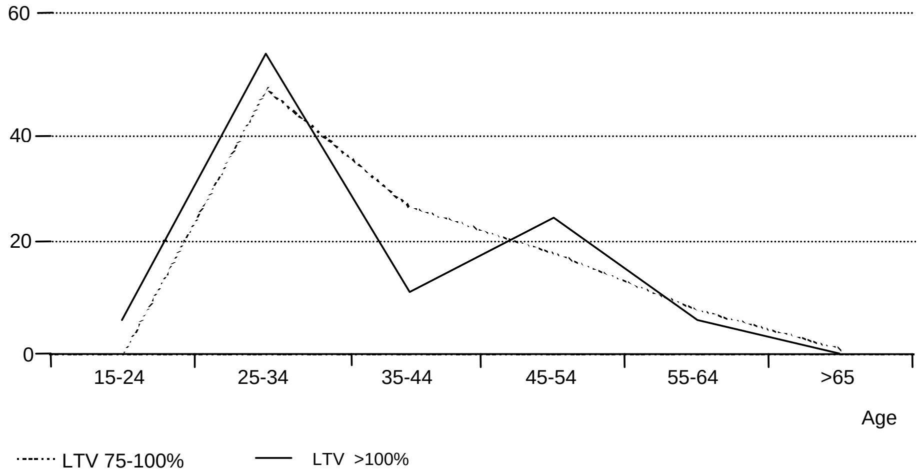
Source: DNB Household Survey, June 2003.

One significant risk for homeowners is entailed by interest rate movements. A rising mortgage interest rate would lead to higher housing costs if, at the time of renewal of a fixed rate contract, the prevailing rate is clearly higher. For variable rate contracts, a higher mortgage interest rate automatically leads to higher mortgage payments. The most common fixed interest period is still 10 years. This is the term agreed for 32% of the mortgages outstanding, while the rate is fixed for five years in 23% of the cases, and variable for 15%. For the interest rate related risks incurred by households, it is not the term of the contracted interest period that is relevant though, but the expiry date. Graph 4 shows that about half the outstanding mortgage loans have a remaining term to maturity of four years or less (even while only 20% opt for a variable or fixed interest period of less than five years). For over one quarter of mortgage contracts the rates will be adjusted before the end of 2004. The rates contracted for a significant number of these mortgage loans are a great deal higher than the prevailing rates. However, a large group of households stand to be facing higher housing charges in the near future, should mortgage interest rates begin to surge.

Graph 3

Distribution of age per LTV group

Percentage of households in mentioned LTV group

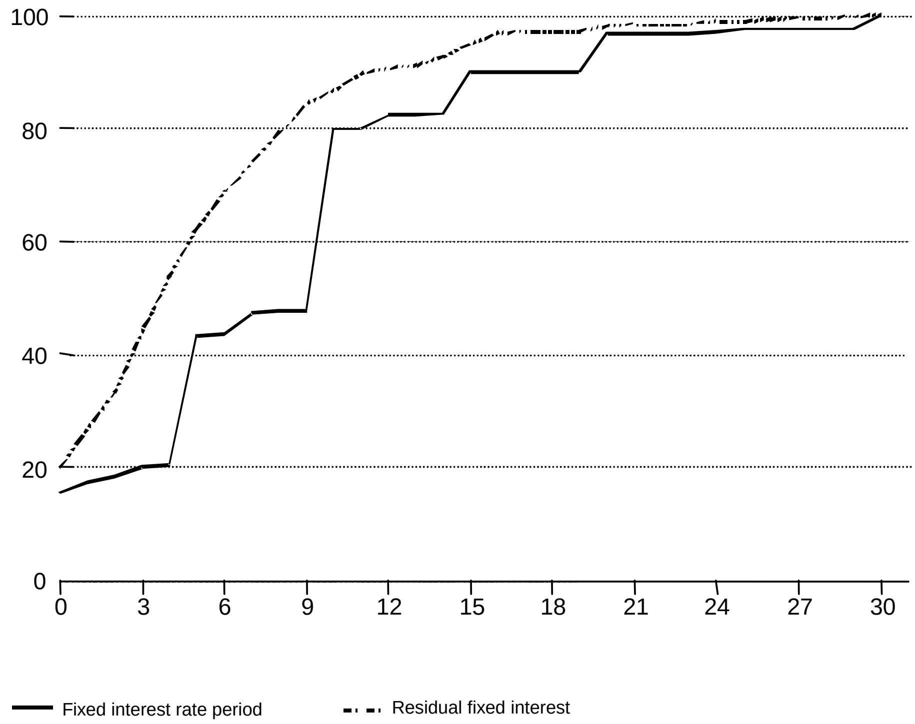


Source: DNB Household Survey, June 2003.

Graph 4

Cumulative distribution of interest rate period and years until rate adjustment

Percentage of the number of mortgages outstanding



Source: DNB Household Survey, June 2003.

3.2 Mortgage loans for balance sheet restructuring

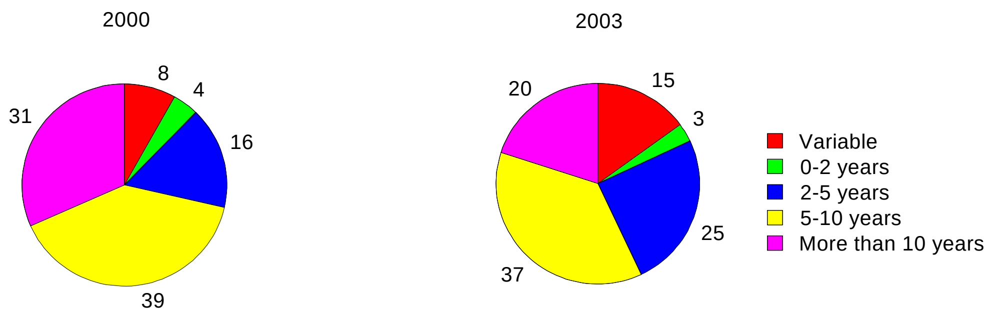
Although at first sight household balance sheets have become more imbalanced owing to the persistent debt rise, the ample availability of mortgage loans also offers opportunities to improve liquidity or reorganise the financial position by restructuring the various assets and liabilities, or extending the contractual interest period. In the United States, balance sheet restructuring is one of the main reasons for cashing out home equity (Federal Reserve (2002)). The most recent DNB Household Survey was designed to examine whether this was also the case in the Netherlands.

The survey first sought to assess how households had responded to the fallen interest rates. In a climate of declining mortgage interest rates, two considerations may play a role in setting the interest rate terms in the mortgage loan. On the one hand, it is possible that further interest rate declines are expected and that households speculate on this by opting for a variable rate or a brief fixed interest period. This is exactly what happened in recent years, causing the percentage of mortgage loans with a variable interest rate to rise. At present, 15% of all mortgage loans outstanding were contracted at variable rates, against 8% according to the survey commissioned by the Bank in 2000. This increase has made Dutch homeowners more vulnerable to interest rate movements. On the other hand, the interest rate decline may be expected to have bottomed out and prompt households to fix interest rates. This expectation may have been fuelled by the fact that in July 2003 mortgage interest rates reached the lowest level in over 40 years. Rather than inducing households to opt for a longer fixed interest period, however, the percentage of mortgage loans with a fixed period of 10 years or longer dropped compared to three years earlier in favour of mortgage loans with a term to maturity of five years or less (Graph 5). In other words, Dutch households have not profited from the fallen rates by reducing their interest rate vulnerability. This may be related to the steepening of the interest rate curve seen since 2000, as a result of which rates for short-term loans fell relatively more than those for long-term loans. There is a risk that households fixing mortgage interest periods for a short term are underestimating the odds of an interest rate rise. According to the survey, only 3% of households regard a sizeable rate hike as a source of mortgage payment problems. Unexpected personal circumstances are considered as having a greater impact on a household's capacity to defray housing costs. As principal factors in payment problems, 50% of households name unemployment or disability, and 18% divorce.

Graph 5

Fixed interest rate period of mortgages

Percentage of the number of mortgages outstanding



Sources: DNB Survey of March 2000; DNB Household Survey of June 2003.

Also, a mortgage loan top-up (by way of an equity withdrawal) may be used to release funds for reorganising the financial balance sheet, eg by redeeming other, often more expensive loans. About a quarter of the surplus value realised in the United States has been parlayed for paying off relatively expensive consumer credit and credit card debts. In the Netherlands, this practice is relatively less common. Of the home equity cashed out since 1998, just 6% was used to repay other loans. By using equity withdrawal for investment or portfolio investments, rather than to restructure debt, households are rendering themselves more vulnerable to financial setbacks. After all, they secure a (higher) mortgage loan with financial assets of fluctuating value. An important drive behind taking out mortgage loans for the purpose of portfolio investments is interest arbitrage. Interest arbitrage is lucrative as long as the effective yield on investments (minus any capital yield tax) is higher than the effective mortgage interest rate (interest rate after possible taxes). Driven by the declining mortgage rate, this circumstance has in recent years increasingly influenced the borrowing behaviour of homeowners. It is, in part, due to this behaviour that in the past five years about 10% of cashed-out home equity was turned into financial assets, almost two thirds of which consisting of equities and other portfolio investments. Expenditure on non-financial assets mainly concerns investment in owner-occupied homes by way of home improvements. While, as a rule, such investment enhances the value of the owner-occupied home, just as with financial investments it holds that households investing on the housing market with borrowed money make themselves more vulnerable to movements in asset prices and interest rates. Indeed, being disadvantageous for the value of houses and equity wealth, (in the course of time) a higher rate will lead to higher mortgage payments.

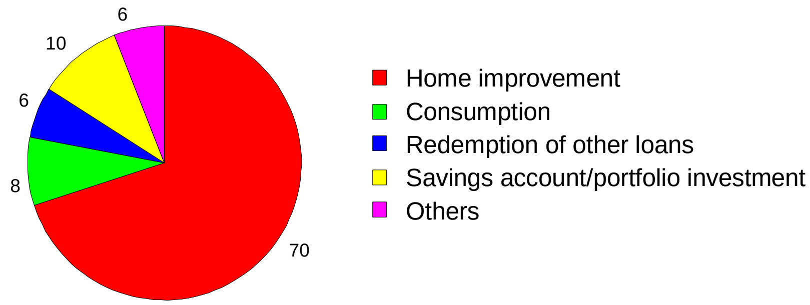
3.3 Effects of home equity withdrawal on expenditure

Withdrawing home equity has partly been permitted and encouraged (feel-good factor) by the sharp price rises on the housing market in the period around the turn of the century. Indeed, becoming, and often feeling, more affluent, homeowners tend to adjust their consumption patterns. Besides for investment in owner-occupied homes, equity withdrawal is spent on durable consumer goods, electronics and holidays (Graph 6). Former calculations using the macroeconomic mode l MORKMON showed that these expenditures could exert a considerable effect on the economic development of the Netherlands, with contributions to economic growth varying from roughly 1 percentage point in 1999 and 2000 to –0.5 percentage points in 2001 (DNB 2002)).

Graph 6

Expenditure out of home equity withdrawal

Percentage of surplus value realised



Source: DNB Household Survey, June 2003.

The present data are used to extend the analyses of the macroeconomic effects with the years 2002 and 2003 (Table 2). Rather surprisingly, the survey results indicate that equity withdrawal related expenditure in 2002 equalled or even slightly exceeded the previous year's level. This notwithstanding, the model calculations show a negative contribution to growth in 2002 and 2003, by about 0.5 percentage points and 0.25 percentage points, respectively. This negative growth contribution reflects the sharp fall in equity withdrawal related spending after 2000.

The continuation of the level of home equity cash out of 2001 into 2002 is surprising against the background of the declining consumer confidence and house price rises. It is worth noting that it looks as if equity withdrawal related expenditure in 2003 will again turn out to be at least on a par with the preceding year's level. This trend is not only indicated by the survey data for the first six months, but also confirmed by the increase in refinancing and second mortgages registered by Statistics Netherlands. By all appearances, the mortgage rate fall by almost 2 percentage points since the second quarter of 2002 has promoted the mortgage-related finance of specific expenses. Indeed, a low interest rate also means low finance charges. Consequently, interest rate movements may have a considerable bearing on (the timing of) such expenses. Graph 7 in any case shows that there is a clear relation between the interest rate movements, on the one hand, and refinancing and second mortgages, on the other hand.

Table 2

Effects of spending impulse from mortgage equity withdrawal

In percentage points, unless stated otherwise

	1998	1999	2000	2001	2002	2003
Assumptions						
Domestic expenditure impulse (level, EUR billions)[1]	3.1	6.8	9.9	4.5	4.7	5.0
Results according to MORKMON						
Volume growth of domestic expenditure	1.0	2.0	2.2	–1.1	–0.9	–0.4
– of which directly from expenditure impulse	0.6	1.0	0.7	–1.4	0.0	0.0
GDP volume growth	0.5	1.0	1.1	–0.5	–0.5	–0.3
Development of remuneration per employee	0.1	0.2	0.4	0.5	0.7	0.5
Inflation	0.0	0.1	0.2	0.3	0.8	0.5
Private employment growth	0.1	0.3	0.5	0.5	0.1	–0.1

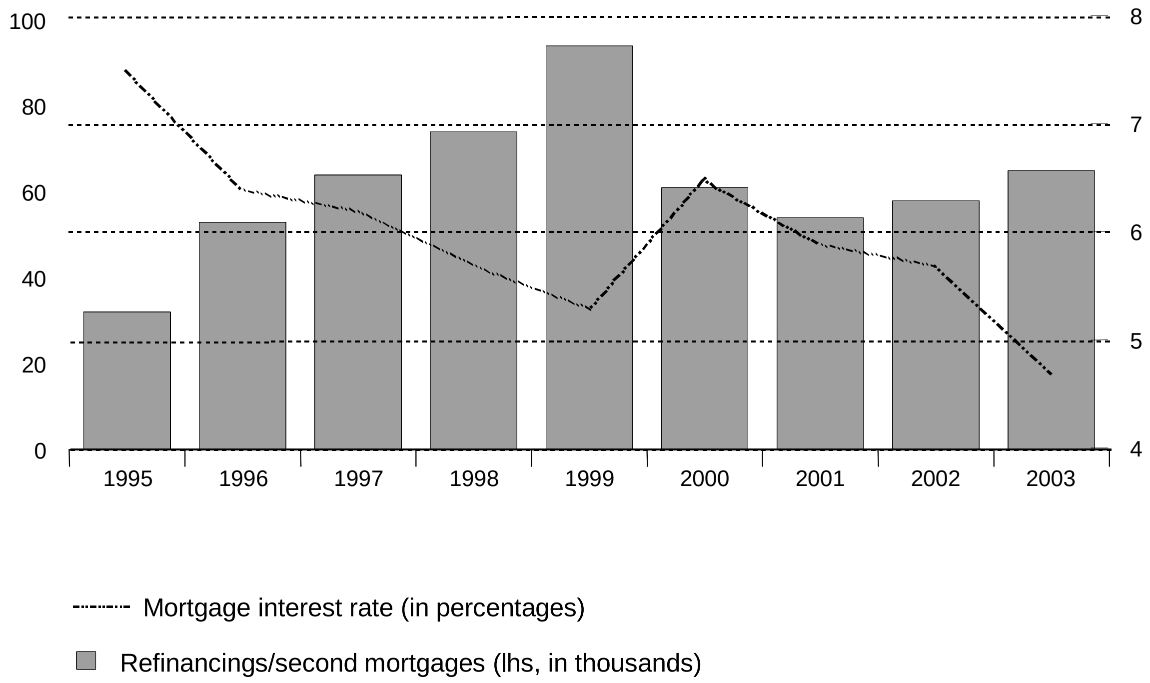
[1] For an accurate estimation of the effects for 1998, the calculation was based on equity withdrawal related spending impulses of EUR 1.2 billion and EUR 0.9 billion in 1996 and 1997, respectively. The figure for 2003 is an estimation based on the first six months of that year.

Sources: DNB surveys (March 2002 and June 2003).

Graph 7

Mortgage interest rate and refinancings/second mortgages

Quarterly averages



Sources: DNB; Statistics Netherlands.

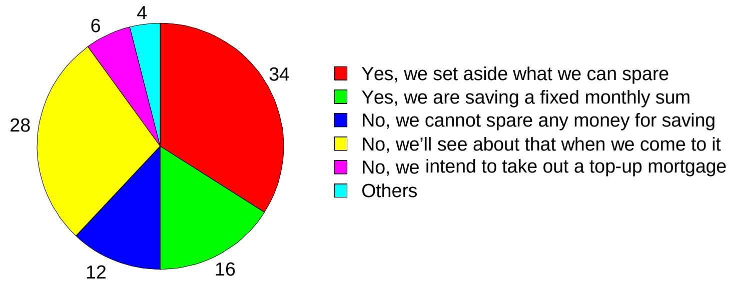
3.4 Expenditure effects: homeowners versus tenants

A few brief comments on the above analysis seem in order. Firstly, it is not clear how high expenses would have been, had they needed to be financed in ways other than mortgage loans. Secondly, as is inherent in a survey, the reliability of the results depends on the accuracy of the respondents' replies. Thirdly, tenants may have stepped up their saving as, owing to the price rises, the type of dwelling they have in mind is increasingly moving beyond their reach. This dampens the expenditure effects of equity withdrawal by homeowners. Recent empirical research even arrives at the conclusion that, on balance, the effect of higher house prices on the economy will be limited (Alessi and Kapteyn (2002)). This conclusion is not supported by the survey results. Over a quarter of households living in rented accommodation would like to purchase a home. A large part of this category, however, states that they do not specifically save towards this goal (Graph 8). It is not very likely that rising house prices will have an effect on the saving behaviour of this category of tenants. However, more than half of those intending to buy a home - in the shorter or longer term - put aside money for this purpose, the majority saving what money they can spare. In this case, too, the house price movement does not affect their saving behaviour either. In reply to the direct question whether higher house prices lead to additional saving, only 9% of renting households that are saving in order to be able to buy a house reply in the affirmative.

Graph 8

Are you saving with a view to buying your own home?

Percentage of tenants intending to buy a home



Source: DNB Household Survey, June 2003.

4. Equity holdings and the stock market crisis

4.1 The effects of the stock market crisis on financial behaviour

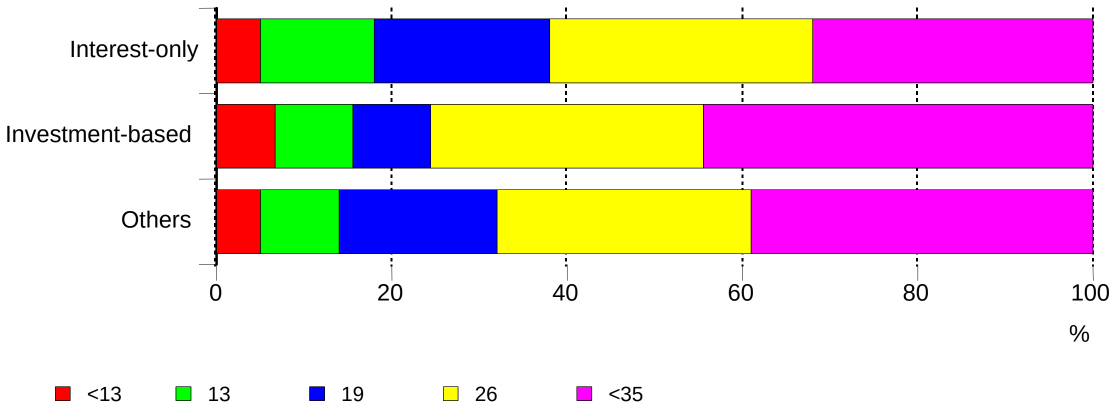
Besides the development of the prices of dwellings, the stock market crisis has also influenced Dutch households' financial behaviour. Two per cent of the respondents who owned equities or trust fund units indicate that they made a profit (EUR 2,000 on average) on their equity portfolio. They succeeded in doing so despite the fact that share prices are now substantially lower than the peak values they reached in the bullish period three years ago. More than three quarters of private investors state they have sustained evident losses (EUR 20,000 on average). Nonetheless, private investors have not turned their backs on the stock market in droves. According to the survey, since the onset of the crisis, some 10% of investing households have largely or wholly disposed of their equities, while roughly 10% have reduced their equity portfolios. The latter category were also asked in what year they shed most of their equities. It turns out that most did not sell until long after the stock market slump set in; 19% disposed of the largest block of shares in 2000, 33% in 2001, 38% in 2002, and 10% in 2003. Against the group of investors that reduced their equity portfolios (20% of those holding shares) is a small part (4%) that increased theirs, and a large group (more than 60% of equity holders) who held on to most of their portfolios, hardly extending their holdings, if at all. Apparently, a great many are able and willing to absorb the decline in their equities' value. This is probably accounted for by the fact that equity holdings are concentrated in the wealthiest households (DNB 2002)). This group has relatively large capital buffers to be able to absorb asset price shocks. Nonetheless, it is conceivable that the drop in share prices has also altered the financial behaviour of households that left their equity holdings intact. The survey reveals that the stock market crisis has made investing households more risk-averse. One third has changed to investing less or saving more, whereas another 10% indicate that they have adopted a more conservative spending pattern. Households have not proceeded to borrow less, though. This is confirmed by the macroeconomic trend of a persistent rise in household debt.

The effect of the stock market crises may also make itself felt through mortgages. Households investing in equities by monthly contributions towards investment-based or endowment mortgages have been confronted with a drop in the value of their investment trust. An undervalue usually needs to be supplemented when the mortgage matures, or when the contract is prematurely cancelled. It is also conceivable that a bank will require additional contributions when the mortgage is renewed. According to the survey, by mid-2003 19% of the mortgages outstanding consisted of investment-based or endowment mortgages. Graph 9 shows that the higher income categories exceed the lower income categories as to number of investment-based mortgages. Mortgage owners in the low income brackets are the most vulnerable to disappointing yields on their investment trust, as they have fewer buffers to cushion any residual debt or a rise in monthly costs. Such risks rarely manifested themselves in the previous two years. According to the survey, in a mere 1% of the cases an additional payment or higher contribution was required as a result of the fall in share prices. This is related to the fact that, generally, investment-based mortgages do not involve a contractual obligation to make additional deposits. This is why homeowners do not regard the stock market crisis as a potential source of payment problems. Only 3% of households indicate that in the event of a further sharp fall in share prices they would start having difficulty meeting their mortgage payment obligations.

Graph 9

Breakdown of mortgage types by income bracket

Income in thousands of euros



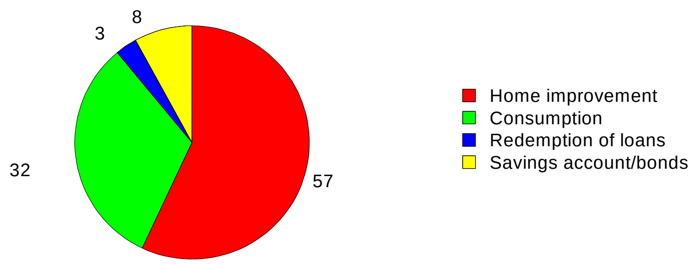
4.2 Expenditure effects of equities sales

Private investors who had disposed of (part of) their equity holdings were asked what they had done with the proceeds. The respondents stated they had used a large share for expenditures and only a smaller share to redeem debts or invest more safely (Graph 10). Compared to the home equity withdrawal related expenditures, a larger share is used purely for consumption (eg the purchase of furniture, electronics and vacations). The economic interest of home equity withdrawal related spending is many times larger than that of spending related to equities sales. On the one hand, the number of households with owner-occupied homes is more than twice as high as the number of households with equity (or unit trust) holdings (52% against 23%) and, on the other hand, the amount involved by home equity withdrawal is considerably higher than that realised in equities sales (on average, more than EUR 30,000 against EUR 10,000).

Graph 10

Expenditure from equities sales

Percentage of total proceeds



Source: DNB Household Survey, June 2003.

4.3 Consumer credit

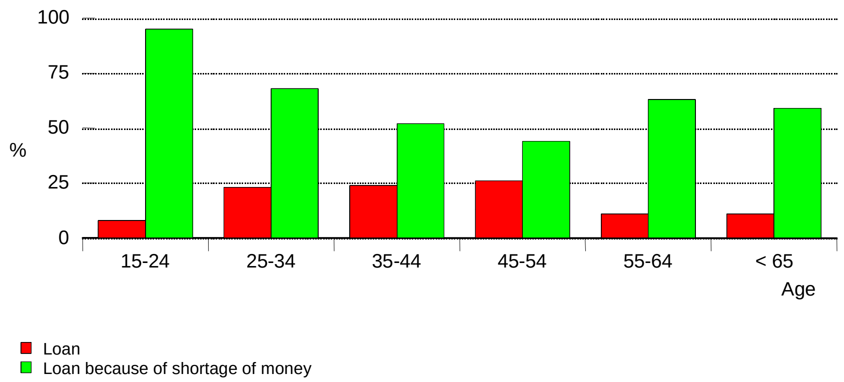
The ascent of the stockholding culture is sometimes related to the exuberant borrowing behaviour of households (Haliassos and Hassapis (2002)). One of the underlying causes of this development is that the equity risk premium has led households to cherish higher expectations of the growth of their wealth than households that do not hold equities. This equals out the additional risk on equities, making the first group of households more strongly inclined to step up spending and borrowing. This theory would appear to hold true for Dutch households, too. According to the DNB Household Survey, the percentage of personal loans, continuous personal loans or credit card debt is higher among households that invest in equities than among households that do not (27% against 19%). Having a mortgage debt also appears to be a factor in taking out consumer credit. Of mortgage-burdened households, 18% have consumer credit and of households without mortgage loans, 21%. This indicates that homeowners sometimes use their mortgage as an alternative to consumer credit.

In the majority of cases (56%), households resort to borrowing on account of a (temporary) lack of money. Not surprisingly, this motive is stronger for the low income brackets (72% of households with a disposable income up to EUR 19,000) than for the high income bracket (45% of households with a disposable income over EUR 35,000). The difference in motive for borrowing widens if also the age of households is taken into account (Graph 11). Young people (age 15-24) practically always borrow because they are short of money, whereas the over-50s do so in less than half of the cases. It is no coincidence that the fewest borrowing restrictions apply to the middle age group. According to the DNB Household Survey, 30% of total consumer credit is outstanding at households in the 45-54 age bracket.

Graph 11

Breakdown of non-mortgage loans by age category

Percentage of households with loans and of households in age categories with loans respectively



Source: DNB Household Survey, June 2003.

5. Pensions

From the balance sheet of Dutch households it appears that pension rights are a sizeable wealth component. The declines in the stock market have decreased the assets of pension funds by EUR 33 billion since the end of March 2000, putting the second pillar pension schemes under pressure. Practically all pension funds have taken measures towards improving their solvency position. Also, measures concerning first pillar pensions (public pension scheme) cannot be precluded against the background of the ageing population. In the survey, Dutch residents were asked to give their opinion of these measures and indicate what they had noticed so far in this context. Furthermore, the survey inquired after the respondents' insight into their own pension situation, their expectations for the future and their preferences regarding old age provisions.

5.1 Awareness

It turns out that only 13% are aware of measures related to eroded pension savings, such as raising pension contributions for employees (8%), for employers (3%), and the partial execution of the customary indexing (2%). Although it may well be that a number of respondents have hardly noticed anything of any measures by pension funds, eg when measures taken are of little financial consequence or - as in the case of a non-contributory pension - that only the employers' contributions to social insurance have increased, this percentage suggests that people give their pension arrangements relatively little thought (Table 3). In addition, 44% do not know whether their pension schemes are final or average pay based, or whether they depend on the yields on the contributions deposited; 45% cannot tell if their pension rights are indexed; while as many as 61% have no idea how much they have accrued, despite the pension overviews they have received. Finally, 65% have no idea as to what they may expect to receive on turning 65. Insight into the individual pension arrangement improves with age, though. This points to an increasing interest in the pensions arrangement as the retirement age gets closer.

Table 3

Awareness of own pension provisions

Percentages

Age (in years)	Number of respondents aware of:[1]			
	Type of pension scheme	**Indexing**	**Current pension rights**	**Eventual pension rights**
16-24[2]	25	11	0	18
25-34	35	36	21	29
35-44	52	57	34	35
45-54	64	58	48	39
55-64	65	57	67	42
65 and older	77	79	na	na
Total	56	55	39	35

[1] Those that have not ticked "don't know/no reply" in reply to the following questions: (1) How are your pension rights built up (based on final pay/average pay/available contributions/other system/don't know)?, (2) Is your pension indexed (yes/no/don't know)?, (3) What pension rights do you estimate you have accrued at your current/previous employer (... euros per year/don't know)?, (4) How high do you estimate will your net pension be as a percentage of your last net income prior to your retirement (… %, don't know)? [2] The percentages in this age bracket are based on a few observations, as the labour participation rate in this category is relatively low and many companies do not permit pension accrual before the age of 25.

Source: DNB Household Survey, June 2003.

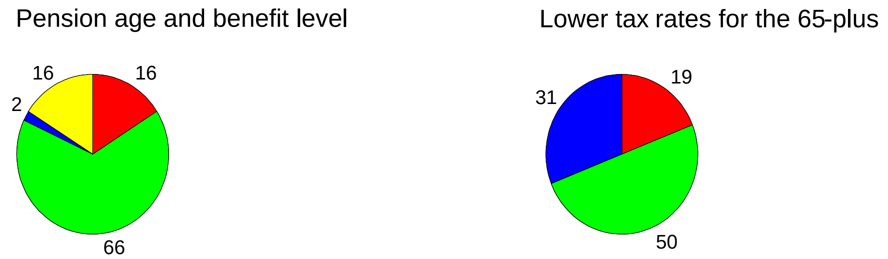
5.2 Concern about pension income

The fact that employees are familiar with their own pension rights to a certain extent only does not mean that the Dutch public are not well aware of the more general developments. The national discussion of the sustainability of the current pension arrangements against the background of an ageing population has in any case not escaped their notice (Graph 12). At least, a majority expect that the public pension scheme will be cut down in about 10 years from now. Two thirds foresee that the pensionable age will be raised and/or the benefits will represent less purchasing power, while only one in six respondents expect the situation to be similar to today's. This suggests that, compared to last year's survey, worries about the public pension scheme have increased, since according to that survey half of the respondents expected later and/or lower benefits in the future. Moreover, many people expect that in due course the difference in tax rates in favour of the 65-plus (as this age group no longer pays state pension contributions) will narrow, if not disappear.

Graph 12

Expectations regarding the public pension scheme 10 years ahead

Percentage of respondents



Similar to present situation
Will be raised and/or will be relatively lower
Will be lowered and/or will be relatively higher
Don't know/no reply

Similar to present situation
No or smaller difference
Don't know/no reply

Source: DNB Household Survey, June 2003.

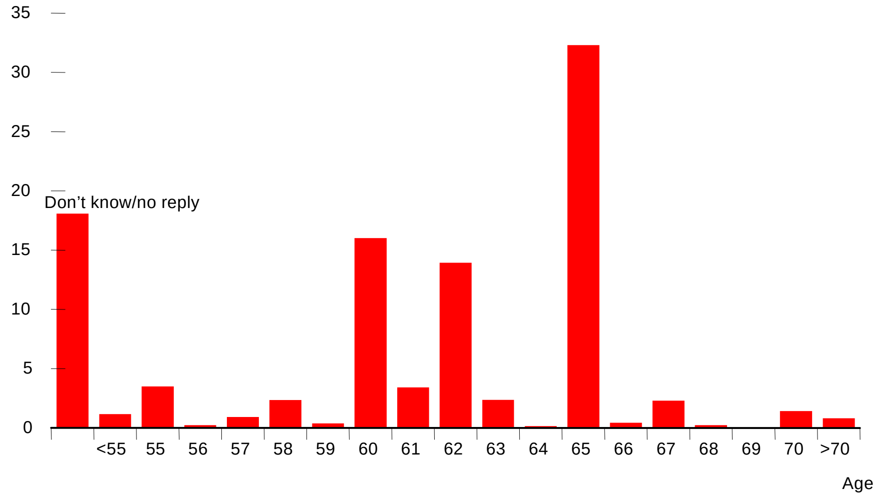
5.3 Retirement

That said, over 40% of employed people under the age of 65 expect to go on (early) retirement at the age of 62 at the latest, while just under 40% reckon to do so at the age of 65 or later (Graph 13). One third of the people in work have made other pension arrangements besides the regular pension scheme they participate in through their employer; in most cases, by way of annuity and single premium insurance policies. These may serve, on the one hand, to bridge the gap between the date of early retirement and the date on which the pension becomes payable, and on the other hand, to supplement their pension rights. Obviously, for a great many workers the pension build-up falls short of 40 years. A part of this category are single income household members keeping house. Roughly 15% of those interviewed state they will not be able to get by on just the (expected) income after their retirement (Graph 14). It turns out that half of the remaining percentage will just about manage to make ends meet and that the other half expects also to succeed in saving some money. Strikingly enough, those that claim still to be able to lay money by make up the majority in the 65-plus bracket, while most of those barely expecting to make ends meet are in the 65-minus category. The survey provides no answer to the question whether the 65-minus perhaps set higher demands on their lifestyle, whether they are cautious in their expectations, or whether they factor in a less favourable income pattern following measures designed to ensure that the ageing wave remains affordable. Another striking outcome is that those who do not get by on their income are eating into their savings, while the others largely manage on their partner's income.

Graph 13

Expected age of retirement

Percentage of respondents

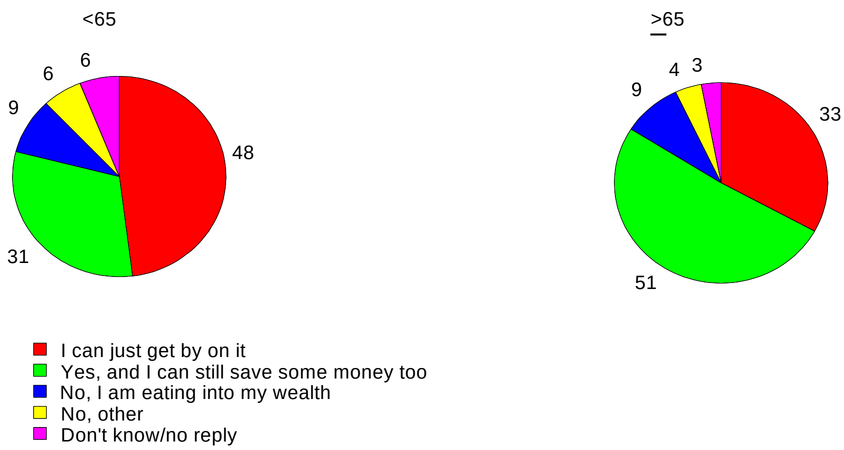


Source: DNB Household Survey, June 2003.

Graph 14

Can you get by on your (expected) income after 65

Percentage of respondents



Source: DNB Household Survey, June 2003.

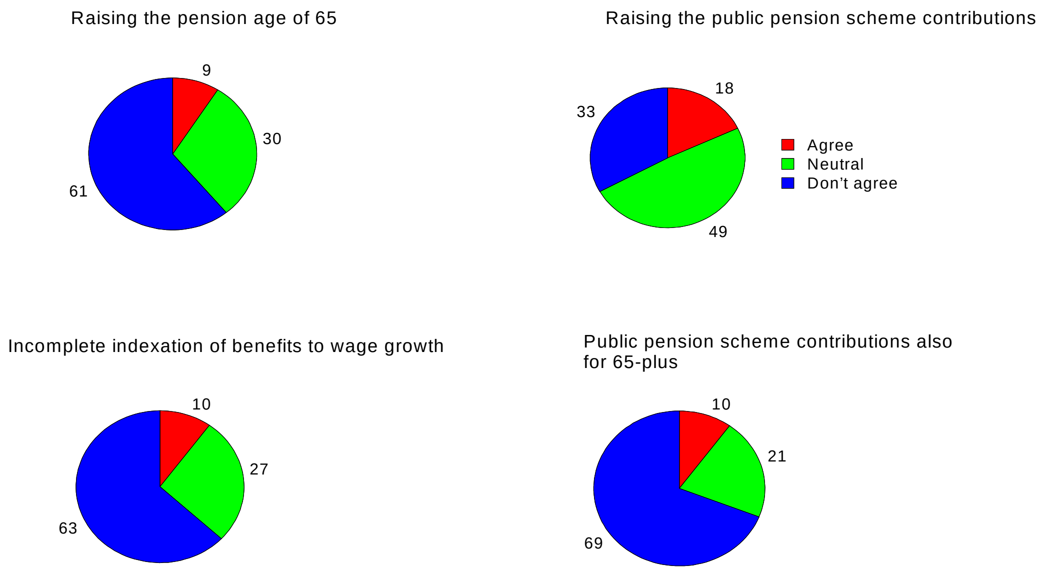
5.4 Measures to keep the public pension scheme affordable

From the foregoing it appears that the Dutch public are aware of the discussions about the increasing burden of the public pension scheme on public finance and about such measures as may be required to keep the first pillar pension scheme affordable. The panel members of DHS were asked for their opinion regarding four much suggested measures. From their replies it emerged that, without exception, these measures meet with much resistance (Graph 15). Notably, especially those three measures cutting down on the existing regulations (raising the age of retirement; incomplete indexation of benefits to wage increases; and levying old-age pension contributions on the 65-plus) are not much favoured. Two thirds are flatly opposed, while only one in 10 cannot see anything wrong in them. The least opposition is met by the option to impose a higher contribution on persons under the age of 65. It should be noted here that no amounts were specified in the questions regarding the higher contribution and other related measures. If the amounts realistically involved by the measures concerned were specified, though, and the respondents were confronted with a compulsory choice, the outcome might turn out differently. For example, if it appears that the public pension scheme contributions would need to rise more than is being envisaged, the replies might perhaps be coloured by self-interest. Even in the current replies a vague pattern can already be discerned reflecting this tendency, old-age pensioners or those nearing the age of retirement being sooner inclined to oppose measures they regard as an encroachment on acquired rights. Young people, on the other hand, are more inclined to support cutting down on these rights, and would rather not see the public pension contribution raised for the 65-minus. Self-evidently, what may come into play here is that the young are better able to absorb a retrenchment of these rights. Also, the “we'll deal with that when we come to it” attitude that many appear to have with regard to pension provisions may be playing a role here. For example, one third of the respondents, in their reply to the question whether they would adjust their saving behaviour if the existing pension scheme were retrenched, display such an attitude (Graph 16).

Graph 15

Opinions about possible measures towards ensuring the affordability of the public pension scheme

Percentage of respondents

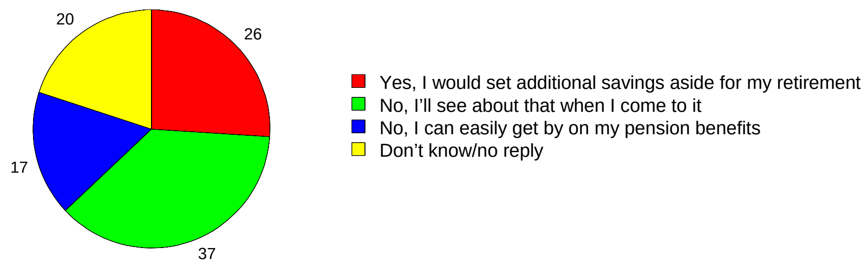


Source: DNB Household Survey, June 2003.

Graph 16

Would you adjust your saving behaviour if pension schemes were retrenched?

Percentage of respondents



Source: DNB Household Survey, June 2003.

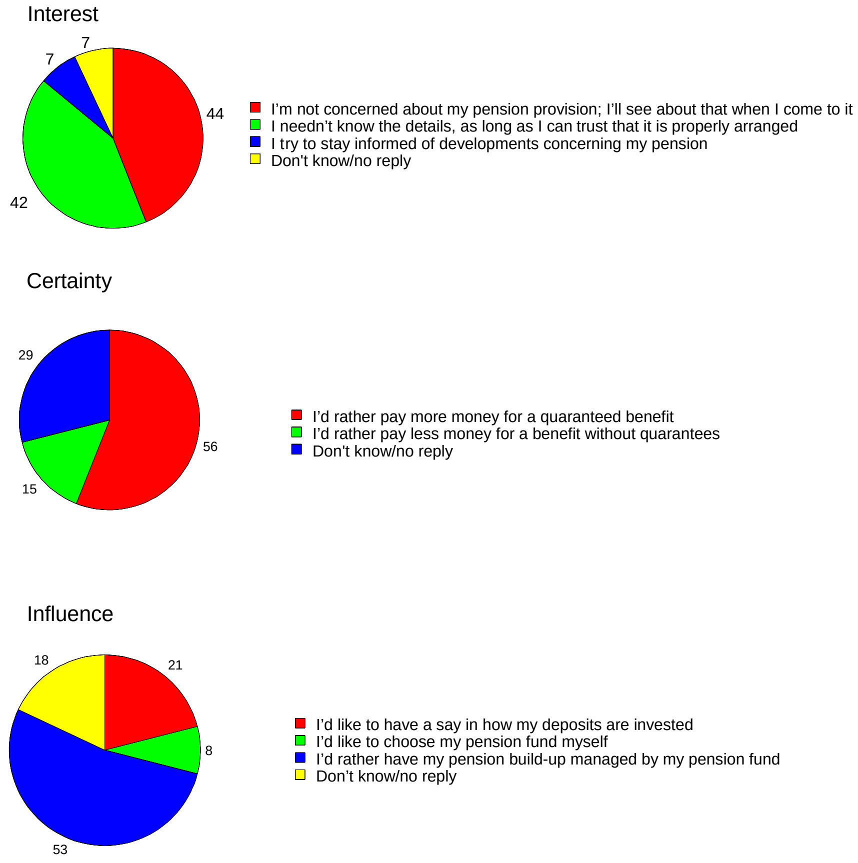
5.5 Preferences

The persons interviewed were requested to comment on three terms reflecting attitudes towards second pillar pension arrangements and thus to provide an insight into their preferences in this context (Graph 17). From this it emerged that the majority prefer having their pension build-up managed by their pension fund. Approximately 85% of the respondents would rather not go too deeply into the details of their pension schemes, half of this group taking the “we’ll deal with that when we come to it” position. Nevertheless, three out of 10 respondents would prefer having more freedom of choice than they have now. One in 10 would like to be free to choose a pension fund that manages their pension contributions. Besides, two in 10 feel the need to exert influence on the way in which their deposits are being managed, making the eventual benefits dependent on their own decisions. The latter very much resembles a defined contribution pension scheme, where only the deposit is fixed, without there being any guarantee of the amount of the eventual benefit. Most pension contracts in the Netherlands, however, are defined benefit-based, with the pension funds making conditional or unconditional pledges as to the height of the pension benefit of the average pay or final pay pension system. The majority of the population feels rather comfortable with the defined benefit system. The survey results show that people would rather pay a higher contribution for a guaranteed pension than a lower contribution in exchange for greater uncertainty about the eventual benefit. Note that households were not asked how much extra contribution they would like to pay in this respect.

Graph 17

Attitude towards own pension arrangements

Percentage of respondents



Source: DNB Household Survey, June 2003.

6. Saving behaviour

This section focuses on two recent developments that have an effect on people's saving behaviour. Firstly, it deals with the expectations regarding deflation, which may lead to a postponement of expenditures. Secondly, it goes on to discuss the consequences of the recent retrenchment of company saving schemes and the related additional unfreezing of earlier saving deposits.

6.1 Deflation

Besides the developments in the mortgage market and on stock markets, the financial behaviour of households depends on specific macroeconomic developments and risks. In a period of stock market declines, overcapacity at companies and historically low yields, one realistic risk being feared by financial market participants is deflation. Also among policymakers, deflation is a frequent discussion item (IMF (2003)). In this context, a drop in asset prices (asset price deflation) should be distinguished from general price deflation (goods deflation). The first form of deflation was seen in the stock markets in 2000-02. Deflation in the commodities sector is taking place in Japan, where consumer prices have been falling in recent years. Both forms of deflation affect the household sector. Goods deflation, notably the expectation thereof, is decisive for consumers' spending and saving behaviour. Indeed, deflation is attended by uncertainty about economic prospects, inducing households to step up saving. Also the expectation that consumer durables will become cheaper in time is a saving incentive. Asset price deflation undermines the financial positions of households, as it diminishes the value of their assets, such as their equities and homes. As a consequence, the value of collateral will drop, making creditors more cautious. While assets are decreasing in value, the level of the outstanding debt will remain the same, causing the balance sheet position of the household sector to deteriorate. If asset price deflation is followed by goods deflation, the debt level will even rise in real terms, while the same will hold for interest charges.

Against this background, Dutch households have been polled for their deflation expectations. According to the outcome of the DNB Household Survey, the chance of the general price index falling in the next two years is estimated to be relatively low (16% on a scale of 0-100%). One third of households rule out the possibility of deflation altogether. This explains why, according to the survey, their spending and saving behaviour is hardly affected by deflation expectations. Of the households proceeding from a more than 50% chance of deflation, only 5% are in effect adapting their spending and saving pattern. Households appear to factor in deflation more in their borrowing behaviour (12% of the overall population). Three quarters of these households own their own home. This category is probably more aware of the consequences of borrowing and more anticipatory in their financial planning than are tenants. With regard to borrowing, households take deflation risks into account by not taking out new loans. According to the survey, deflation risks barely prevent households from taking out interest-only loans. At times of deflation, such loans are disadvantageous as due to the declining price level the outstanding nominal debt will mount in real terms in the course of time. However, interest-only mortgages are very popular with Dutch households (according to the survey results, 41% of outstanding mortgages are interest-only); this carries a measure of susceptibility to deflation risks.

6.2 Company saving schemes

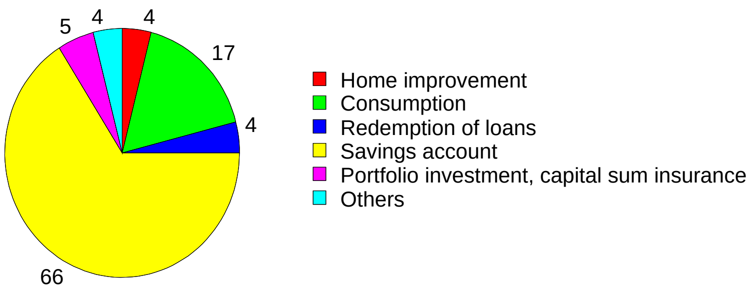
As of 1 January of 2003, the premium savings scheme was abolished and the salary savings plan was made less generous. By way of compensation, part of the savings - which initially were to be blocked on the savings account for four years - were prematurely released. One third of the respondents - representative of the Dutch population aged 16 and older - participate in one or more salary savings schemes and/or premium savings schemes. This amounts to roughly 4 1/2 million participants. For over 60% of this group, at least part of the savings were released. The others had not participated long enough, had withdrawn their money at an earlier date for specific expenditures, or just did not know whether their savings have been unfrozen. The average amount released is estimated at more than EUR 1,500. On a macro level, this amounts to roughly EUR 4-5 billion. It turns out that two thirds of this amount was transferred to other savings accounts (Graph 18). Only one fifth thereof went into expenditures, such as daily errands and (durable) consumer goods, or home improvement. The amount spent towards these ends comes to between 0.3 and 0.4% of the annual total of consumer expenditure and investment in owner-occupied homes (approximately EUR 250 billion). Hence, the macroeconomic effects are limited.

Nevertheless, about half of the company saving scheme participants indicate they intend to set less money aside, adducing as the main reason that saving has now become less attractive. Other reasons are that savings are more easily spent if they are not frozen and that the loss of income entailed by the retrenchment measures reduces the opportunities for saving. From last year's household survey it emerged that many people set aside the money released from their company savings schemes for the benefit of their own pension plans. From the present survey it appears that no fewer than one third of all respondents automatically transferred their deposits to annuity or single premium policies. Of this group, 70% indicate they intend to continue channelling deposits to the said policies, while a quarter have ceased making deposits or have agreed lower amounts.

Graph 18

Expenditure of released company saving scheme balances

Percentage of total amount released



Source: DNB Household Survey, June 2003.

7. Conclusions

From an international perspective, the debt of Dutch households is extraordinarily high. And it is still rising. Due in part to the low interest rate, second mortgage loans are raised or existing mortgage loans are renegotiated on a massive scale for the purpose of withdrawing equity. Unlike in the United States, the Dutch seldom do so to repay more expensive (consumer) loans. Often, funds from equity withdrawal are invested in owner-occupied homes. While this implies that the rising debt is attended by a higher property value, it also increases dependence on asset price fluctuations. Besides, homeowners more and more opt for a variable interest rate, rendering themselves increasingly susceptible to rate changes. At the moment, 15% of the mortgages outstanding have variable rates. Although the fixed interest periods of five and 10 years (accounting for 23% and 32%, respectively) are still the most popular, potentially, over a quarter of mortgages may be hit by an interest rate increase before the end of 2004.

On a macroeconomic level, mortgage equity withdrawal represents large amounts. While equity withdrawals appear to be on the rise again, their volume is evidently still below the level recorded for 1999 and 2000. At EUR 5 billion, released equity-related spending is estimated to make up half of that realised in 2000. Account being taken of carry-over effects, this will on balance reduce economic growth by more than 0.25 percentage points this year. Incidentally, it turns out that tenants, even those indicating they would like to buy a home, barely adjust their saving behaviour in response to the price movements on the housing market.

The stock market decline did not prompt investors to divest their stock holdings in droves, but it has made them more cautious. The vast majority have hardly bought additional equities, while a proportion of the investors indicate they have stepped up saving. Only 20% have substantially reduced their portfolios in recent years. Most did not do so until the crisis had lasted several years. While the proceeds from these transfers largely went into expenditures, this did not produce any major macroeconomic effects as the group concerned is relatively small.

The Dutch public are well aware of the discussions about the sustainability of the present pension system. Two thirds expect the public pension scheme to be retrenched within 10 years from now, ie in that it will become payable at a later age and/or represent less purchasing power. However, the public find it hard to reconcile themselves to measures that infringe on what they regard as acquired rights. They would rather pay higher contributions until the age of 65 and start enjoying today's level of benefits from then onwards. Although it remains to be seen if this will still hold once it is clear how much more will need to be contributed in such a situation, it is typical of the general attitude towards pension provisions. The public like to have the build-up of their pension rights managed by pension funds and would accept having to pay higher contributions in exchange for guaranteed benefits. Incidentally, a substantial minority (circa 30% of respondents) advocates a greater freedom of choice. Many, however, are as yet not concerned about their pension rights, adopting a "we'll see about that when we come to it" attitude. This is evidenced, among other things, by the lack of insight into the individual pension situation. It should not be ruled out, though, that this is not only caused by a lack of interest among the interviewed, but also by the information supplied regarding this subject. Here, a role may be reserved for pension funds and national authorities. It is in everybody's interest that people have a realistic image of what they may be expecting to receive on turning 65.

Bibilography

Alessie, P J A and F A Kapteyn (2002): "Huizenprijzen en besparingen", in R J M Alessie, P J A van Elsen and L H Hoogduin (eds), *De rol van vermogen in de economie*, Preadviezen 2002, Koninklijke Vereniging voor de Staathuishoudkunde, 31-56.

De Nederlandsche Bank (2000): "Survey among Dutch mortgage-holders on the use of mortgage credit", *Quarterly Bulletin*, June, 31-43.

——— (2002): "Spotlight on household wealth management in the Netherlands", *Quarterly Bulletin*, June, 27-36.

Federal Reserve (2002): "Mortgage refinancing in 2001 and early 2002", *Federal Reserve Bulletin*, December.

Haliassos, M and C Hassapis (2002): "Equity culture and household behaviour", *Oxford Economic Papers*, 54, 719-45.

International Monetary Fund (2003): *Deflation: determinants, risks, and policy options, findings of an interdepartmental task force*.

Van Rooij, M C J (2002): "Financial vulnerability of mortgagors and mortgagees", *Research Memorandum WO*, no 706.

The impact of financial variables on firms' real decisions: evidence from Spanish firm-level data

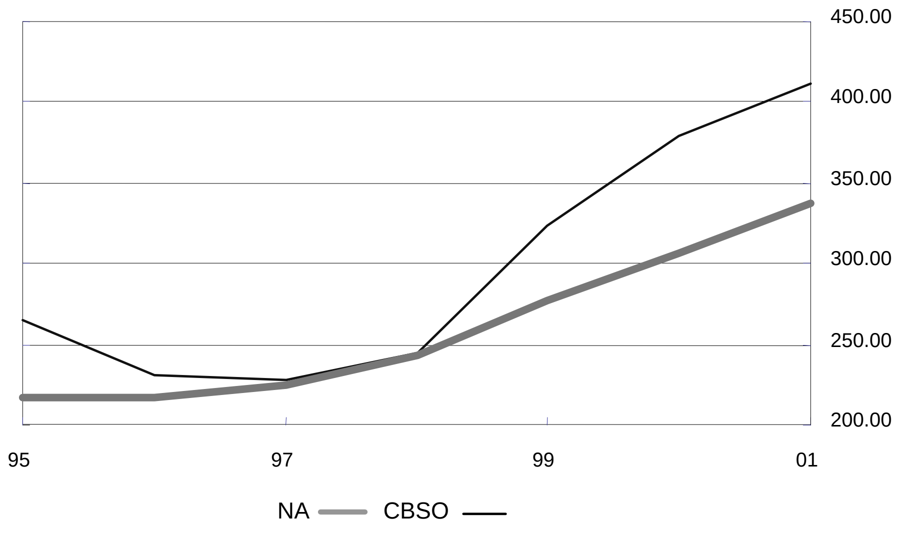
Ignacio Hernando and Carmen Martínez-Carrascal[1]
Bank of Spain

1. Introduction

The financial position of the corporate sector may influence the performance of the real economy and the stability of the financial system through its contribution to aggregate demand and its links to the banking system and capital markets. This paper analyses some measures of firms' financial health and assesses their impact on some real decisions of firms, bearing in mind that basing the assessment of the financial position of companies on an analysis of aggregate sectoral indices may, while being informative, occasionally cover up some vulnerabilities that only a study at a greater level of detail can reveal. In this sense, the implications for the financial strength of the Spanish corporate sector of the increasing debt ratios observed at an aggregate level (Graph 1) may differ depending on the distribution of indebtedness across firms. Therefore, in this paper the emphasis is placed on the analysis of disaggregated data on such financial indicators. For this purpose, itemised data from a sample of the non-financial firms reporting to the Bank of Spain Central Balance Sheet Data Office Annual Database for the period 1985-2001 are used.

Graph 1

Debt over gross revenue



Note: NA: National Accounts; CBSO: Central Balance Sheet Data Office.

Adjustment by companies to changes in the financial pressure they face (for instance, as a result of a monetary policy shift) can potentially involve a wide range of activities, with the most prominent

---

[1] This paper represents a follow-up of previous joint work with Andrew Benito, to whom we are very grateful. We also thank Juan Ayuso, Roberto Blanco, Jorge Martínez-Pagés, Fernando Restoy, Javier Vallés and participants in the BIS meeting and in the seminar held at the Bank of Spain for helpful comments, and the Central Balance Sheet Data Office of the Bank of Spain for providing the data. The views expressed are those of the authors and should not be attributed to the Bank of Spain.

relating to their investment decisions, human resources policies and financial policies. Benito and Hernando (2002) examine, making use of microeconometric methods (panel techniques), the sensitivity of a number of aspects of corporate behaviour (namely, investment in physical capital, employment, inventories and dividend policies) to changes in financial pressure. In this paper, using a similar methodological framework, we conduct a more in-depth study of the response of fixed investment and employment to a relatively broad set of indicators that are usually considered to characterise the financial position of firms. Among these, we include variables providing information on corporate profitability, financial burden and indebtedness (or leverage).

Additionally, we evaluate whether the impact of the financial position on business decisions is non-linear. In particular, our conjecture is that this relationship becomes relatively more intense when financial pressure exceeds a certain threshold. Furthermore, we analyse whether the relevant threshold differs depending on the real decision considered.

Finally, in the light of the estimated impacts of the different financial variables on firms' real decisions, we construct a composite indicator of financial pressure as a weighted average of these variables. Again, we investigate to what extent the weights attached to the different financial proxies differ for employment and for fixed investment.

The remainder of the paper is organised as follows. Section 2 provides a preliminary look at the descriptive information of the cross-sectional distribution of financial variables offering an overall assessment of financial pressure experienced by the Spanish corporate sector over the 1985-2001 period. Section 3 describes the baseline specifications for fixed investment and employment, summarises the estimation method and presents the basic estimation results. Section 4 analyses whether the impact of the financial position on corporate decisions becomes relatively more intense when financial tightness exceeds a certain threshold, whilst Section 5 constructs composite indicators of financial pressure, in the light of the estimated impacts of the different financial variables on firms' real decisions. Section 6 concludes.

2. The financial position of the Spanish corporate sector: a preliminary look at firm-level data

The financial performance and financing decisions of firms as well as their responses to financial pressure are important to both a country's macroeconomic conditions and the stability of its financial system. Thus, for instance, excessive indebtedness may adversely affect investment spending or, in the face of an unexpected shock, prompt sharp portfolio switching. However, from the standpoint of identifying the risks to macroeconomic and financial stability, it should be borne in mind that the fragility of certain firms need not be offset by the soundness of others. Accordingly, the use of aggregate indicators to assess the financial position of the corporate sector and its impact on real activity may be inadequate and thus a study at a greater level of detail may be required. Indeed, the behaviour of the companies that are most exposed financially is, for these purposes, as relevant (if not more so) as the average behaviour of the sector. Against this background, the purpose of this section is twofold. First, we attempt to provide an overall picture of the financial position of Spanish non-financial companies and its evolution over the period considered. Second, we try to assess to what extent the real behaviour - more precisely, the demand for factors of production - of the more financially vulnerable firms differs from that of firms with an average financial position.

The data employed are derived from an annual survey of non-financial firms conducted by the Central Balance Sheet Data Office of the Bank of Spain (Bank of Spain (2002)). This is a large-scale survey used extensively by the Bank in forming its assessment of the Spanish corporate sector. In terms of gross value added, the survey respondents jointly represent around 35% of the total gross value added of the non-financial corporate sector in Spain, and the pattern of evolution of the aggregate values for the main variables used here (employment, investment) is quite similar to that observed in the whole economy. This paper employs data for the period 1985-2001, for which the coverage of the survey has been relatively stable. Data are only used when there are at least five consecutive time-series observations per company. This produces an unbalanced sample of 7,547 non-financial companies and 70,625 observations with between five and 17 annual observations per company (see Data appendix).

Table 1 presents median values for the different variables used in our analysis for subsample periods.[2] The most important aggregate variation observed in (pro)cyclical variables such as fixed investment and cash flow reflects the recession in Spain, the trough of which was experienced in 1993. Also clear from Table 1 is the declining debt service burden apparent in the late 1990s. A median value for the interest debt burden term *idb*, of 0.216 and 0.214 for 1989-92 and 1993-96, respectively, compares to a figure of 0.100 for 1997-2001. This reduction primarily reflects reductions in nominal interest rates and the entry of Spain into the European monetary union.[3]

Table 1

Sample medians

		1985-88	**1989-92**	**1993-96**	**1997-2001**	**1985-2001**
I/K	Investment rate	0.118	0.103	0.076	0.111	0.100
N	Employment	65	47	35	37	43
Y	Real sales (1995 prices)	7,580.6	5,525.9	4,213.9	4,357.3	5,088.8
Δy	Sales growth	0.038	–0.007	0.013	0.041	0.021
Δw	Wage growth	0.012	0.022	0.004	0.005	0.010
B/A	Debt	0.301	0.247	0.269	0.249	0.263
$(B – m)/A$	Net indebtedness	0.207	0.164	0.173	0.140	0.168
B/GR	Debt over gross revenue	1.500	1.424	1.645	1.489	1.514
idb	Interest debt burden	0.188	0.216	0.214	0.100	0.167
tdb	Total debt burden	1.052	1.037	1.013	0.714	0.944
GR/A	Gross revenue	0.216	0.188	0.162	0.168	0.179
CF/A	Cash flow	0.130	0.105	0.095	0.115	0.110
pd	Probability of default	0.007	0.009	0.012	0.007	0.009
Observations		12,444	18,294	19,448	20,439	70,625

Note: See Data appendix for definitions.

This section presents, in primarily graphical form, preliminary data analysis of the sample of Spanish non-financial firms. This analysis first illustrates variation in the cross-sectional distributions of financial and real variables and how these distributions have varied over time. Then, a comparison is made of the behaviour of investment and labour demand for various sets of firms defined in terms of their financial position, using alternative indicators to proxy the degree of financial tightness of the companies.

First, we consider a narrow definition of the debt service burden that is defined as the ratio of interest payments on debt to the company's gross revenue (interest debt burden). The cross-sectional distribution of this variable and how it varies over time is shown in Graph 2.1. Different percentiles (ie the 25th, 50th, 75th and 90th) in the cross-sectional distribution in each year are displayed. The experience of the median company (the 50th percentile) is indicative of the typical Spanish company in each year, whilst the higher percentiles indicate the experience of those companies facing more intense financial pressure. Consider the median company (the 50th percentile) first. Its interest

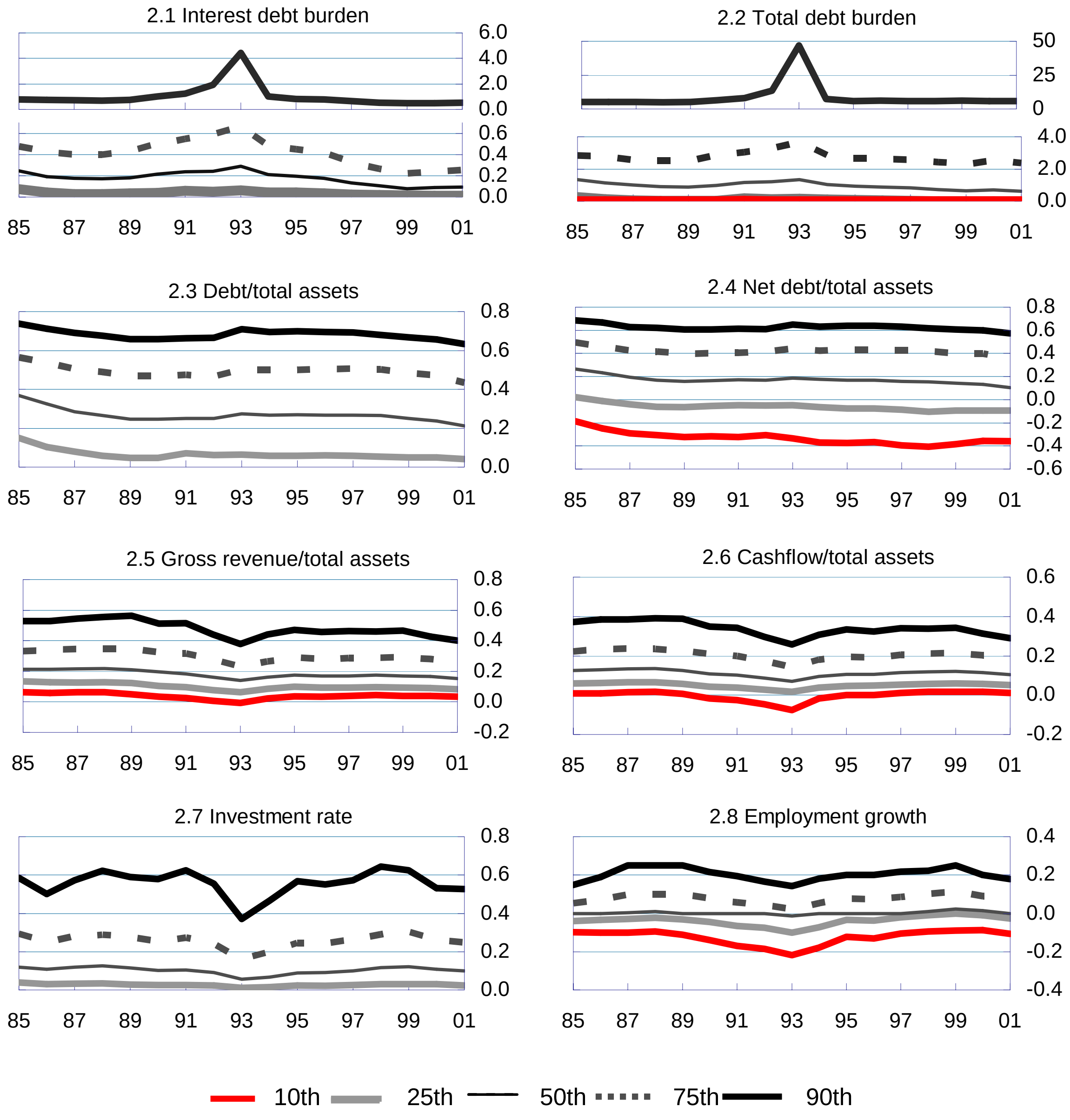
[2] See Data appendix for more precise definitions of the variables used in the paper.

[3] Nominal short-term interest rates in Spain were in the range of 12-16% (annual averages) in the period from 1985 to 1990, from which point they were reduced steadily to reach 4% by 2000, with Spain being one of the economies adopting the euro on 1 January 1999.

payments relative to gross operating profit fell during the mid-1980s but then began to increase at the end of the 1980s before again declining as growth resumed following the recession of the early 1990s. Variation in this ratio reflects a combination of variation in interest rates, company profitability and indebtedness. The variable peaked in 1993, from which point it has declined steadily. An important finding from Graph 2.1 is that as interest rates have fallen from the mid-1990s, the implied reduction in financial pressure has been felt throughout the cross-sectional distribution of firms in Spain and, indeed, is strongest for the more financially vulnerable. At the 75th percentile of the distribution, the interest debt burden fell from 0.66 in 1993 to 0.25 in 2001. This is a positive development for financial stability associated with the corporate sector in Spain. It also contrasts with the experience during the recession, at its deepest in 1993, when the financial pressure on the most vulnerable companies increased relative to the more typical companies suggesting that aggregate data on debt burdens at the time understated the vulnerability of the most fragile companies and hence of the system as a whole.

Graph 2

Percentiles of distributions of financial variables



A very similar pattern emerges when considering a broader measure of the debt service burden as a proportion of gross revenue that includes not only interest payments but also the stock of short-term debt. As can be seen from Graph 2.2, where the cross-sectional distribution of the total debt burden is displayed, the highest variation in this ratio is experienced by the most vulnerable companies, ie those in the upper decile of this distribution.

The cross-sectional distribution of corporate indebtedness, defined as the ratio of total outstanding debt to total assets, is illustrated in Graph 2.3. Similarly, Graph 2.4 depicts the cross-sectional distribution of net indebtedness. Both graphs show a remarkable stability in the cross-sectional distribution of indebtedness of firms. It should be noted that stability in the company-level cross-sectional distribution can be consistent with aggregate movements in a variable and in variation for individual companies. For instance, the aggregate data corresponding to those in Graph 2.4 indicate an increase in indebtedness from 32.4% in 1997 to 38.6% by 2001. This is explained by large firms increasing their debt levels. The stability of the cross-sectional distribution of indebtedness among Spanish firms also contrasts with findings for UK quoted firms, where a marked increase in dispersion in recent years has been found (Benito and Vlieghe (2000)).

Graphs 2.5 and 2.6 illustrate two measures of profitability: gross revenue and cash flow, in both cases divided by total assets. Two key observations arise from these graphs. First, profitability is clearly procyclical as we would expect. At the median, gross revenue (cash flow) over total assets declined from 21.9% (13.5%) in 1988 to 13.9% (7.1%) in 1993, from which point it has since recovered steadily, reaching 15.2% (10.3%) in 2001. Second, the experience of the median firm understates variation at the upper tail of the cross-sectional distribution, and in the case of the cash flow measure also at the lower tail. For financial stability issues it is the lower tail that is more relevant and here (ie at the 10th percentile) cash flow over total assets fell from 1.7% in 1988 to –7.5% in 1993.

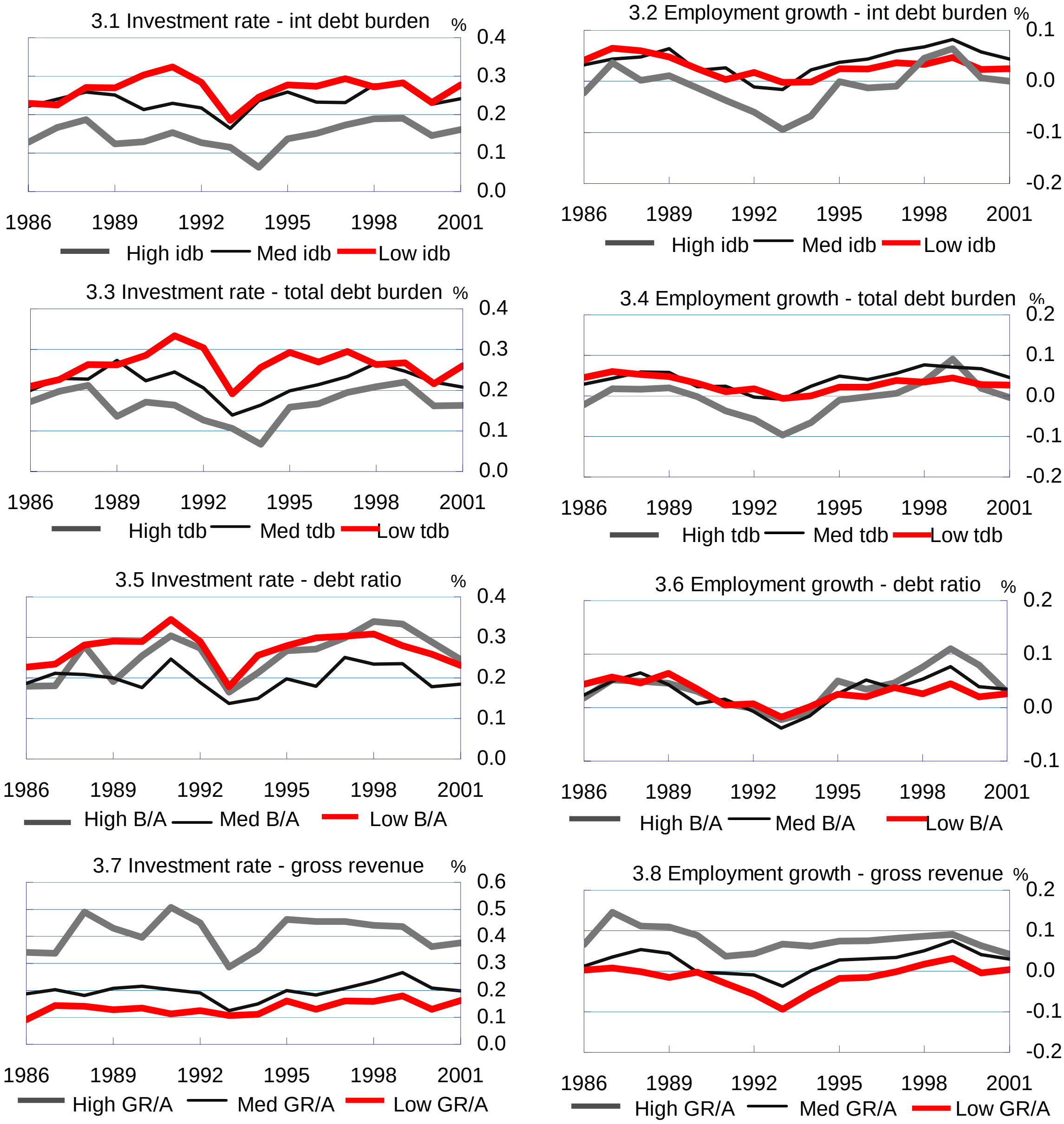
The cross-sectional distributions of fixed investment and employment growth are also considered, in Graphs 2.7 and 2.8, respectively. Investment is procyclical as expected. In particular, it declines in the recession of 1993 and especially so at the top of the cross-sectional distribution, namely at the 90th percentile. Employment growth at the median firm varies relatively little during the sample period, but becomes negative for the only time during the period in 1993. This disguises more significant variation at both the upper and lower tails of the distribution, which show even stronger declines in the recession of 1993 coinciding with increases in the financial pressure of borrowing costs, as shown above.

This descriptive analysis has shown that there is substantial cross-sectional variation in the distribution of Spanish firms for each of the variables examined. To the extent that real behaviour differs across companies facing different degrees of financial pressure, the assessment of the financial position of the corporate sector should ideally adopt a disaggregated perspective. To emphasise the relevance of this issue, in what follows we illustrate how investment in physical capital and labour demand differ across companies with different financial positions. For this purpose, Graph 3 compares the average level of both real variables in different corporate groupings defined on the basis of their financial position, proxied by alternative indicators. Each panel of the graph presents the average value of a real variable (the investment rate or the growth rate of employment) for the firms belonging to three different deciles of the distribution defined in terms of a financial indicator (the interest debt burden, the total debt burden, the debt ratio or gross revenue over total assets). The median decile (that including the firms between percentiles 45 and 55) can be regarded as representative of the behaviour of a firm in an average financial position. Analogously, the top (bottom) decile includes the 10% of firms with the highest (lowest) value of the corresponding financial indicator.

First, Graphs 3.1 and 3.2 compare the behaviour of firms facing different degrees of financial pressure, this being proxied by means of a measure of the relative burden of debt (or, in other words, of the firms' capacity to meet interest payments), ie our interest debt burden (*idb*) variable, which is defined as the ratio of interest payments to gross revenue. This variable, being the net result of changes in interest rates, in corporate profitability and in corporate debt, is a relevant indicator of the financial pressure firms may be facing. In Graphs 3.1 and 3.2, no marked differences in demand for factors of production are observed between the firms with lowest financial pressure and those with average financial pressure. However, firms with a higher financial burden in relation to their capacity to generate funds from operations have substantially lower investment and employment growth rates. Moreover, in the case of employment, this difference seems more marked in recessionary phases.

Graph 3

Financial position and level of activity



According to Graphs 3.3 and 3.4, similar conclusions can be drawn when the comparison is established in terms of our total debt burden variable (*tdb*). Thus, those companies facing a higher total financial burden display substantially lower investment and employment growth rates. Differences are less marked between the firms with the lowest total financial burden and those subject to average financial pressure, especially in the case of employment growth.

Interestingly, the pattern of results changes when the level of indebtedness is used as the indicator of financial tightness. Thus, in Graph 3.5, the observed relationship between the investment rate and the debt ratio is not monotonic. Similarly, no significant differences in employment growth are observed among the three deciles considered (Graph 3.6). This absence of a clear relationship between the debt level and the level of activity at the company level may be interpreted as the consequence of two opposite effects. On the one hand, firms with high indebtedness may experience difficulties in gaining access to additional credit to finance new investment projects, but on the other hand, those companies with higher levels of investment and employment growth are those that have been successful in attracting external funds to take advantage of their growth opportunities.

Finally, Graphs 3.7 and 3.8 show a clear link between the level of profitability and the demand for factors of production. Firms with higher levels of gross revenue over total assets have substantially higher investment and employment growth rates.

Overall, the evidence in this section suggests: first, that there is a substantial dispersion in the distribution of Spanish firms in terms of several indicators of the degree of financial tightness they face; second, that financial position affects business activity; and third, that this impact is not linear and becomes relatively more intense when financial pressure exceeds a certain threshold.

3. The impact of financial variables on firms' real decisions

The estimation analysis in this section considers the responsiveness of fixed investment and employment to changes in the financial conditions facing a company, proxied by a set of financial variables. These variables include indicators providing information on corporate profitability, indebtedness (or leverage) and relative burden of debt and try to capture the degree of financial pressure firms may be facing. More precisely, the financial variables considered are: two measures of the debt service burden, *tdb* and *idb*, two measures capturing the indebtedness of the company, (B/A) and ($(B-m)/A$), and two measures of corporate profitability, (GR/A) and (CF/A). Finally, we also consider an indicator of the probability of default that has been constructed using the estimated coefficients of a probit model for the probability of default estimated by Benito et al (2003) for a similar sample of Spanish non-financial firms.

3.1 Baseline specifications

The model estimated for fixed investment is an error correction model which specifies a target level for the capital stock and allows for flexible specification of short-run investment dynamics, in which we add different financial indicators as potential explanatory variables. The error correction model is standard in the investment literature. As is emphasised in Bond et al (1999), this type of model tends to produce more reasonable parameters than more structural models, such as Q models, which may be significantly affected by measurement error.[4] Assuming long-run constant returns to scale, subsuming the depreciation rate into the unobserved firm-specific effects and assuming that variation in the user cost of capital can be controlled for by including both time-specific and firm-specific effects, the following specification for the investment rate can be obtained:[5]

$$\left(\frac{I}{K}\right)_{it} = \alpha_i + \beta_1\left(\frac{I}{K}\right)_{it-1} + \beta_2\Delta y_{it} + \beta_3\Delta y_{it-1} + \beta_4(k-y)_{it-2} + X'_{it}\gamma + \theta_t + \varepsilon_{it} \tag{1}$$

where i indexes companies $i = 1... N$ and t indexes years $t = 1… T$. Δ denotes a first difference, I/K is the investment rate, y is the log of real sales, k is the log of real fixed capital stock, α_i are company-specific fixed effects, and X represents a vector of financial variables. θ_t are time effects that control for macroeconomic influences on fixed investment common across companies and ε_{it} is a serially uncorrelated, but possibly heteroskedastic error term. The coefficients β_2 and β_3 indicate the short-run responsiveness of fixed investment to sales growth, whilst the coefficient β_4 indicates the speed of adjustment of the capital stock towards its desired level.

The labour demand equation, derived by Nickell and Nicolitsas (1999) from a quadratic adjustment cost model which then adds financial factors, takes the following form:

$$n_{it} = \phi_i + \lambda_1 n_{it-1} + \lambda_2 k_{it} + \lambda_3 w_{it-1} + \lambda_4 \Delta w_{it} + \lambda_5 \xi_{it} + X'_{it}\eta + \Psi_t + \mu_{it} \tag{2}$$

where i indexes companies $i = 1,2… N$ and t indexes years $t = 1,2… T$. n is (log) average company employment during the year, w is the (log) average real wage at the company, k denotes (log) real

4 In any case, a Q model is not possible here, since most of the Spanish firms are not quoted and thus the usual Q variable cannot be constructed.

5 See Bond et al (1999) or Bond et al (2003) for details on the derivation of the investment model.

fixed capital stock. ξ is a demand shock proxy which consists of the growth in log real sales, and ψ_t represent a set of common time effects (year dummies) which will control for aggregate effects including aggregate demand.[6] μ_{it} is a serially uncorrelated but possibly heteroskedastic error term.

Two elements in equation (2) depart from what is considered a standard specification for labour demand. First, financial factors, represented by the regressors X_{it}, are included. Despite the extensive literature considering a potential role for financial conditions in shaping fixed investment (see Hubbard (1998)), there are few studies which allow for such a role in the context of labour demand models.[7] Second, the model includes a demand shock variable, ξ_{it}, following Bentolila and Saint-Paul (1992).

3.2 Estimation method

The estimation method consists of the GMM-System estimator proposed by Arellano and Bover (1995) and examined in detail by Blundell and Bond (1998). These models control for fixed effects with the estimator being an extension of the GMM estimator of Arellano and Bond (1991) estimating equations in levels as well as in first differences.

Apart from the bias that would arise if fixed effects were not controlled for, it is also necessary to note that most current firm-specific variables are endogenous. In order to avoid the bias associated with this endogeneity problem, we use a GMM estimator taking lags of the dependent and explanatory variables as instruments.

The use of a GMM-System estimator is justified because where there is persistence in the data such that the lagged levels of a variable are not highly correlated with the first difference, estimating the levels equations with a lagged difference term as an instrument offers significant gains, countering the bias associated with weak instruments (see Blundell and Bond (1998)). Several variables display high levels of serial correlation. The estimation method requires the absence of second-order serial correlation in the first-differenced residuals for which the test of Arellano and Bond (1991) is presented (labelled M_2). If the underlying model's residuals are indeed white noise then first-order serial correlation should be expected in the first-differenced residuals for which we also present the test of Arellano and Bond (1991), labelled M_1. We also report the results of the Sargan test for instrument validity in the GMM-System equations.

3.3 Basic results

Table 2 reports estimation results for fixed investment. Column 1 reports the results of the basic specification without financial variables. We generally find insignificant levels of persistence in company-level investment, a result quite consistent with results reported by Bond et al (2003). The error correction term $(k-y)_{it-2}$ is correctly signed and statistically significant with coefficient (robust standard error) of –0.175 (0.022) implying a reasonable speed of adjustment, comparable to that obtained in similar studies. The sales growth terms are positive and significant and their magnitude is in the upper range of the values usually obtained in the literature. We find the expected first-order serial correlation in our first-differenced residuals, while there is no evidence of second-order serial correlation, the key requirement for validity of our instrumentation strategy, and the Sargan test statistics are insignificant at conventional levels.[8]

We then consider adding the financial variables to the basic specification. Columns 2 to 8 of Table 2 report the estimates of the basic specification augmented with one financial variable at a time. First, columns 2 and 3 add debt variables to the standard specification. The expected negative coefficient is obtained although it is only at the margin of significance (p-value = 0.15) in the case of the B/A_{it-1} term.

---

6 The demand shock variable is not considered in the analysis of Nickell and Nicolitsas (1999), but it was included in a similar specification by Bentolila and Saint-Paul (1992).

7 Some exceptions are Nickell and Wadhwani (1991), Nickell and Nicolitsas (1999) and Ogawa (2003).

8 In our preferred estimates (those reported in the tables) we selected the instrument set such that the Sargan test statistic reported was not significant at conventional levels, although these estimates proved very similar to those where the instrument set included instruments dated $t-2$ to $t-6$ in the first-differenced equation and $t-1$ in the levels equation.

Table 2

Fixed investment

Basic specification augmented with one financial variable at a time

$(I/K)_{it}$	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
$(I/K)_{it-1}$	–0.057 (0.099)	–0.020 (0.083)	–0.084 (0.057)	–0.094 (0.085)	–0.055 (0.085)	–0.099 (0.090)	–0.113 (0.087)	–0.079 (0.076)
Δy_{it}	0.358 (0.124)	0.365 (0.111)	0.347 (0.109)	0.329 (0.095)	0.294 (0.098)	0.312 (0.111)	0.386 (0.113)	0.293 (0.101)
Δy_{it-1}	0.334 (0.112)	0.313 (0.106)	0.379 (0.088)	0.271 (0.086)	0.260 (0.086)	0.321 (0.103)	0.290 (0.104)	0.214 (0.057)
$(k-y)_{it-2}$	–0.175 (0.022)	–0.164 (0.020)	–0.171 (0.017)	–0.168 (0.020)	–0.162 (0.020)	–0.161 (0.020)	–0.158 (0.019)	–0.163 (0.018)
$(B/A)_{it-1}$		–0.070 (0.050)						
$((B-m)/A)_{it-1}$			–0.091 (0.027)					
idb_{it-1}				–0.024 (0.008)				
tdb_{it-1}					–0.004 (0.001)			
$(GR/A)_{it-1}$						0.201 (0.097)		
$(CF/A)_{it-1}$							0.331 (0.126)	
pd_{it-1}								–0.537 (0.204)
Year effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
M_1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
M_2	0.785	0.469	0.903	0.927	0.739	0.840	0.747	0.907
Sargan	0.188	0.170	0.402	0.091	0.374	0.142	0.156	0.145
Companies	7,547	7,547	7,547	7,547	7,547	7,547	7,547	7,547
Observations	55,531	55,531	55,531	55,531	55,531	55,531	55,531	55,531

Note: Estimation by GMM-System estimator using the robust one-step method (Arellano and Bond (1998). Blundell and Bond (1998)). Sargan is a Sargan test of over-identifying restrictions (p-value reported), with a chi-square distribution under the null of instrument validity. M_j is a test of jth-order serial correlation in the first-differenced residuals (p-values reported). These are both distributed as standard normals under the null hypotheses. Asymptotic robust standard errors reported in parentheses. Instruments: in the first-differences equation, the following lagged values of the regressors: Δy, B/A, GR/A, $CF/A(t-4,\ t-5)$, $(k-y)$ $(t-5,\ t-6)$ $(B-m)/A(t-2$ to $t-5)$, idb, tdb, $pd(t-3$ to $t-5)$; in the levels equations, the first differences of the regressors dated as follows: I/K, Δy, B/A, $(B-m)/A$, idb, $tdb(t-2)$, $pd(t-1)$, $(k-y)$, GR/A, $CF/A(t-3)$.

Table 3

Fixed investment

Simultaneously including several financial variables

$(I/K)_{it}$	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
$(I/K)_{it-1}$	–0.054 (0.075)	–0.101 (0.053)	–0.107 (0.079)	–0.022 (0.076)	–0.085 (0.053)	–0.076 (0.079)	–0.117 (0.057)	–0.133 (0.057)
Δy_{it}	0.316 (0.090)	0.329 (0.088)	0.302 (0.089)	0.277 (0.092)	0.296 (0.089)	0.270 (0.092)	0.292 (0.084)	0.322 (0.083)
Δy_{it-1}	0.257 (0.083)	0.336 (0.071)	0.278 (0.082)	0.245 (0.084)	0.338 (0.072)	0.285 (0.082)	0.364 (0.071)	0.357 (0.070)
$(k-y)_{it-2}$	–0.159 (0.019)	–0.167 (0.017)	–0.158 (0.018)	–0.154 (0.019)	–0.166 (0.017)	–0.155 (0.018)	–0.160 (0.017)	–0.161 (0.017)
$(B/A)_{it-1}$	–0.033 (0.048)			–0.020 (0.049)				
$((B-m)/A)_{it-1}$		–0.075 (0.027)			–0.071 (0.027)		–0.044 (0.031)	–0.046 (0.031)
idb_{it-1}	–0.024 (0.009)	–0.018 (0.009)	–0.017 (0.010)					–0.015 (0.009)
tdb_{it-1}				–0.004 (0.001)	–0.003 (0.001)	–0.003 (0.001)	–0002 (0.001)	
$(GR/A)_{it-1}$			0.162 (0.094)			0.151 (0.094)	0.153 (0.065)	0.155 (0.065)
$(CF/A)_{it-1}$								
Year effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
M_1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
M_2	0.706	0.671	0.754	0.449	0.907	0.962	0.465	0.312
Sargan	0.097	0.230	0.087	0.362	0.514	0.275	0.257	0.115
Companies	7,547	7,547	7,547	7,547	7,547	7,547	7,547	7,547
Observations	55,531	55,531	55,531	55,531	55,531	55,531	55,531	55,531

Note: See note to Table 2.

These estimates, in particular that including the net indebtedness term $((B-m)/A)_{it-1}$,[9] suggest that a high level of debt can lead to balance sheet adjustment in the form of companies deferring or forgoing investment projects (see also Vermeulen (2002) for an industry-level study). Second, in columns 4 and 5 two indicators of the relative debt service burden are included. For both variables (the interest debt burden term idb_{it-1} in column 4 and the total debt burden tdb_{it-1} in column 5) a significantly negative and well determined effect is found. This suggests that the financial pressure of debt servicing plays an important role in influencing investment levels of firms. Third, the estimates in columns 6 and 7 include two indicators of corporate profitability. In both cases, $(GR/A)_{it-1}$ in column 6 and $(CF/A)_{it-1}$ in column 7, the coefficients are significantly positive, which is consistent with studies of investment for other countries. As has been extensively discussed in the literature on investment and financial constraints, the cash flow terms might be either picking up the relevance of internal finance for investment or acting as a proxy for investment opportunities. Finally, the results reported in column 8 show that the indicator for the probability of default, pd_{it-1}, displays the expected negative and statistically significant effect on investment.[10] As this indicator is a composite measure based on several financial indicators, each of them weighted by its influence on the probability of default, its estimated coefficient in the investment equation reflects the impact of the financial situation on corporate investment through its incidence on the probability of default.

Nevertheless, the relative importance of different financial variables in explaining the probability of default or the probability of failure might differ from their relative contribution to explaining real decisions of companies. Thus, in order to get a more precise picture of the global impact of financial conditions on corporate behaviour, it is worth directly and simultaneously including several financial indicators in the estimated equations. Thus, it is possible to ascertain which specific financial features (indebtedness, profitability, financial burden ...) are more relevant for each specific corporate decision. However, the close links between the different financial indicators imply that few indicators are likely to turn out to be simultaneously significant. As a consequence, the interpretation of the results of this exercise is not a trivial task. Table 3 reports the estimates of specifications of the investment equation, simultaneously including several financial variables. As can be seen from the tables, those variables measuring the burden of servicing debt, both *tdb* and *idb*, remain significant in all specifications and their coefficients are quite robust. As regards the indicators of indebtedness, the gross measure (B/A) is never significant. In the case of the net debt term ($(B-m)/A$), it retains its significance in most cases. However, a notable decline in the point estimate of its coefficient is observed when a profitability indicator is included. Finally, the coefficients for the corporate profitability terms remain significant in all specifications although their point estimate is lower whenever the net debt term is included in the specification.[11]

Our first set of estimation results for the employment equation is presented in Table 4. Column 1 reports the results of the basic specification without financial variables whereas columns 2 to 8 report the results obtained when a financial variable is added to the specification. These results show the importance that financial factors have in explaining labour demand. The results in columns 2 and 3 show that debt has a negative (although non-significant) impact on labour demand. However, when considering the two indicators of the relative burden of debt, both of them have a negative and highly significant impact on labour demand. The results of the estimation when an indicator of profitability is included are reported in columns 6 and 7, and show a positive and significant impact of the profitability indicator on employment demand, for a 95% confidence level. Finally, as in the case of the investment equation, a negative and significant coefficient is found for the indicator of the probability of default, pd_{it-1}.

---

[9] By including this indicator we want to analyse whether debt is important once adjusted for liquidity. An indicator of liquidity (liquid assets divided by short-term debt) turned out to be insignificant when included in both the investment and the employment equations.

[10] The estimate for this variable should be viewed with some caution since the reported standard errors do not take into account that it is an estimated regressor.

[11] Table 3 reports results for specifications including the gross revenue term (GR/A). The pattern of results is qualitatively similar when the cash flow term (CF/A) is included instead of GR/A.

Table 4

Employment

Basic specification augmented with one financial variable at a time

	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
n_{it-1}	0.915 (0.020)	0.924 (0.015)	0.910 (0.017)	0.943 (0.016)	0.941 (0.017)	0.930 (0.019)	0.927 (0.019)	0.920 (0.018)
k_{it}	0.039 (0.008)	0.037 (0.007)	0.042 (0.007)	0.030 (0.007)	0.030 (0.007)	0.031 (0.007)	0.034 (0.007)	0.041 (0.008)
Δw_{it}	–0.535 (0.118)	–0.533 (0.109)	–0.522 (0.104)	–0.416 (0.097)	–0.507 (0.101)	–0.491 (0.096)	–0.501 (0.099)	–0.462 (0.111)
w_{it-1}	–0.017 (0.053)	–0.023 (0.044)	–0.002 (0.048)	–0.053 (0.042)	–0.037 (0.043)	–0.024 (0.046)	–0.012 (0.046)	0.016 (0.049)
Δy_{it}	0.303 (0.047)	0.305 (0.044)	0.301 (0.044)	0.300 (0.046)	0.299 (0.044)	0.285 (0.042)	0.306 (0.043)	0.311 (0.043)
$(B/A)_{it-1}$		–0.012 (0.021)						
$((B-m)/A)_{it-1}$			–0.010 (0.013)					
idb_{it-1}				–0.022 (0.007)				
tdb_{it-1}					–0.003 (0.001)			
$(GR/A)_{it-1}$						0.085 (0.031)		
$(CF/A)_{it-1}$							0.113 (0.041)	
pd_{it-1}								–0.450 (0.199)
Year effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
M_1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
M_2	0.082	0.088	0.075	0.117	0.068	0.059	0.088	0.089
Sargan	0.443	0.242	0.444	0.471	0.647	0.339	0.362	0.298
Companies	7,547	7,547	7,547	7,547	7,547	7,547	7,547	7,547
Observations	55,531	55,531	55,531	55,531	55,531	55,531	55,531	55,531

Note: Estimation by GMM-System estimator using the robust one-step method (Arellano and Bond (1998), Blundell and Bond (1998)). Sargan is a Sargan test of over-identifying restrictions (p-value reported), with a chi-square distribution under the null of instrument validity. M_j is a test of jth-order serial correlation in the first-differenced residuals (p-values reported). These are both distributed as standard normals under the null hypotheses. Asymptotic robust standard errors reported in parentheses. Instruments: in first-differences equation, following lagged values of the regressors: n, B/A, $(B-m)/A(t-5)$, a, Δy, $\Delta w(t-5, t-6)$, w, GR/A, $CF/A(t-4, t-5)$, idb, tdb, $pd(t-4$ to $t-6)$; in the levels equations,the first differences of the regressors dated as follows: n, w, B/A, $(B-m)/A(t-2)$, idb, tdb, pd, $CF/A(t-3)$, $GR/A(t-3)$.

Table 5 shows the results obtained when more than one financial variable is included in the estimation. As can be seen in columns 1 and 4 for debt and 2 and 5 for net indebtedness, both indicators are also non-significant when they are combined with another financial variable. In contrast, indicators of debt burden maintain their significance when they are included in the estimation with an indebtedness or profitability measure. The same applies to profitability indicators: they remain significant when they are combined with another indicator. Finally, columns 7 and 8 show that when three financial indicators are included in the regression (one for indebtedness, another for debt burden and the third one for profitability) the first is no longer significant, as was also the case when it was combined with only one additional indicator, whereas the indicators of debt burden remain significant at a 95% confidence level and the profitability terms are also significant although their point estimates are somewhat reduced.

4. Non-linear effects

The evidence presented in Section 2 shows that firms with a weaker financial situation - ie those firms belonging to the decile of the distribution characterised by the highest values of alternative proxies of financial pressure - have substantially lower investment and employment growth rates. However, in general, no significant differences in demand for factors of production are observed between the firms with least financial tightness and those with an average financial pressure. This evidence suggests that the relationship between the real activity of firms and their financial position is non-linear. Moreover, it seems reasonable that there will be a more pronounced impact of this position on real activity once the financial pressure reaches a certain threshold. In this section, we provide a more formal analysis of this hypothesis. For this purpose, we estimate the investment and labour demand equations described in Section 3, but now allowing for a differential impact of financial conditions depending on the relative level of the corresponding financial indicator. As in Tables 2 and 4 we estimate the investment and employment models considering one financial indicator at a time. In each regression, we test whether the companies facing high financial pressure - ie those firms in the upper decile (or quartile) of the distribution defined in terms of the corresponding financial indicator - are more sensitive to the financial conditions. More precisely, we estimate the following specifications:

$$\begin{aligned}\left(\frac{I}{K}\right)_{it} &= \alpha_i + \beta_1\left(\frac{I}{K}\right)_{it-1} + \beta_2\Delta y_{it} + \beta_3\Delta y_{it-1} + \beta_4(k-y)_{it-2} \\ &\quad + \gamma_1 F_{it} D^F_{0-75} + \gamma_2 F_{it} D^F_{75-90} + \gamma_3 F_{it} D^F_{90-100} + \theta_t + \varepsilon_{it}\end{aligned} \tag{3}$$

and

$$\begin{aligned}n_{it} &= \phi_i + \lambda_1 n_{it-1} + \lambda_2 K_{it} + \lambda_3 w_{it-1} + \lambda_4 \Delta w_{it} + \lambda_5 \xi_{it} \\ &\quad + \eta_1 F_{it} D^F_{0-75} + \eta_2 F_{it} D^F_{75-90} + \eta_3 F_{it} D^F_{90-100} + \psi_t + \mu_{it}\end{aligned} \tag{4}$$

where D^F_{0-75}, D^F_{75-90} and D^F_{90-100} are dummy variables for observations below the 75th percentile, between the 75th and 90th percentiles, and above the 90th percentile, respectively, of the distribution defined in terms of the financial variable *F*. When a corporate profitability measure - either (*GR*/*A*) or (*CF*/*A*) - is used as a financial indicator, we replace these dummies by D^F_{0-10}, D^F_{10-25} and $D^F_{25-100,}$ which are dummy variables for observations below the 10th percentile, between the 10th and 25th percentiles, and above the 25th percentile. In these cases, the lower the percentile, the lower the profitability, and the higher, a priori, the degree of financial tightness.

4.1 Results

Table 6 reports the results obtained for investment when non-linearities are considered. As can be seen, debt is not significant in either of the groups, although the comparison of the magnitude of the coefficients for the three groups shows that it goes in the expected direction (negative coefficient and higher, in absolute value, for those companies with higher indebtedness). When we consider net indebtedness instead of debt, we obtain evidence in favour of the existence of differences in the impact of this variable on investment, depending on its magnitude: net indebtedness is irrelevant for firms with moderate levels of net indebtedness (below the 75th percentile), whereas for those firms above this threshold it has a negative and significant impact both for the group between the 75th and 90th percentiles and for the group in the upper decile.

Table 5

Employment

Simultaneously including several financial variables

	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
n_{it-1}	0.944 (0.013)	0.936 (0.015)	0.950 (0.016)	0.940 (0.013)	0.930 (0.016)	0.944 (0.015)	0.951 (0.015)	0.950 (0.015)
k_t	0.029 (0.006)	0.032 (0.007)	0.026 (0.007)	0.030 (0.007)	0.035 (0.007)	0.025 (0.007)	0.022 (0.007)	0.025 (0.007)
Δw_{it}	–0.435 (0.090)	–0.433 (0.086)	–0.454 (0.082)	–0.503 (0.095)	–0.500 (0.082)	–0.512 (0.078)	–0.515 (0.074)	–0.453 (0.077)
w_{it-1}	–0.036 (0.038)	–0.033 (0.039)	–0.048 (0.039)	–0.022 (0.038)	–0.037 (0.044)	–0.030 (0.038)	–0.024 (0.037)	–0.038 (0.037)
Δy_{it}	0.292 (0.043)	0.291 (0.043)	0.307 (0.042)	0.288 (0.041)	0.271 (0.038)	0.285 (0.036)	0.283 (0.034)	0.295 (0.039)
$(B/A)_{it-1}$	0.016 (0.024)			0.016 (0.026)				0.005 (0.027)
$((B-m)/A)_{it-1}$		0.007 (0.014)			0.021 (0.015)		0.011 (0.017)	
idb_{it-1}	–0.023 (0.007)	–0.022 (0.008)	–0.017 (0.008)					–0.014 (0.008)
tdb_{it-1}				–0.003 (0.001)		–0.003 (0.001)	–0.003 (0.001)	
$(GR/A)_{it-1}$					0.097 (0.019)	0.079 (0.019)	0.091 (0.020)	
$(CF/A)_{it-1}$			0.084 (0.044)					0.114 (0.044)
Year effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
M_1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
M_2	0.086	0.084	0.113	0.054	0.042	0.048	0.044	0.084
Sargan	0.201	0.400	0.525	0.191	0.416	0.425	0.365	0.230
Companies	7,547	7,547	7,547	7,547	7,547	7,547	7,547	7,547
Observations	55,531	55,531	55,531	55,531	55,531	55,531	55,531	55,531

Note: See note to Table 4.

Table 6

Investment

Non-linear effects

	[1]	[2]	[3]	[4]	[5]	[6]	[7]
$(I/K)_{it-1}$	–0.089 (0.066)	–0.076 (0.069)	–0.102 (0.069)	–0.098 (0.074)	–0.114 (0.076)	–0.134 (0.075)	–0.141 (0.050)
Δy_{it}	0.293 (0.096)	0.346 (0.093)	0.356 (0.090)	0.261 (0.093)	0.344 (0.090)	0.362 (0.092)	0.284 (0.077)
Δy_{it-1}	0.245 (0.095)	0.354 (0.091)	0.339 (0.088)	0.335 (0.087)	0.356 (0.088)	0.290 (0.090)	0.361 (0.067)
$(k-y)_{it-2}$	–0.171 (0.018)	–0.166 (0.018)	–0.170 (0.018)	–0.170 (0.019)	–0.166 (0.019)	–0.159 (0.018)	–0.162 (0.016)
$(B/A)_{it-1}(<p75)$	0.072 (0.077)						
$(B/A)_{it-1}(>75; <p90)$	–0.013 (0.059)						
$(B/A)_{it-1}(>p90)$	–0.052 (0.054)						
$((B-m)/A)_{it-1}(<p75)$		–0.061 (0.047)					–0.052 (0.030)[1]
$((B-m)/A)_{it-1}(>p75; <p90)$		–0.147 (0.062)					–0.052 (0.030)[1]
$((B-m)/A)_{it-1}(>p90)$		–0.127 (0.048)					–0.052 (0.030)[1]
$(idb)_{it-1}(<p75)$			–0.081 (0.096)				
$(idb)_{it-1}(>p75; <p90)$			–0.100 (0.060)				
$(idb)_{it-1}(>p90)$			–0.031 (0.009)				
$(tdb)_{it-1}(<p75)$				–0.007 (0.008)			–0.004 (0.007)
$(tdb)_{it-1}(>p75; <p90)$				–0.005 (0.010)			0.011 (0.010)
$(tdb)_{it-1}(>p90)$				–0.004 (0.001)			–0.002 (0.001)
$(GR/A)_{it-1}(>p25)$					0.202 (0.101)		0.165 (0.063)[1]
$(GR/A)_{it-1}(>p10; <p25)$					0.662 (1.103)		0.165 (0.063)[1]
$(GR/A)_{it-1}(<p10)$					0.658 (0.727)		0.165 (0.063)[1]
$(CF/A)_{it-1}(>p25)$						0.311 (0.135)	
$(CF/A)_{it-1}(>p10; <p25)$						3.470 (2.770)	
$(CF/A)_{it-1}(<p10)$						0.890 (0.447)	
Year effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
M_1	0.000	0.000	0.000	0.000	0.000	0.000	0.000
M_2	0.978	0.988	0.756	0.812	0.643	0.486	0.201
Sargan	0.068	0.254	0.032	0.259	0.082	0.395	0.187

Note: See note to Table 2. Number of companies: 7,547. Number of observations: 55,531.

[1] Coefficients restricted to be equal.

When indicators of debt burden are considered, results strongly support the existence of non-linearities: both indicators are significant for firms above the 90th percentile, whereas for firms between the 75th and the 90th percentile total debt burden is found to be insignificant and interest debt burden is only at the margin of significance (p-value = 0.09). For firms below the 75th percentile, neither of these indicators has a significant impact on investment.

As for profitability indicators, a positive and significant coefficient is obtained for those firms with higher profitability (those above the 25th percentile). However, the coefficients for these variables are rather imprecisely estimated for the other two groupings. As expected, we obtain a higher coefficient for those companies in the lower tail of the distribution (a priori those facing higher financial pressure). However, this coefficient is only significant for (*CF*/*A*).

Ideally, we would like to allow for non-linearities in the effects of more than one financial variable at a time. However, when simultaneously including different financial variables in a non-linear fashion, there is a sharp drop of significance in the interaction terms. For this reason, we opted for a mixed strategy by including one financial variable in a non-linear way and the rest of the financial variables linearly. Using this approach, the results of our preferred specification are reported in the last column of Table 6. In this specification, a linear effect is allowed for gross revenue over total assets and for net debt, while total debt burden enters in a non-linear way. We find, as expected, a positive coefficient for (*GR*/*A*) and a negative one for net debt.[12] Finally, a negative impact of total debt burden is only found for firms that are in the upper tail of the distribution.

Results for employment are shown in Table 7, and corroborate the existence of a non-linear impact of financial variables on firms' real decisions. We find, however, some differences with respect to investment: both indicators of indebtedness and debt burden are significant for firms in the upper decile of the distribution, for a 99% confidence level, but for firms between the 75th and 90th percentile only indicators of interest debt burden have a significant impact on employment. When profitability is considered, lower and upper bounds are found to be significant and, as was also the case for investment, the coefficient estimated for the lower decile is higher than that estimated for firms with higher profitability (above the 25th percentile). As in the case of investment, we also adopted a mixed strategy in the specification of the financial variables in the employment equation. The results of our preferred specification are reported in the last column of Table 7. In this specification, a linear effect is allowed for gross revenue over total assets while total debt burden enters in a non-linear way. A positive and significant coefficient is found for the profitability term and a negative and significant one for total debt burden only for firms that are above the 90th percentile.

Overall, these results corroborate the descriptive evidence in Section 2 and point to the existence of threshold effects on the impact of financial variables on investment and employment.[13] The specific threshold and the different sensitivities to the financial position seem to depend on the particular financial variable considered.

5. Composite indicators of financial pressure

In Section 3, we obtained evidence in favour of the existence of a significant impact of financial variables on the demand for factors of production. The results in Section 4 suggest that this impact is more pronounced for the upper tail of the distributions defined in terms of the proxies for financial pressure. Now, in this section, we wish to construct synthetic indicators that summarise the non-linear influence that financing conditions have on investment and employment. Moreover, on the basis of these composite indicators we wish to assess how the impact of financial conditions has evolved over time with a special emphasis on the distribution across companies of this impact. For this purpose, we compute linear combinations of alternative sets of financial variables, where the relative weights are given by the estimated coefficients in the investment and the employment equations.

---

[12] Profitability and net debt enter linearly in the specification, although in the table we present the coefficient for each of the three groups (which is equal for all of them) separately.

[13] Although the results clearly support this conclusion, it has to be mentioned that the results reported in this section are more sensitive to the set of instruments used than those obtained for the linear specifications presented in the previous section.

Table 7

Employment

Non-linear effects

	[1]	[2]	[3]	[4]	[5]	[6]	[7]
n_{t-1}	0.922 (0.014)	0.905 (0.015)	0.934 (0.014)	0.931 (0.014)	0.926 (0.016)	0.940 (0.026)	0.958 (0.013)
k_{it}	0.035 (0.007)	0.041 (0.007)	0.033 (0.007)	0.034 (0.007)	0.032 (0.007)	0.029 (0.007)	0.019 (0.006)
Δw_{it}	–0.510 (0.092)	–0.492 (0.092)	–0.635 (0.087)	–0.550 (0.087)	–0.637 (0.080)	–0.519 (0.077)	–0.554 (0.075)
w_{i-1}	0.000 (0.039)	0.043 (0.041)	–0.021 (0.037)	–0.003 (0.035)	0.012 (0.040)	–0.026 (0.039)	–0.050 (0.032)
Δy_{it}	0.297 (0.038)	0.293 (0.042)	0.313 (0.041)	0.286 (0.038)	0.280 (0.038)	0.280 (0.037)	0.280 (0.035)
$(B/A)_{it-1}(<p75)$	0.032 (0.034)						
$(B/A)_{it-1}(>75; <p90)$	0.015 (0.029)						
$(B/A)_{it-1}(>p90)$	–0.042 (0.023)						
$((B-m)/A)_{it-1}(<p75)$		–0.001 (0.015)					
$((B-m)/A)_{it-1}(>p75; <p90)$		0.001 (0.025)					
$((B-m)/A)_{it-1}(>p90))$		–0.052 (0.023)					
$(idb)_{it-1}(<p75)$			–0.039 (0.051)				
$(idb)_{it-1}(>p75; <p90)$			–0.109 (0.033)				
$(idb)_{it-1}(>p90)$			–0.034 (0.005)				
$(tdb)_{it-1}(<p75)$				0.006 (0.004)			0.005 (0.005)
$(tdb)_{it-1}(>p75; <p90)$				–0.001 (0.004)			0.006 (0.005)
$(tdb)_{it-1}(>p90)$				–0.004 (0.001)			–0.003 (0.001)
$(GR/A)_{it-1}(>p25)$					0.090 (0.012)		0.085 (0.019)[1]
$(GR/A)_{it-1}(>p10; <p25)$					0.116 (0.060)		0.085 (0.019)[1]
$(GR/A)_{it-1}(<p10)$					0.304 (0.090)		0.085 (0.019)[1]
$(CF/A)_{it-1}(>p25)$						0.067 (0.044)	
$(CF/A)_{it-1}(>p10; <p25)$						–1.350 (1.134)	
$(CF/A)_{it-1}(<p10)$						0.549 (0.185)	
Year effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
M_1	0.000	0.000	0.000	0.000	0.000	0.000	0.000
M_2	0.067	0.052	0.093	0.039	0.045	0.049	0.048
Sargan	0.151	0.334	0.107	0.155	0.567	0.564	0.440

Note: See note to Table 4. Number of companies: 7,547. Number of observations: 55,531.

[1] Coefficients restricted to be equal.

Thus, a financial conditions indicator for investment (FCII) can be defined as follows:

$$FCII_{it} = -\sum_k \hat{\gamma}^k X_{it}^k \tag{5}$$

where $\hat{\gamma}^k$ is the estimated coefficient for financial variable X^k in the investment equation. Analogously, a financial conditions indicator for employment takes the following form:

$$FCIE_{it} = -\sum_k \hat{\eta}^k X_{it}^k \tag{6}$$

where $\hat{\eta}^k$ is the estimated coefficient for financial variable X^k in the employment equation. These indicators measure the contributions of the financial variables in the investment and employment equations. As the sign of these contributions is changed, the higher the indicator, the tighter the financial conditions faced by companies, ie the larger the negative impact of financial conditions on investment or employment. Since we have allowed for a non-linear impact of financial variables, the differences in the indicator across firms will reflect not only differences in the financial position but also differences in the sensitivity of the real variables to this position.

Our starting point is to construct financial conditions indicators for investment and employment on the basis of the estimated coefficients of our preferred models in the previous section. In particular, our benchmark models are those in column 7 of Table 6 for fixed investment and column 7 of Table 7 for employment. Both models allow for a non-linear effect of the total debt burden tdb_{it-1}, while restricting the impact of the gross revenue term (GR/A) to be linear. In addition, the investment model also includes a linear net debt term ($(B - m)/A$).

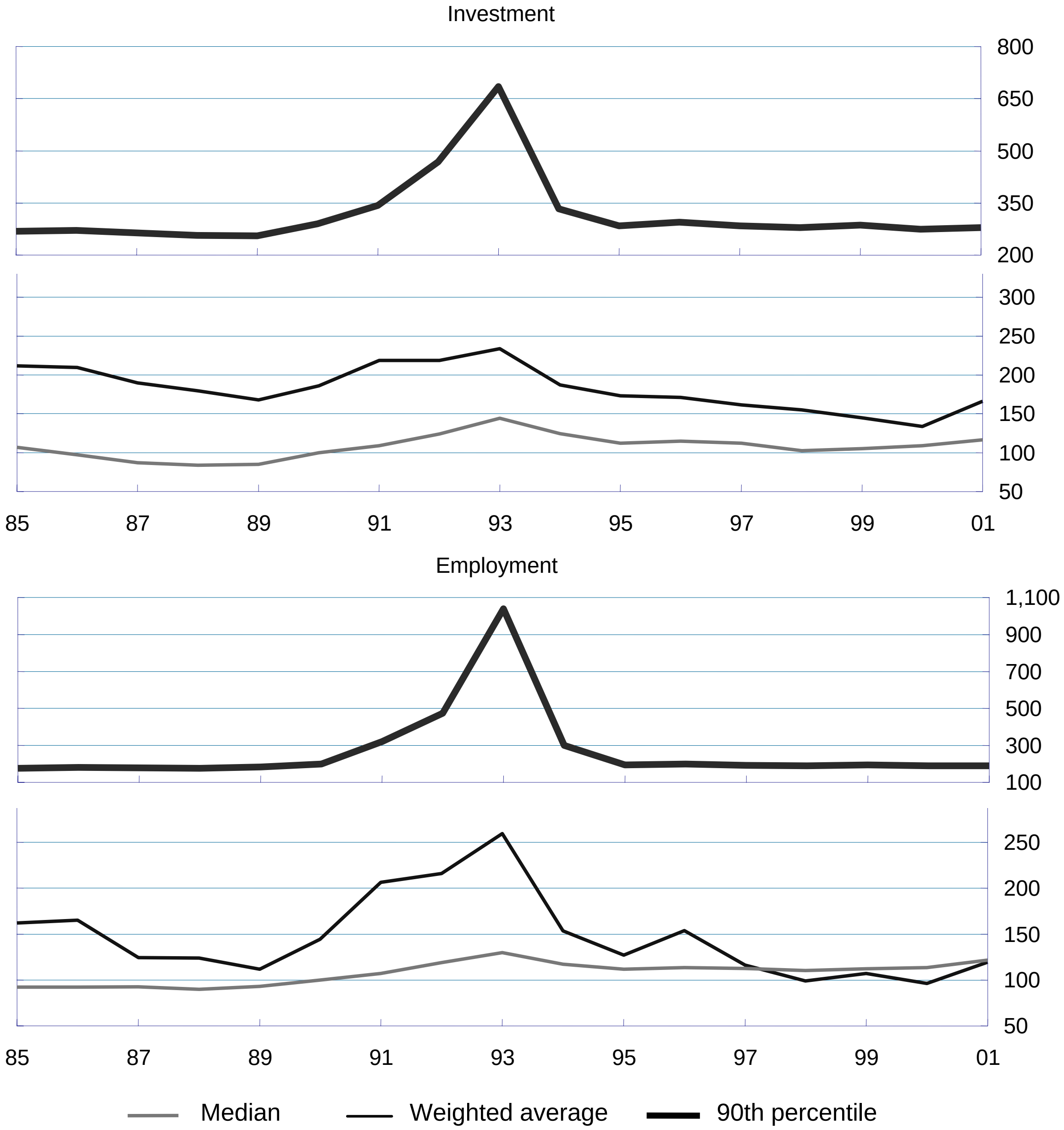
In order to ascertain the relevance of financial variables for companies in different financial positions, it is useful to focus on different percentiles of the distribution of these indicators. More precisely, we present the evolution of the median value of these indicators as representative of the average financial pressure faced by the companies in our sample. We also show the evolution of the 90th percentile, to assess the time profile of the vulnerability of the companies facing high financial pressure. Finally, we report the weighted average as an aggregate indicator of the position of the corporate sector as a whole. The weight for each firm in this indicator will be given by its contribution to total (aggregate) fixed assets, in the case of investment, or to total employment, in the case of employment. To compare the different percentiles and the weighted average of the financial indicators we normalise them by setting $FCII_{1990}^{median} = 100$ and $FCIE_{1990}^{median} = 100$.

Graph 4 displays the different percentiles and the weighted average of the indicators for the impact of financing conditions on investment and employment. In the case of fixed investment (Graph 4, upper panel), the different percentiles and the weighted average display a similar countercyclical pattern. According to the median FCII, the second half of the 1980s was characterised by a relaxation of financial conditions which was mostly explained by the reduction in corporate debt in a period of high nominal interest rates and, to a lesser extent, by a certain recovery in corporate profitability. In the early 1990s, this indicator shows a tightening of financial conditions as a result of an intense deterioration of corporate profitability.[14] After reaching a peak in 1993, the median FCII declined continuously until 1998, owing to the reduction in the level of debt. In this period, there is also a modest improvement in corporate profitability. Finally, the median FCII displays a slight increase in the last three years of the sample owing to a slight reduction in corporate profits.

[14] Interestingly, if we consider FCIIs derived from models excluding measures of profitability (for instance, the models in columns 2 and 5 of Table 3), the tightening of financial conditions during the cyclical downturn of the early 1990s is less severe.

Graph 4

Composite indicators of the impact of financial conditions



The comparison of the median and the weighted average FCII shows that the weighted average presents higher values for the entire period, implying that the financial position for those firms that are more relevant for investment is weaker than that of the median. Furthermore, in some periods a different evolution pattern is observed for the representative (median) firm and the weighted average. For instance, the significant tightening in financial conditions observed in 2001 for the weighted average is not so clearly seen in the median, which displays a more stable evolution in the last part of the sample.

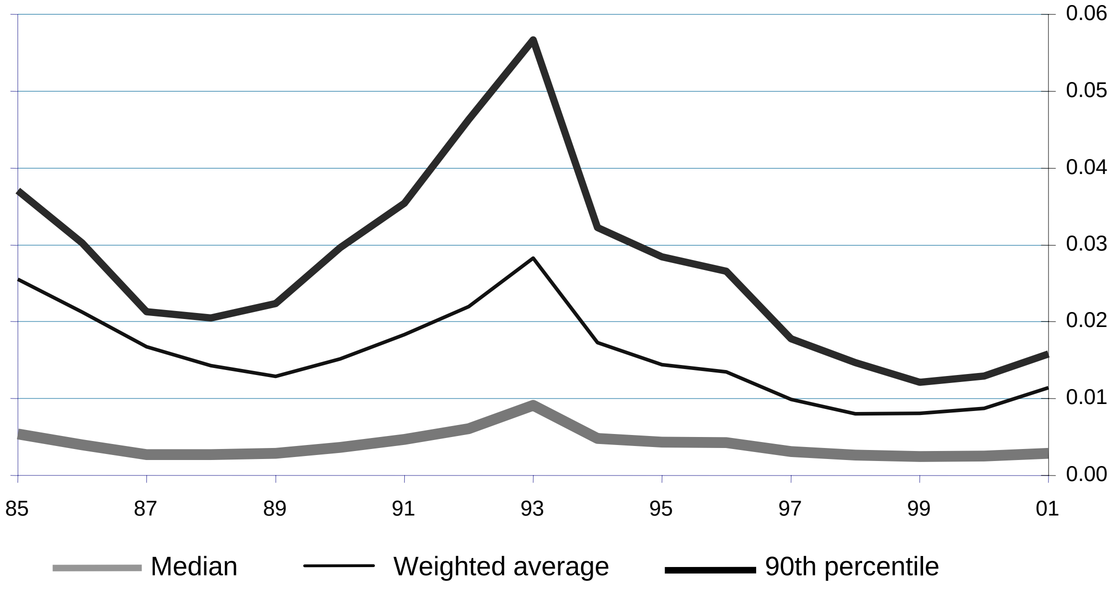
Again, the comparison of the median FCII with the higher percentiles reveals that it is in the recessions, especially in the cyclical trough of 1993, when the impact of the financing conditions on investment increased relatively more for the most vulnerable companies than for companies with an average financial pressure.[15] It is also worth noting that the observed increase in the median in the last years of the sample is not observed for the firms in the upper decile of the distribution.

In the case of employment (Graph 4, lower panel), our preferred financial indicator is a weighted average of the *total debt burden* and the *gross revenue term*. As previously mentioned, a non-linear effect is allowed for the total debt burden term and the debt variables are no longer significant once additional financial indicators are included in the equation. The profile of the different percentiles of the FCIE is quite similar to that of the FCII. First, the different percentiles display a countercyclical pattern and second, this pattern is more evident in the case of the highest percentile. Again, the median indicator is not a good proxy of the position of the sector as a whole, although in this case the difference between the median and weighted average indicator diminishes in the last part of the sample. In fact, the median exceeds the weighted average in the last part of the sample period (after 1998), something that is not seen in the FCII. The tightening in financial conditions observed in 2001 for the weighted average FCII is also seen in the FCIE.

Finally, for the sake of comparison, we show in Graph 5 the indicator of financial fragility based on the model of Benito et al (2003) for the probability of default. As in the case of our indicators of the impact of financial conditions, we display the median, the 90th percentile and the weighted average. In this case, the weights are given by total assets of the firm with respect to the aggregate level of assets. The cyclical profile of the different percentiles of the distribution of this indicator is quite similar to those reported in Graph 4. The weighted average value of the indicator has a range of 0.008 to 0.028 while the 90th percentile varies between 0.012 and 0.057. There is a slight difference regarding the timing of the most recent deterioration in financial conditions. This financial stability indicator dates it to 1998, while according to our indicators it is only in 2001 that we observe a tightening in financing conditions.

Graph 5

Financial fragility indicator



Note: The indicator of financial fragility is an indicator of the probability of default, based on Benito et al (2003). See the Data appendix for a brief description of this indicator.

[15] As expected, the value of the 90th percentile of the indicator based on a non-linear specification is higher, over the whole sample period, than the value of the 90th percentile of an indicator constructed with the same variables but without considering non-linearities. And, interestingly, it is in the recession when this difference is larger. For the weighted average indicator, a linear specification also yields a degree of fragility persistently lower than that reported here, including non-linearities.

Overall, this evidence shows the relevance of using firm-level data when analysing the evolution of the financial position and suggests that financing conditions do not affect all companies equally. A tightening of financial conditions will have a significantly greater effect on the real decisions of those firms with lower financial soundness. This is particularly relevant in episodes where the financial pressure on a significant number of firms breaches the threshold at which it has a more intense influence on business activity. In these episodes, indicators based on aggregate data may not reliably reflect the system's financial soundness since they do not adequately reflect the deterioration of the financial position of the more vulnerable companies.

6. Conclusions

This paper has aimed to assess the impact of financial conditions on firms' real decisions, using a large-scale company-level panel data set for the period 1985-2001. The analysis has focused on the behaviour of fixed investment and employment, which are conceivably two of the most important aspects of adjustment by firms in response to changes in financial conditions. Within the general topic of the relationship between financial conditions and real activity, we have addressed three specific issues: first, the assessment of the relative importance of different financial variables in explaining the real decisions of firms; second, the analysis of the non-linearity in the relationship between financial proxies and real variables; and, finally, the construction of a synthetic indicator to capture the impact of financing conditions on investment (and, alternatively, on employment).

Our results strongly indicate that financial position is important in explaining corporate decisions on fixed investment and employment. Several financial indicators turn out to be significant in the estimated equations. In particular, measures of the debt service burden (both including and excluding the stock of short-term debt) remain significant when additional financial indicators are incorporated and their coefficients are quite robust. As regards the indicators of corporate profitability, they are significant in all specifications, although their point estimates depend on the additional financial variables included in the specification. Finally, the evidence for the indicators of indebtedness is less conclusive. In the investment equation the net debt term is significant in most cases. In the employment equation, the debt terms are never significant in the linear specifications but they are significant for the upper decile of the distribution when considering non-linear specifications.

We have found evidence in favour of the hypothesis of a non-linear relationship between financial conditions and real activity. At a purely descriptive level, we have shown that the group of firms facing a higher degree of financial pressure, which we identify as those in the upper decile of the cross-sectional distribution of firms defined in terms of alternative financial indicators, have substantially lower investment and employment growth rates. The regression analysis corroborates this result: the sensitivity of investment and employment to financial conditions is substantially larger for those firms in the upper quartile (or decile) of the distribution defined in terms of the corresponding financial indicator. Moreover, in some specifications, the financial variable is not significant for the companies facing a moderate (or low) degree of financial tightness. Overall, this evidence suggests that the real impact of financial conditions is non-linear and becomes relatively more intense when financial pressure exceeds a certain threshold. As a consequence, from the standpoint of identifying the risks to macroeconomic and financial stability, the use of firm-level data seems to be particularly relevant in episodes where the financial pressure on a significant number of firms reaches levels at which it has a more pronounced influence on real activity. In these episodes, indicators based on aggregate data may not reliably reflect the system's financial soundness since they do not adequately reflect the vulnerability of the most fragile companies. In addition, the analysis of our composite indicators constructed at the firm level reveals that neither the level nor the evolution of the financial pressure experienced by the representative (median) firm is a good measure of the financial pressure faced by the corporate sector. In fact, in the last year of our sample (2001) the observed increase in our median indicators is much lower than that observed for the weighted average.

As regards the most recent data, our composite indicators for the impact of financial conditions on investment and employment remain at moderate levels, in historical terms. At an aggregate level, Spanish firms have shown an increase in debt ratios, although this has not been translated into a higher debt service burden due to the declining path of interest rates. Thus, the financial position of the corporate sector will not foreseeably represent, on average, a significant obstacle to the recovery in investment and employment. Moreover, a more disaggregated analysis shows that, in the most recent period, the increase in debt ratios for those firms in a weaker financial position (which are, according to our results, the most sensitive to changes in their financial position) has been lower than that observed in the aggregate. Furthermore, the available information for 2003 reveals that the companies with the highest indebtedness have indeed experienced reductions in their debt ratios. Nonetheless, the high level of debt at some of these firms suggests that their scope to obtain additional external funds is now lower and that their exposure to potential shocks is higher. Additionally, our analysis has shown that financial conditions for those firms that are more relevant for investment and, to a lesser extent, for employment, are tighter than those for the median (representative) firm and therefore these companies may be more influenced by disturbances affecting their financial position.

Data appendix

Table A1

Number of time-series observations per company

Panel structure

No of records	5	6	7	8	9	10	11
Companies	1,268	1,109	913	658	581	379	352
No of records	12	13	14	15	16	17	Total
Companies	411	365	415	400	234	462	7,547

Investment (*I*)

Purchase of new fixed assets.

Capital stock (*K*)

Fixed assets at replacement cost (calculated by the Central Balance Sheet Data Office (CBSO) of the Bank of Spain). When introduced in real terms, *K* is deflated by the Gross Fixed Capital Formation deflator.

Total assets (*A*)

This is given by the sum of fixed assets at replacement cost *K* and working capital less provisions.

Employment (*N*)

The number of employees during the year.

Real sales (*S*)

Total company sales, deflated by the GDP deflator.

Wages (*W*)

The average company wage is given by direct employment costs (not including social security contributions) divided by the employment headcount and deflated by the GDP deflator.

Gross revenue over total assets (*GR/A*)

Gross operating profit plus financial revenue divided by total assets.

Debt (*B/A*)

Total outstanding debt divided by total assets.

Debt over gross revenue (*B/GR*)

Total outstanding debt divided by gross revenue, *GR*.

Net debt ((*B – m)/A*)

Total outstanding debt less cash and its equivalents divided by total assets.

Interest debt burden (*idb*)

Interest payments divided by gross revenue.

Total debt burden (*tdb*)

Interest payments plus short-term debt over gross revenue.

Cash flow (*CF/A*)

Post-tax profit plus depreciation of fixed assets divided by total assets.

Probability of default (*pd*)

Based on Benito et al (2003), this indicator is obtained from the estimation of a probit model which has as explanatory variables real sales, debt, interest debt burden, short-term debt without cost over total debt, profitability, liquidity, a dummy indicating if the firm pays dividends and the growth rate of gross domestic product.

For interest debt burden and total debt burden, where companies have a negative or zero value for the denominator and a positive value for the numerator the ratio is set equal to the value of the 99th percentile that year; where the numerator is zero, the ratio is set equal to zero, for any value of the denominator. Additionally, for all the variables used as regressors (except those that enter in levels), when the value is higher than the 99th percentile, it is changed for the value of this percentile.

References

Arellano, M and S Bond (1991): "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations", *Review of Economic Studies*, 58, pp 277-97.

——— (1998): *Dynamic panel data estimation using DPD98 for GAUSS: a guide for users*, Institute for Fiscal Studies, mimeo.

Arellano, M and O Bover (1995): "Another look at the instrumental-variable estimation of error-components models", *Journal of Econometrics*, 68, pp 29-52.

Bank of Spain (2002): *Results of non-financial corporations. Annual Report 2001*, Central Balance Sheet Data Office, Bank of Spain.

Benito, A, J Delgado and J Martínez-Pagés (2003): *A synthetic indicator of Spanish non-financial firms' fragility*, Bank of Spain, mimeo.

Benito, A and I Hernando (2002): "Extricate: financial pressure and firm behaviour in Spain", Banco de España Working Papers, no 0227.

Benito, A and G W Vlieghe (2000): "Stylised facts of UK corporate health: evidence from micro-data", *Bank of England Financial Stability Review*, June, pp 83-93.

Bentolila, S and G Saint-Paul (1992): "The macroeconomic impact of flexible labor contracts with an application to Spain", *European Economic Review*, 36, pp 1013-53.

Blundell, R W and S Bond (1998): "Initial conditions and moment restrictions in dynamic panel data models", *Journal of Econometrics*, 87, pp 115-43.

Bond, S, J Elston, J Mairesse and B Mulkay (2003): "Financial factors and investment in Belgium, France, Germany and the United Kingdom: a comparison using company panel data", *The Review of Economics and Statistics*, 85(1), pp 153-65.

Bond, S, D Harhoff and J Van Reenen (1999): "Investment, R&D and financial constraints in Britain and Germany", Institute for Fiscal Studies, *working paper*, 99/5.

Hubbard, R G (1998): "Capital market imperfections and investment", *Journal of Economic Literature*, 36, pp 193-225.

Nickell, S and D Nicolitsas (1999): "How does financial pressure affect firms?", *European Economic Review*, 43, pp 1435-56.

Nickell, S and S Wadhwani (1991): "Employment determination in British industry: investigations using micro-data", *Review of Economic Studies*, 58, pp 955-69.

Ogawa, K (2003): "Financial distress and employment: the Japanese case in the 90s", NBER *Working Papers*, no 9646.

Vermeulen, P (2002): "Business fixed investment: evidence of a financial accelerator in Europe", *Oxford Bulletin of Economics and Statistics*, 64, pp 213-31.

Financial constraints and real activity: a non-structural approach using UK survey data[1]

Ulf von Kalckreuth[2]
Deutsche Bundesbank

1. Introduction and summary

Understanding the causes and effects of financial constraints for firms is of key importance for a variety of policy issues. In monetary transmission theory, the credit channel is supposed to condition and amplify the "neoclassical" relative price effects of interest rate changes on firm activity. Monetary policy may affect the ability of banks to finance firms (bank lending channel), or else influence firms' ability to attract external finance by affecting the value of their equity (balance sheet channel). Second, financial constraints on real activities form one crucial link that determines the real consequences of financial imbalances of various types: banking crises, asset price bubbles, or government debt. Ultimately, financial constraints due to asymmetric information are especially important for those future-oriented activities that deal with generating new knowledge: research, development and the introduction of innovative products and processes. These activities are fundamental to the long-run performance of any economic system.

For all these reasons, the study of firms' financial constraints at a micro level is a major topic on the agenda of central bank research. A recent coordinated research effort by the European System of Central Banks (ESCB) on the basis of large national balance sheet databases shows that financial constraints do seem to matter for firm investment and the monetary transmission process (see Chatelain et al (2003a) for an overview). However, unlike much of the literature on US firms, size does not seem to be a good indicator of informational asymmetries and the assorted financial constraints in European countries. Among some of the larger euro area countries - France, Germany, Italy and Spain - only Italian small firms show an excess sensitivity of investment with respect to cash flow.[3]

It is conceivable that the importance of financial constraints for the real activity of firms also depends on the financial system. Allen and Gale (2001) argue that intermediaries and markets may have different comparative advantages. A market-based system deals better with situations where innovations occur and where there is a fundamental diversity of opinion, whereas intermediaries are able to save transaction costs when a large amount of experience has been gained and things are no longer changing. The empirical patterns of financial constraints and their importance for monetary policy, financial stability and innovation and growth may therefore depend on economic institutions.

This paper is part of a larger research effort based on large panels of survey data which aims to compare the significance of financial constraints for firm behaviour in (bank-based) Germany and the (capital market based) United Kingdom. With respect to the United Kingdom, we are able to explore

---

1 This paper was written while the author was at the Bank of England. Encouragement and support from Charles Bean, Peter Brierley, Heinz Herrmann and Garry Young were pivotal. The CBI gave access to its rich micro database, and I would like to thank, in particular, Ian McCafferty, Jonathan Wood and Jamie Morrison for their crucial help. Ongoing discussions with many people were productive. Thanks are therefore due to Nick Bloom, Steve Bond, Harald Stahl, Christian Upper, Geoffrey Wood, Garry Young and Mike Young. Especially, I would like to thank Emma Murphy. She is co-author of a companion paper and allowed me to draw on our joint work. Ultimately, I am grateful for comments on presentations at the Bank of England in London, at the BIS in Basel, at the Deutsche Bundesbank in Frankfurt and at the CES-ifo in Munich.

2 The views expressed in this paper do not necessarily reflect those of the Deutsche Bundesbank. All errors, omissions and conclusions remain the sole responsibility of the author.

3 The key results have been collected in Angeloni et al (eds; 2003): see Chatelain and Tiomo (2003) on France, von Kalckreuth (2003b) on Germany, Gaiotti and Generale (2003) on Italy, as well as Chatelain et al (2003b) for a comparative study of the euro area. On Germany, see also the study by Breitung et al (2003).

the database for the CBI *Industrial Trends Survey* (ITS), which is the most important survey for business cycle analysis in the United Kingdom. For the 11 years between January 1989 and October 1999, our cleaned unbalanced panel contains 49,244 quarterly observations on 5,196 firms. According to the CBI, the ITS represents around 33% of the total current employment within UK manufacturing.

Apart from its size and coverage, the data set has two important characteristics. First, it contains many small firms, on which very little information is available from micro data sets based on quoted companies. More than 63% of the ITS observations refer to firms with less than 200 employees. Second, the data-set contains detailed information on financial constraints that firms face in their investment decisions. Notably, a number of firms (around 20.8% of respondents) explicitly state two things: that they are constrained by the lack of either internal or external financial resources, and that these constraints have an influence on their investment behaviour.

This is exactly what the bulk of the empirical literature on financial constraints, following the seminal article by Fazzari et al (1988), tries to prove. The standard procedure in this literature is to split the sample by some criterion that a priori identifies firms as being financially constrained or unconstrained, such as size, dividend behaviour or the risk of default, and then to test whether the observed differences in investment behaviour between the two types of firm are consistent with what is to be expected from a better or worse financial standing in a situation of asymmetric information.[4] Armed with the CBI data, this complicated and very indirect procedure, heavily criticised on theoretical grounds by Kaplan and Zingales (1997, 2000), seems to be unnecessary: a subset of respondents explicitly claims to be constrained. However, it needs to be examined whether they have told the truth, ie whether or not there is informational content in their assertions. If this is the case, we have the chance to take a closer look at the interrelationship between financial constraints and investment demand.

Section 2 is dedicated to the presentation of our data set. The econometric part of the paper, Section 3, examines the informational content of our data on financial constraints. Our focus is on capacity adjustment, as the ITS data on capacity gaps, planned expansion and rates of capacity utilisation are especially rich. First, we look at the association between two types of constraints: capacity restrictions and financial constraints. Then we undertake a *duration analysis* with respect to spells of capacity constraints. Firms report whether their capacity is insufficient with respect to demand. Those firms which indicate financial constraints should take longer to close a capacity gap if there is informational content in their answers - either because they are less able to finance their investments or else because they have bigger gaps to fill. In fact, financially constrained firms do take longer to end a period of insufficient capacity. The paper ends with a conclusion in Section 4.

2. The data set

The CBI *Industrial Trends Survey* (ITS) is a qualitative survey that looks at short- and medium-term trends in the UK manufacturing and processing industries. The survey is a postal questionnaire aimed at a senior level within firms and is usually completed by either the Chairman or the Chief Executive. The CBI produces both a monthly and a quarterly survey, the latter providing more in-depth analysis. It covers a wide range of subject areas including optimism regarding the general and export business situation, investment, capacity, order books, numbers employed, output, deliveries, stocks, prices, constraints to output, export orders, competitiveness regarding the domestic, EU and non-EU market, innovation and training. The quarterly survey is the empirical basis for our analysis. Mitchell et al (2002a,b) have used the ITS micro data to show that disaggregate survey-based indicators they developed can outperform traditional aggregate indicators. The full text of the questionnaire can be found in Wood (2001).

According to the CBI, the ITS represents around 33% of the total current employment within UK manufacturing. Our research focuses on 11 years of data between January 1989 and October 1999.

---

4 See, for example, Chirinko and von Kalckreuth (2002).

The cleaned, unbalanced panel contains 49,244 quarterly observations on 5,169 firms. We exclude any divisions of a company, as their information might not be truly relevant to questions regarding size or financial constraints. Furthermore, we exclude all anonymous responses because these companies cannot be tracked over time.

Apart from its size and coverage, the data set has a number of important characteristics. First, the survey consists of four employment size groups, the largest of which looks at small firms with less than 199 employees. As can be seen in Table 1, 63% of the ITS observations refer to these small firms. Second, the ITS has a wide-ranging base of firms from the UK manufacturing and processing industries; Table 2 shows the breakdown of two digit SIC codes by observation.

The question on constraints on investment is of key importance for our study. We therefore quote the exact wording here for the sake of convenience:

Question 16c

What factors are likely to limit (wholly or partly) your capital expenditure authorisation over the next 12 months?

(If you tick more than one factor, please rank in order of importance)

☐ inadequate net return on proposed investment

☐ shortage of internal finance

☐ inability to raise external finance

☐ cost of finance

☐ uncertainty about demand

☐ shortage of labour, including managerial and technical staff

☐ other

☐ n/a

Table 3 shows both the overall frequency with which firms cite a given constraint (any rank) to investment expenditure and the frequency with which this constraint was given the first rank. Firms had the opportunity to name more than one constraint on capital expenditure, but they were asked to rank the importance of their constraints. We interpret the answers to this question as information on marginal investment. For the entire sample, uncertainty about demand is the most common impediment mentioned by all firms. It is cited by most firms (55% of respondents) as a significant constraint over the time period we studied. An interpretation of these figures in the light of theory, however, has to take into account the possibility that many firms focus only on “downside risks”, such as an unanticipated decrease in demand, rather than on uncertainty in the sense of imprecise expectations. For a recent review of the microeconometric literature on investment and uncertainty see von Kalckreuth (2003a). The second most important constraint is inadequate net return, cited by 39% of firms as an important constraint. Other constraints seem to have been less important. Costs of finance were cited frequently in the early 1990s, but have been mentioned significantly less often since then.

Turning to financial issues, we see that 4.3% of firms cite inability to raise external finance as a factor likely to limit their capital expenditure over the next 12 months. However, it is also interesting to note that only 1.96% mentioned this particular factor as their foremost constraint. 18.9% of firms cite “shortage of internal finance” as an impediment to investment, and for 13.6% of firms it is the most important barrier.

For inferential purposes, it is important to know whether there is sizeable individual variation in the financing constraints data. Table 4 conditions on whether in the preceding period a firm reported either a shortage of internal finance or an inability to raise external finance, and it shows the transition to the next period. It is easy to see that the reports on financial constraints are strongly autocorrelated. Among the firms that do not report financial constraints in a given period, a share of 87.6% will not report any in the next period either, and 12.4% switch to reporting constraints. But only 36.7% of the firms that report financial constraints in one period will state that they are unconstrained next time; the remaining two thirds will claim to be still constrained. However, the state of financial constraints is far from being determined by the state in the preceding period - there is a great deal of individual movement in both directions.

3. Is there informational content in the financial constraints data?

As highlighted in the previous section, a sizeable proportion of firms in the CBI *Industrial Trends Survey* state that their investment is constrained either by insufficient internal funds or by the inability to raise external finance. These statements are interesting and potentially very rich: as we shall see below, they permit identification of the financial regime of a firm. Weighted averages of survey questions are often used for forecasting and evaluation purposes at a sectoral or macro level and in many cases turn out to be surprisingly accurate.[5] However, it is not clear a priori how well the survey responses reflect the individual economic situation of the answering firm. Therefore, we need to check the informational content of the statements on financial constraints at a micro level. In other words, we want to see whether the statements on financial constraints relate to other information in the data set in a way that is consistent with theory.

3.1 The endogeneity problem

This, however, is no easy task. Capital accumulation and financial constraints are determined simultaneously: financial constraints depend not only on the financial situation of the firm, but also on the size of the planned investment.

With complete markets and a type of uncertainty common to all agents, the net present value of a firm does not depend on the way it is financed. The Modigliani-Miller separation theorem holds that a firm's real capital allocation decision can be analysed independently of the financing decision - the structure of the asset side of the balance sheet is independent of the liability side. With asymmetric information, however, there will be a premium on external financing over and above a fair default premium which simply compensates for the fact that the debtor will not have to pay in certain states of nature. The creditor is less able than the debtor to evaluate the situation of the firm and the prospects of the investment project to be financed. The finance premium covers expected dead-weight losses caused by monitoring, costs of litigation, adverse selection and moral hazard. The important thing is that its *size depends on the financial structure of the firm*. Investment and the cost of external finance are therefore *jointly endogenous*.

Graph 1, adapted from Bernanke et al (1999), shows that the costs of external finance depend on the difference between the actual capital demand and what can be financed internally. By means of this graph, we can interpret the responses to the questions on financial constraints in terms of three regimes which are ordered in a natural way: a state of no financial constraints, a state of limited internal finance (the firm needing external finance) and a state of unavailability of external finance. If a firm states that its capital expenditure authorisations are limited by a shortage of internal finance, it is saying that it has to pay an external finance premium because the internal resources are insufficient. And if it reports that no further external finance can be raised, the firm may find itself in the regime described by Stiglitz and Weiss (1981): at a certain credit volume, the interest rate cannot be raised beyond a certain value. Then the firm is credit-rationed. Under certain circumstances, this is the equilibrium outcome of a situation where the severity of the agency problems is a function of the

5 Mitchell et al (2002a,b) show that survey responses contain information that is useful in generating indicators of manufacturing output ahead of the publication of survey data. Furthermore, they show that disaggregate indicators for output growth can outperform traditional aggregate measures with respect to their predictive content.

interest rate itself. In the graph, the existence of such a regime would make the schedule break off at some maximum interest rate.

We see that shocks to the financial structure will affect real decisions and vice versa. In any equation describing the capital accumulation decision, the error term will be correlated with the financial constraints variable. If we had continuous variables describing the accumulation of capital, this problem could be resolved using instrumental variable techniques, such as the GMM method developed by Arellano and Bond (1991). Breitung et al (2003) explore the simultaneity between investment decision and financial conditions by estimating a VAR on a large panel of German manufacturing firms. However, instrumental variable analysis is made difficult by the fact that the ITS data on investment and expansion are qualitative: we know whether or not the firm expands or steps up investment, but not by how much.

We therefore want to test the informational content of the data on financial constraints by looking at a relationship where both lines of causality point in the same direction. To this end, we investigate the occurrence and the duration of spells of capacity constraints.

3.2 Occurrence and duration of capacity restrictions

If there are adaptation costs such as delivery lags or time to build constraints, the move to a higher desired capital stock will be spread over several periods. In order to achieve tractability, it is often assumed that marginal adaptation costs increase linearly with the size of investment.[6] Second, the external finance premium might also be an increasing function of the investment intensity. Creditors might want to give finance in instalments, cutting the project into several phases, in order to monitor feasibility and the willingness of the management to comply with the terms of the credit contract. This may induce a sequential and "evolutionary" development of a project from a smaller to a larger size even in cases where, in a world without information asymmetry, a massive parallel investment effort might have been optimal. In the extreme case, when a firm has no access to external finance, the amount of investment per period is quite simply limited by the firm's cash flow.

The ITS survey gives us information on whether or not a firm experiences capacity constraints in a given period by asking the following question:

Question 14

What factors are likely to limit your output over the next four months?

(Please leave completely blank if you have no limits to output)

☐ orders or sales	☐ skilled labour	☐ other labour	☐ plant capacity
☐ credit or finance	☐ materials or components	☐ other	

Both directions of causation between financial constraints and the expansion decision lead us to predict that a state of capacity restrictions is more probable and will be of longer duration if the respondent also reports financial constraints to investment. With a given marginal valuation of capital, a large external finance premium will induce the firm to spread investment over a longer time horizon, inducing and prolonging capacity constraints. On the other hand, with a given financial structure, a shock in the marginal valuation function will not only trigger financial constraints, but also lead to a longer adaptation process. Larger gaps simply take more time to fill. Below, we shall compare the occurrence and duration of capacity constraints for restricted and unrestricted financing, with a particular emphasis on the distinction between small and large firms. Our analysis shows that the financial constraints data actually do have informational content at the micro level.

[6] See Hayashi's (1982) neoclassical micro-foundation of the Q model.

3.3 Association analysis for capacity restrictions and financial constraints

Table 5 compares the frequency of capacity restrictions for three groups of firms: those that do not seem to be limited by the lack of either internal or external finance (Group 1), those that complain about shortages of internal finance but not about inability to raise external finance (Group 2) and, finally, those that report being rationed on the market for external finance (Group 3). Whereas only 12.99% of the first group claims to be capacity restricted, the corresponding figures are 22.52% of the second group and 19.17% of the third group. The two latter groups are clearly different from the first group. We perform three statistical tests of association: the well known Pearson test, a likelihood ratio test and Fisher's exact test. Given two discrete (multinomial) variables, all three tests focus on how strongly the realised shares for one variable, conditional on the values that the other variable may take, deviates from the overall shares. Pearson's test and the likelihood ratio test are easily calculated and rely on asymptotic properties of the test statistic: for large numbers their distribution converges against the Chi(2) with $(r - 1)(s - 1)$ degrees of freedom, r being the number of rows and s being the number of columns in the contingency tables. Fisher's test exploits the exact distribution of the test statistic, but computation can take a very long time for larger tables.[7] All tests reject the null hypothesis of independence with a p-value of less than 0.0005.

It is also interesting to look at *changes* of states, as the association between the levels of the financial constraints and capacity restrictions might be the result of a special sensitivity to constraints in general on the part of the individual respondents. To put it differently: some individuals might have a special propensity to complain. Therefore we want to condition on the state of capacity restrictions in the preceding period. This examination also prepares our duration analysis: by definition, a switch from the state of not restricted to restricted initiates a spell of restricted capacity. If the restricted state is maintained, the spell goes on, and a reverse switch will end it.

Table 6 performs the three above-mentioned non-parametric association tests separately for firms that reported capacity restrictions in the preceding period and those that did not. Generally, capacity restrictions are cited much more frequently when there were restrictions in the previous quarter: whereas only 7.2% of the unrestricted firms switch to the restricted state, 53.3% of the restricted firms remain restricted. However, under both conditions the probability of capacity restrictions clearly becomes higher when financial constraints are present. Again, the three association tests mentioned above reject the null hypothesis of independence with a p-value of less than 0.0005.

3.4 The design of the duration analysis

The econometric analysis of duration data began only in the late 1970s (see Heckman and Singer (1984), Kiefer (1988) and Lancaster (1990) for overviews). Not only the statistical models, but also a good part of the terminology have been borrowed from biostatistics. The classical focus of "survival analysis" is the evaluation of survival times of human patients or animals after the contraction of a specific disease, with the aim of testing the effects of medical treatments and other factors that might potentially be of relevance. Among the economic applications have been the analysis of the duration of unemployment, for example by Steiner (1990), or of fiscal behaviour, as in the study by von Hagen et al (2001). To the best of our knowledge, the duration of capacity constraints has never been investigated before at a microeconometric level. This makes our exercise interesting and worthwhile in its own right, as capacity constraints may play an important role in the propagation of inflationary shocks.[8]

Here, we wish to consider the duration of states of restricted capacity. For a firm in this state, the probability of switching to the unrestricted state may depend on the duration that is already achieved. Such a conditioning on time is called "ageing", and the word itself makes the idea plain. Mortality among human beings is relatively high during the first months of life, and then drops sharply after a couple of years. In advanced age, mortality rises again and reaches extreme levels at the right end of the scale.

---

[7] See, for example, Büning and Trenkler (1994) or any other book on non-parametric statistics.

[8] See Macklem (1997).

In order to estimate survival curves, we therefore need to have information on the time when the period of constrained capacity began. We limit ourselves to contiguous strings of observations that start with a switch of the capacity restrictions variable from zero (no capacity restrictions reported) to one (output is likely to be limited by plant capacity during the next four months). The string is interrupted if the state is left, ie the "spell" ends, or else if there is no further information on the firm. One missing survey is enough to cut the string off. For inferential reasons, we can use only those observations which are not censored immediately after entry. That is, after the initial switch from zero to one, we need at least one more consecutive observation on the firm if the string is to contain any information on duration other than that it was non-negative. The cleaned CBI survey data for the period between January 1989 and November 1999 contain 49,244 observations on 5,169 firms. In this data set, we observe 1,431 of such strings, with a total of 5,153 observations,[9] taken from 862 firms.

We need to pay special attention to three important features of our data set. First, our duration data are censored considerably. From our 1,431 cases, we observe the end of the spell 1,210 times, but in the remaining 221 spells the string is cut off by missing observations. In these cases, we know that the spell has lasted *at least* until the end of the string, and this information has to be used appropriately. Second, we have *grouped data*. We do not observe the end of the spell in continuous time, but only know that it falls in an interval between two discrete points of time. Our observations are quarterly, and the vast majority of observed periods of capacity constraints are less than four quarters. This means that the granularity of our observations is rather high, and we believe that it would not be correct to use standard models and estimation procedures which assume observed duration times to be continuously distributed in time. Third, as already stated, we are working with a *panel* of survival time data. For many firms, we observe more than one spell. These cannot be assumed to be stochastically independent, and special care has to be taken with testing procedures.

3.5 Kaplan-Meier survival curves

We start by looking at the estimated *survivor functions*. A survivor function is defined for both discrete and continuous distributions by the probability that the duration T exceeds a value t in its range, that is:

$$F(t) = P(T > t), \quad 0 < t < \infty. \tag{1}$$

For each hypothetical duration t, the survivor function gives the share of individuals with duration of t or more. In our context, the survivor function depicts the process of firms liberating themselves from capacity constraints, once they have entered into this state. It gives the mass on the *right tail* of the distribution of duration times. This is convenient, because the right tail is the important component for the incorporation of right censoring.

The Kaplan-Meier[10] (or *product limit*) estimator is a non-parametric maximum likelihood estimator of the survivor function. The estimator is given by:

$$\hat{F}_t = \prod_{j \le t} \left(1 - \hat{\lambda}_j\right), \quad \text{with } \hat{\lambda}_j = \frac{d_j}{n_j}. \tag{2}$$

The index j enumerates observed times to completion, ie time spans passed since the observational unit entered into the risk pool. We only observe firms at discrete intervals, therefore the j can be thought of as quarters. The $\hat{\lambda}_j$ are estimated probabilities for the observational unit to complete at j, given that it has reached $j - 1$, the last observed time to completion. The estimate of these conditional probabilities is obtained by dividing the observed number of completions, d_j, by the number of observational units that have neither completed nor been censored before j.

As can be seen, the survivor function is estimated recursively. The expression $(1 - \hat{\lambda})$ is an estimation of the conditional probability that an individual "survives" in the state, given that it has lasted until $j - 1$.

9 This number of observations includes the initial zero and the initial one for each string.

10 For the derivation of the Kaplan-Meier estimator as a maximum likelihood estimator, see Kalbfleisch and Prentice (2002).

The unconditional probability that the duration is at least *j* is then computed as a product of all the contemporaneous and prior conditional survival probabilities. For this estimate to be unbiased, the censoring mechanism needs to be independent; that is, the completion probabilities of non-censored and censored individuals must be identical. This will be assumed throughout below.

Table 7 not only describes termination and censoring over time, but also gives the numerical values for the survivorship and completion rates in the entire sample. The first column, time, is the number of quarters *after* the original switch from unconstrained to constrained. If, for example, the capacity state of a firm switches from unrestricted to restricted in the third quarter of 1991, then for this firm the fourth quarter of 1991 assumes the value of one. The second column gives the number of firms "at risk", for which we have information in this quarter. The third column gives the number of completions, and the fourth column the number of firms censored in this quarter, on which there is no further information thereafter. The sixth column is the estimated Kaplan-Meier survivor function, based on the estimated hazard rates in the fifth column according to equation (2). According to this estimate, about 40% of firms that start out with capacity constraints remain in this state for more than one quarter, 20% for more than two quarters, etc. After the fifth quarter, the survivor function has dropped to 6.4%. The longest observed duration is completed after 13 quarters. Completion probabilities seem to be falling, ie there is negative age dependence. The more time a firm has spent in a state of constrained capacity, the less likely it is to leave in the next quarter. The size of the sample, on which duration information is based, decreases rapidly with time. After the fifth quarter, not more than 3.7% of the original set of firms is left in the sample. It therefore seems inappropriate to draw any conclusions from survival times longer than that. The last column gives the standard deviation of the survivor function, taking into account the stochastic dependence of the duration experiences for a given firm. The standard deviations are simulated on the basis of a maximum likelihood estimation of the parameters using 20,000 replications. Numerically, they differ only very slightly from what is obtained assuming all duration experiences to be independent.

Next we wish to look at survival experiences of financially constrained and unconstrained firms. The relative sizes of the groups and some global statistics are given in Table 8. The state is measured at the *start of the spell*. As before, there are two natural ways to analytically distinguish financially constrained and unconstrained firms. First, we can group a firm as financially constrained if it reports that it has to scale down investment because of insufficient internal funds. Second, we can classify it as financially constrained if it cites either shortages of internal finance or inability to obtain external finance. The difference between the two groupings is in those 44 spells where firms cite the inability to obtain external finance as a limitation to investment, without indicating shortages of internal finance at the same time. As such a pattern is incompatible with the standard pecking order view of corporate finance under financial constraints or the natural ordering that results from costly monitoring models as shown in Graph 1, we prefer the less ambivalent first grouping. Ultimately, the answer "costs of finance" as a limit to capital expenditure might indicate the working of the classical user cost mechanism. Therefore we do not use it as a sorting criterion.

We see that the prevalence of financial limitations is clearly higher among those firms that cite capacity restrictions. Whereas 25.3% of all capacity restriction experiences are categorised as "constrained" according to the first criterion, and 28.4% according to the second criterion, the corresponding figures for the entire CBI data set are 19.0% and 20.8%, respectively.

Graph 3 depicts the results for the first criterion (shortage of internal finance) for the whole sample. The survival curves for a split along the other criterion look almost the same. The survival curve for unconstrained firms is always beneath the curve for the financially constrained firms. This means the unconstrained firms are able to complete their spell of restricted capacity faster than the constrained firms. It is convenient to point out again that there are two competing causal explanations for this difference. For a given size of capacity gap, financially constrained firms might take longer to fill it. On the other hand, firms with a huge capacity gap (and accordingly higher financing needs) might be more likely to report financial constraints. Comparing the survival curves essentially tests those two hypotheses jointly. It will be necessary to examine this difference statistically.

3.6 A proportional hazard (Cox) model of duration

In order to test the effect of financial constraints on the duration of capacity restrictions, we need to impose some structure. Let x be a vector of characteristics, among them an indicator variable for financial constraints at the beginning of the spell. As we have little a priori information about the underlying process, we do not want to restrict the form of the baseline survivor function that corresponds to $x = 0$. In what follows, we explicitly recognise (1) that duration is distributed continuously over time, and (2) the measurement of the capacity restrictions for a given unit is taken at discrete intervals (quarters), $j = 1, 2, \dots k$.[11] Let $\lambda(t, x_i)$ be the *hazard* for a unit with characteristics x_i at time t, defined as:

$$\lambda(t,x) = \lim_{h \to 0} P\left(t \le T < t + h \mid T \ge t, x_i\right) / h \qquad (3)$$

The hazard is the instantaneous rate at which spells are completed by units that have lasted until time t, defined in the same way as a mortality rate in demographics or a failure rate in the statistical theory of capital stock dynamics. We want to assume that the characteristics x relate to the hazard rate in a proportional fashion:

$$\lambda(t,x) = \lambda_0(t) \cdot \exp(x_i'\beta), \qquad (4)$$

with β being a vector of coefficients that needs to be estimated. The hazard ratio between an individual with characteristics x_i and the baseline case is given by $\exp(x_i'\beta)$, which is approximately $1 - \beta$ for small β. The hazard ratios between two individuals with characteristics x_1 and x_0 are calculated as $\exp[(x_1 - x_0)\beta]$. Equation (4) constitutes the model of proportional hazard, developed by Cox (1972). In this setup, the baseline hazard remains completely unspecified, which is why the proportional hazard model figures among the semi-parametric approaches.

We assume that the spells of different firms are independent events and that the censoring mechanism is independent of the state of the firm. We can write the probability for the completion of a spell to be registered after j surveys as a product of conditional probabilities. This allows us to derive a likelihood function that contains β as well as further (incidental) parameters describing, for the baseline case, the conditional probability of completing in the time interval between $j - 1$ and j, given that $j - 1$ has been reached. For details, see Hosmer and Lemeshow (1999), Section 7.4, as well as Kalbfleisch and Prentice (2002), Section 5.8. The likelihood function here can be shown to be identical to that for a Bernoulli experiment with probabilities that depend on time as well as on x_i by means of a standard link function. The parameter estimates are asymptotically normally distributed. We take the panel nature of the data into account by computing robust standard errors, with clusters defined by firm identity.

Table 9 contains the maximum likelihood estimations for a Cox model with one covariate, as well as dummy variables carrying information on the sector and the time of origin of the spell. As explained above, we use two alternative definitions of financial constraints. The dummy variable *fin(1)* takes a value of one to indicate that the firm cites insufficient internal finance at the outset of the spell. The dummy variable *fin(2)* will be one if the firm cites either insufficient internal finance or inability to raise external finance. The respective classification is maintained during the entire spell.

In each cell, the first figure gives the estimated coefficients. Below, in curly brackets, this value is translated into a hazard ratio. Column 1, for example, compares the hazard rates for constrained and unconstrained firms according to our first criterion. The hazard rate of a constrained firm is exp(–0.192) times the hazard ratio of a small firm, meaning that financially constrained firms are leaving the state of restricted capacity at a rate which is only about 82.6% that of an unconstrained firm. The third figure, in round brackets, indicates the robust standard deviations, taking into account stochastic dependence between spells generated by the same firm. The last entry, in square brackets, gives the z statistic for statistical significance: under the null hypothesis of no differences, the estimated coefficient divided by its standard error is asymptotically a standard normal variate. Column 2 gives the corresponding estimates with respect to our second indicator of financial constraints, *fin*(*2*). The picture is essentially similar.

[11] The assumption of absolutely continuous time is made only for expositional convenience. A discrete time concept would not invalidate any of our results, after redefining the hazard rate in t as the conditional probability that the spell is completed in $t + 1$, conditional on it having lasted until t. It is possible to conduct duration analysis with distributions of T that have both discrete and continuous portions. See Kalbfleisch and Prentice (2002) for a systematic approach.

It may be argued that the detected differences may be sector-specific. As financial constraints may be sector-specific too, we want to control for sectoral differences in order to avoid a missing variable bias. Columns 3 and 4 repeat the estimates explained above, adding 20 dummies for two digit SIC sectors. This does not lead to a reduction of the financial constraint effects; if anything, the effect is bigger.

A third set of estimates, collected in columns 5 and 6, controls for the position in the business cycle by including dummies for the time of the start of the spell. This is done in order to account for a possible dependence of duration on the general state of the economy. In a time of depression, investors might be less inclined to close capacity gaps. At the same time, internal financial resources might be scarcer and external finance might be more difficult to obtain. In fact, adding the controls for the business cycle situation makes the size effects come out somewhat smaller, as predicted. In our preferred estimate, column 5, lack of internal financial resources lowers the hazard rate by about 18% with respect to the baseline case. The value is significant at a 1% level ($p = 0.006$).

4. Conclusions and outlook

Our association and duration analysis have shown that the CBI financial constraints data are not without informational content - as theoretically expected, financially constrained firms are more often capacity constrained and they take longer to close capacity gaps than unconstrained firms. This means we can take our survey data seriously. They indicate that financial constraints and real activity are indeed interrelated. Survey information on the ups and downs of financial constraint indicators can therefore be a valuable policy tool.

But the precise nature of that interrelationship is still open. Real investment decisions may certainly cause financial constraints, and on the other hand those financial constraints may slow down or prevent expansion plans. Further research is planned to separately identify the two directions of causation using a structural approach.

Finally, it will be interesting to take a more differentiated view. Are there subgroups (large firms, for example) for which financial constraints matter less? Are high-tech firms or innovators different from the rest? What about the importance of the state of the economy? And is it possible to analyse the role of the financial system by making international comparisons? Working with individual level survey data may be demanding, but, so the author believes, it can be highly rewarding.

Table 1

Breakdown of data set by employment size

	Employment size				
	1-199	**200-499**	**500-4,999**	**5,000 and over**	**Total**
No of firms	3,394	1,060	647	68	5,169
No of observations	31,089	10,222	6,994	939	49,244

Source: CBI, *Industrial Trends Survey*.

Table 2

Number of observations split by employment size and two digit SIC code

Two digit SIC code	Employment size				
	1-199	200-499	500-4,999	5,000 and over	Total
Coke ovens	17	6	17	0	40
Mineral oil processing	73	35	38	11	157
Nuclear fuel production	0	0	0	2	2
Extraction and preparation of metalliferous ores	35	0	0	0	35
Metal manufacturing	1,429	460	292	62	2,243
Extraction of minerals not elsewhere specified	493	60	103	9	665
Manufacturing of non-metallic mineral products	1,286	436	443	85	2,250
Chemical industries	1,191	722	641	79	2,633
Production of man-made fibres	142	8	32	1	183
Manufacturing of metal goods not elsewhere specified	3,048	651	308	6	4,013
Mechanical engineering	7,116	1,718	1,028	23	9,885
Manufacturing of office machinery and data processing	103	26	90	7	226
Electrical and electronic engineering	2,991	1,420	808	54	5,273
Manufacturing of motor vehicles and parts thereof	691	409	409	187	1,696
Manufacturing of other transport equipment	315	132	136	111	694
Instrument engineering	838	230	69	0	1,137
Food, drink and tobacco manufacturing industries part 1	473	250	420	43	1,186
Food, drink and tobacco manufacturing industries part 2	689	399	454	151	1,693
Textile industries	2,427	1,098	594	7	4,126
Manufacturing of leather and leather goods	295	63	2	0	360
Footwear and clothing industries	1,439	478	262	39	2,218
Timber and wooden furniture industries	1,258	313	154	1	1,726
Manufacturing of paper and paper products	2,854	668	489	38	4,049
Processing of rubber and plastics	1,698	563	169	22	2,452
Other manufacturing industries	188	77	36	1	302
Total	**31,089**	**10,222**	**6,994**	**939**	**49,244**

Source: CBI, *Industrial Trends Survey*.

Table 3

Investment constraints

	Inadequate net return	**Shortage of internal finance**	**Inability to raise external finance**	**Cost of finance**	**Uncertainty about demand**	**Shortage of labour**	**Other**	**N/a**
Any rank	38.71%	18.89%	4.30%	10.64%	54.88%	5.73%	1.76%	8.89%
Rank 1	28.14%	13.58%	1.96%	5.25%	44.51%	2.76%	1.58%	9.49%

Note: Firms ranking the constraint as a limit on capital expenditure authorisations, as a percentage of all firms, including those who did not answer the question at all. Respondents were able to give one or more responses, hence results do not sum to 100%.

Source: CBI, *Industrial Trends Survey*.

Table 4

Variability and persistence of financial constraints

	Unconstrained in *t*	**Constrained in *t***	**Total**
Unconstrained in $t-1$	19,990 87.61%	2,826 12.39%	22,816 100%
Constrained in $t-1$	2,377 36.68%	4,103 63.32%	6,480 100%
Total	25,162 79.45%	6,510 20.55%	31,672 100%

Note: Number and share of responding firms reporting either a shortage of internal finance or inability to raise external finance as a factor likely to limit capital expenditure over the next 12 months.

Source: CBI, *Industrial Trends Survey*.

Table 5

Association of capacity restrictions and financial constraints

		Capacity restrictions		
		Not restricted	**Restricted**	**Total**
Financial constraints	Not constrained	36,121 87.01%	5,394 12.99%	41,515 100%
	Internal finance	5,012 77.488%	1,457 22.52%	6,469 100%
	External finance	780 80.83%	185 19.17%	965 100%
	Total	41,913 85.63%	7,036 14.37%	48,949 100%

Association tests

Pearson's test:	Chi2(2) = 431.39	P < 0.0005
Likelihood ratio test:	Chi2(2) = 389.00	P < 0.0005
Fisher's exact test		P < 0.0005

Note: Number and share of responding firms reporting a shortage of internal finance or inability to raise external finance as a factor likely to limit capital expenditure over the next 12 months (rows) and number and share of firms reporting plant capacity as likely to limit output over the next four months (columns).

Source: CBI, *Industrial Trends Survey*.

Table 6

Association of capacity restrictions and financial constraints conditional on state of capacity restrictions in the previous period

Case 1: No capacity restrictions in previous period

		Capacity restrictions		
		Not restricted	**Restricted**	**Total**
Financial constraints	Not constrained	20,656 93.69%	1,392 6.31%	22,048 100%
	Internal finance	3,718 89.20%	450 10.80%	4,168 100%
	External finance	1,005 88.55%	130 11.45%	1,135 100%
	Total	25,379 92.79%	1,972 7.21%	27,351 100%

Association tests		
Pearson's test:	Chi2(2) = 137.18	$P < 0.0005$
Likelihood ratio test:	Chi2(2) = 124.07	$P < 0.0005$
Fisher's exact test		$P < 0.0005$

Case 2: Capacity restrictions in previous period

		Capacity restrictions		
		Not restricted	**Restricted**	**Total**
Financial constraints	Not constrained	1,616 49.60%	1,642 50.40%	3,258 100%
	Internal finance	385 39.29%	595 60.71%	980 100%
	External finance	97 38.49%	155 61.51%	252 100%
	Total	2,098 46.73%	2,392 53.27%	4,490 100%

Association tests		
Pearson's test:	Chi2(2) = 39.47	$P < 0.0005$
Likelihood ratio test:	Chi2(2) = 39.76	$P < 0.0005$
Fisher's exact test		$P < 0.0005$

Note: Number and share of responding firms reporting a shortage of internal finance or inability to raise external finance as a factor likely to limit capital expenditure over the next 12 months (rows) and number and share of firms reporting plant capacity as likely to limit output over the next four months (columns).

Source: CBI, *Industrial Trends Survey*.

Table 7

Survivor function and completion probabilities for the entire sample

Time	Beg total	Completed	Net lost	Completion rates	Survivor function	Std dev
1	1,431	856	133	0.5982	0.4018	0.0138
2	442	216	43	0.4887	0.2055	0.0123
3	183	63	16	0.3443	0.1347	0.0107
4	104	40	11	0.3846	0.0829	0.0090
5	53	12	7	0.2264	0.0641	0.0082
6	34	13	4	0.3824	0.0396	0.0074
7	17	3	2	0.1765	0.0326	0.0072
8	12	3	3	0.2500	0.0245	0.0069
9	6	3	0	0.5000	0.0122	.

Table 8

Composition of subsamples

Subsample	No of experiences	Times at risk	Incidence rates
All firms	1,431	2,291	0.528
Shortage of int finance	363	625	0.467
No shortage of int finance	1,068	1,666	0.551
Shortage of int or ext finance	407	703	0.472
No shortage of int or ext finance	1,024	1,588	0.553

Table 9

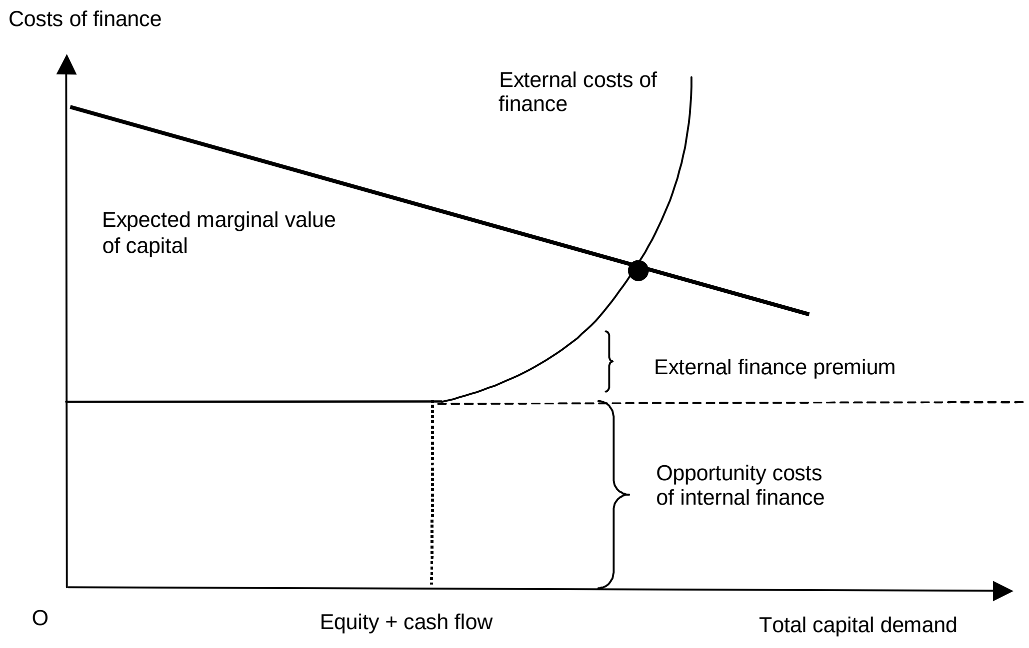
Maximum likelihood estimation of a proportional hazard model with grouped panel data

Coefficient	(1)	(2)	(3)	(4)	(5)	(6)
fin(*1*) (shortage of internal finance)	–0.192 {0.826} (0.072) [–2.65]***		–0.206 {0.814} (0.071) [–2.90]***		–0.199 {0.820} (0.073) [–2.72]***	
fin(*2*) (shortage of internal or external finance)		–0.181 {0.834} (0.068) [–2.68]***		–0.187 {0.830} (0.068) [–2.76]***		–0.172 {0.841} (0.068) [–2.54]**
Duration time dummies	9	9	9	9	9	9
Sector dummies	–	–	20	20	20	20
Dummies for time origin of spells	–	–	–	–	41	41
No of spells	1,431	1,431	1,429	1,429	1,429	1,429
No of firms	862	862	861	861	861	861
No of firm years	2,290	2,290	2,288	2,288	2,288	2,288

Note: Cox duration model with grouped data for spells of capacity constraints, estimated as a binary regression model using the complementary log-log function as link function. A spell is classified as pertaining to a financially constrained firm if, at the time when the spell starts, the firm reports financial constraints. The dummy variable *fin(1)* takes a value of one if a firm reports a shortage of internal finance in the answer to question 16c, otherwise it is zero. The dummy variable *fin(2)* takes a value of one if the firm reports either a shortage of internal finance or inability to raise external finance, otherwise it is zero. Likewise, a spell is classified as belonging to a large firm if the firm has 200 employees or more at the beginning of the spell. One observation had to be dropped because the longest duration interval (13 quarters) predicts the event perfectly. In the regressions reported in columns 3 to 6, two more observations and one sector (manufacturing of office machinery and data processing) were dropped because the sector dummy predicts the event perfectly. ** and *** indicate statistical significance at the 5% and 1% level, respectively.

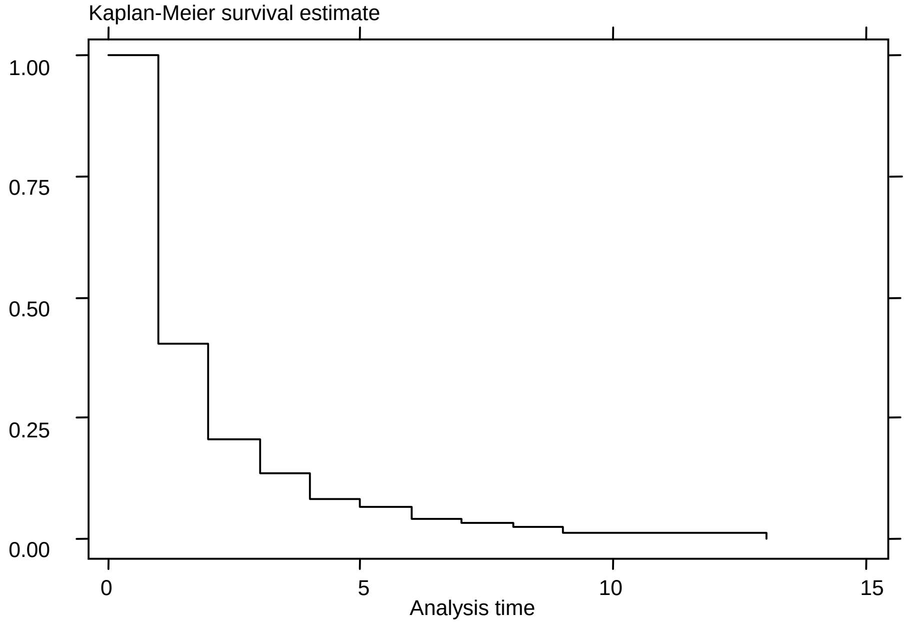
Graph 1

Capital demand and external finance premium



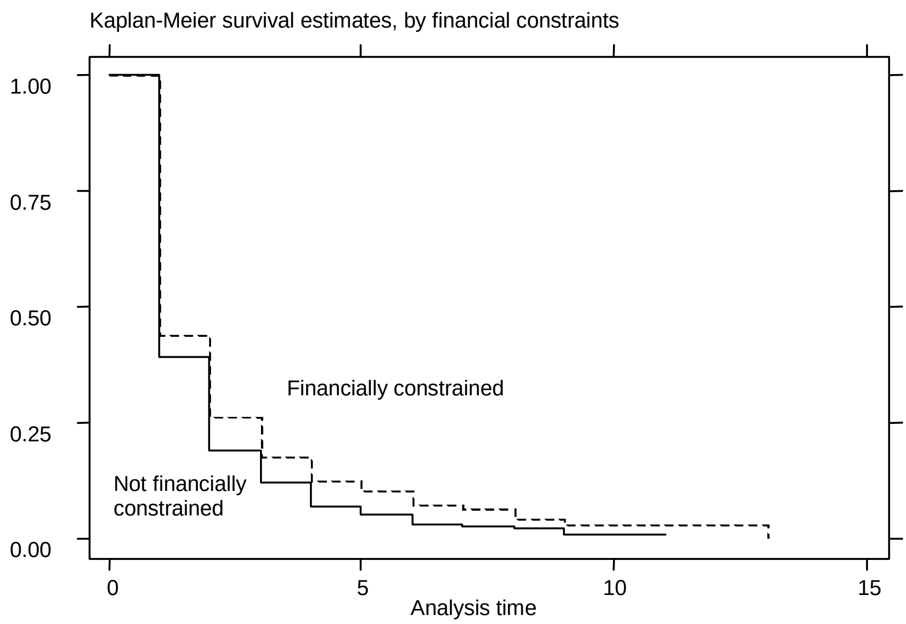
Graph 2

Kaplan-Meier estimates of the survivor function for the entire sample



Graph 3

Kaplan-Meier survival curves for financially constrained and unconstrained firms



References

Allen, Franklin and Douglas Gale (2001): "Comparative financial systems: a survey", *Center for Financial Institutions Working Papers* 01-15, Wharton School Center for Financial Institutions, University of Pennsylvania.

Angeloni, Ignazio, Anil Kashyap and Benoit Mojon (eds; 2003): *Monetary policy transmission in the euro area*, Cambridge, Cambridge University Press.

Arellano, Manuel and Stephen Bond (1991): "Some tests of specifications for panel data: Monte Carlo evidence and an application to employment equations", *The Review of Economic Studies*, 58, April, pp 277-98.

Bernanke, Ben S, Mark Gertler and Simon Gilchrist (1999): "The financial accelerator in a quantitative business cycle framework", in John B Taylor and Michael Woodford (eds), *Handbook of macroeconomics*, vol 1, Ch 21, Amsterdam, New York, etc, North Holland, pp 1341-93.

Berndt, Ernst R (1996): *The practice of econometrics*, Reading, Mass, Addison Wesley.

Breitung, Jörg, Robert S Chirinko and Ulf von Kalckreuth (2003): "A vectorautoregressive investment model (VIM) and monetary policy transmission: panel evidence from German firms", *Discussion Paper of the Economic Research Centre of the Deutsche Bundesbank*, no 6/03, January.

Büning, Herbert and Götz Trenkler (1994): *Nichtparametrische statistische methoden* [Non-parametric statistical methods], second edition, Berlin, New York, Walther de Gruyter.

Chatelain, Jean-Bernard, Ignacio Hernando, Andrea Generale, Ulf von Kalckreuth and Philip Vermeulen (2003a): "New findings on firm investment and monetary transmission in the euro area", *Oxford Review of Economic Policy*, vol 19, pp 73-83.

——— (2003b): "Firm investment and monetary transmission in the euro area", in Ignazio Angeloni, Anil Kashyp and Benoit Mojon (eds), *Monetary policy transmission in the euro area*, Cambridge, Cambridge University Press, forthcoming.

Chatelain, Jean-Bernard and André Tiomo (2003): "Monetary policy and firms' investment in France", in Ignazio Angeloni, Anil Kashyp and Benoit Mojon (eds), *Monetary policy transmission in the euro area*, Cambridge, Cambridge University Press, forthcoming.

Chirinko, Robert S and Ulf von Kalckreuth (2002): "Further evidence on the relationship between firm investment and financial status", *Discussion Paper of the Economic Research Centre of the Deutsche Bundesbank*, no 8/02, November.

Cox, David R (1972): "Regression models and life tables", *Journal of the Royal Statistical Society, Series B (Methodological)*, vol 34, no 2, pp 187-220.

Fazzari, Steven M, R Glenn Hubbard and Bruce C Petersen (1988): "Financing constraints and corporate investment", *Brookings Papers on Economic Activity*, 1, pp 141-95.

Gaiotti, Eugenio and Andrea Generale (2003): "Monetary policy and firm investment in Italy", in Ignazio Angeloni, Anil Kashyp and Benoit Mojon (eds), *Monetary policy transmission in the euro area*, Cambridge, Cambridge University Press, forthcoming.

Hall, Bronwyn (2002): "The financing of research and development", *Oxford Review of Economic Policy*, 18 (1), pp 35-51.

Hayashi, Fumio (1982): "Tobin's marginal q and average q: a neoclassical interpretation", *Econometrica*, 50, January, pp 213-24.

Heckman, James J and Burton Singer (1984): "Econometric duration analysis", *Journal of Econometrics*, 24, pp 63-132.

Hosmer, David W and Stanley Lemeshow (1999): *Applied survival analysis: regression modeling of time to event data*, New York, Chichester, etc, Wiley.

Kalbfleisch, John D and Ross L Prentice (2002): *The statistical analysis of failure time data*, second edition, Wiley.

Kaplan, Steven N and Luigi Zingales (1997): "Do investment-cash flow sensitivities provide useful measures of finance constraints?", *Quarterly Journal of Economics*, 112, February, pp 169-215.

——— (2000): "Investment-cash flow sensitivities are not valid measures of financing constraints", *Quarterly Journal of Economics*, 115, May, pp 707-12.

Kiefer, Nicholas M (1988): "Economic duration data and hazard functions", *Journal of Economic Literature*, 26, pp 646-79.

Lancaster, Tony (1990): "The econometric analysis of transition data", *Econometric Society Monographs* no 17, Cambridge, Cambridge University Press.

Macklem, Tiff (1997): "Capacity constraints, price adjustment and monetary policy", *Bank of Canada Review*, Spring, pp 39-56.

Mitchell, James, Richard J Smith and Martin R Weale (2002a): "Quantification of qualitative firm-level survey data", *The Economic Journal*, 112, March, pp C117-C135.

——— (2002b): "Aggregate versus disaggregate survey-based indicators of economic activity", *NIESR Discussion Paper* 194, March.

Modigliani, Franco and Merton H Miller (1958): "The cost of capital, corporation finance and the theory of investment", *American Economic Review*, 48, June, pp 261-97.

Steiner, Victor (1990): "Long term unemployment, heterogeneity and state dependence: new microeconometric evidence on unemployment persistence", *Empirica*, 17, pp 41-59.

Stiglitz, Joseph E and Andrew Weiss (1981): "Credit rationing in markets with imperfect information", *American Economic Review*, 71, June, pp 393-410.

von Hagen, Jürgen, Andrew Hughes Hallett and Rolf Strauch (2001): "Budgetary consolidation in EMU", *Economic Papers*, 148, Brussels, European Commission.

von Kalckreuth, Ulf (2003a): "Exploring the role of uncertainty for corporate investment decisions in Germany", *Schweizerische Zeitschrift für Volkswirtschaft und Statistik (Swiss Journal of Economics)*, vol 139, pp 173-206.

——— (2003b): "Investment and monetary transmission in Germany: a microeconometric investigation", in Ignazio Angeloni, Anil Kashyap and Benoit Mojon (eds), *Monetary policy transmission in the euro area*, Cambridge, Cambridge University Press.

Wood, Jonathan (2001): *A technical guide to CBI business surveys*, CBI, London.

A note on the recent behaviour of Japanese banks

Nobuo Inaba and Takashi Kozu[1]
Bank of Japan

1. Introduction

This note offers a brief analysis of Japanese banks' behaviour in recent years. Section 2 reviews the current situation at Japanese banks and Section 3 attempts to build a model which describes their behaviour. Although there is no single model that succeeds in explaining banks' behaviour consistently over the longer term, it is possible not only that their behaviour may be significantly affected by different factors in different periods but also that the same factor might have a different degree of impact depending on the period. In order to check the latter possibility, Section 4 focuses on the capital constraint and, making use of simulations within a dynamic model, reviews the influence of the capital constraint on banks' decision-making regarding the amount of write-offs.

2. Japanese banks in recent years

The Bank of Japan has been providing ample liquidity as part of its active pursuit of monetary easing and, as a result, overall financial market stability has been maintained (Figure 1). Within this environment, Japanese banks have been tackling management tasks such as the disposal of non-performing loans (NPLs).

The effects of the active monetary easing on banks' profitability, however, seem complicated. For example, the profitability of deposits, ie the margin between the deposit rate and the market rate, which had been narrowing since the beginning of the 1990s along with the deregulation of deposit rates, finally fell to zero with the introduction of the zero interest rate policy that forced short-term market rates up against the zero bound (Figure 2).

As for the disposal of NPLs, total credit costs at Japanese banks have exceeded operating profits from their core business since fiscal 1993 (Figure 3). In detail, write-offs of past NPLs have been accelerating (Figure 4) and the ratio of NPLs to total loans has started declining, albeit slowly (Figure 5). With regard to loan loss provisions, since fiscal 2002 major banks have adopted the discounted cash flow (DCF) method to calculate loan loss provisions for borrowers, with credit of ¥10 billion or more, classified as "special attention", and the loan loss provision ratio has risen (Figure 4). New NPLs, on the other hand, continue to arise, as Japan's economy is in the midst of structural changes. Under such circumstances, Japanese banks should assume, for the time being, comparatively high credit costs, say around 1% against their loans outstanding. It is, therefore, still very important for banks to earn sufficient profits to cover these credit costs.

Bank capital has become impaired not only because of these high credit costs but also because of stock market weakness. Since fiscal 2000 in particular, net unrealised stock-related gains have actually disappeared (Figure 6) and hence any losses that occur tend to impair capital. This tighter constraint on capital may have affected bank behaviour. For instance, during this process banks seem to have become more sensitive about the size of their loan assets, reducing overseas loans in the late 1990s and subsequently even domestic ones (Figure 7).

---

[1] We are grateful to the Department staff for the analyses in this note, especially Mr Junichi Suzuki for Section 2, Mr Shinobu Nakagawa for Section 3 and Appendix 1, and Mr Yutaka Soejima for Section 4 and Appendix 2. The views expressed here do not necessarily reflect those of the Bank of Japan. (Corresponding author: Takashi Kozu, e-mail address: takashi.kouzu@boj.or.jp).

3. Modelling banks' behaviour

This section attempts to build a theoretically grounded model to describe bank behaviour consistently. Considering the issues discussed in the previous section, it may be expected that building such a model would prove problematic, and in fact it proves not to be possible to build a model capable of providing a fully satisfactory explanation of the observed reality.

The model applied here is based on the assumption that the bank acts to maximise its present value and that its decision regarding the amount of loans to extend is dependent mainly on the loan margin. The following additional factors are also taken into account: (1) costs on loans, including losses from NPL disposal; (2) land prices, reflecting the value of collateral; (3) the constraint on capital; (4) net unrealised stock-related gains/losses; and (5) developments in the real economy. Appendix 1 explains the details of the model.

As bank behaviour may depend upon balance sheet size, the model was estimated for both major banks and regional banks. We also carried out estimations for four different periods: (a) the whole period, fiscal 1985-2001; (b) the bubble period, fiscal 1985-89; (c) the first half of the 1990s, ie fiscal 1990-96; and (d) the period from the second half of the 1990s onwards, ie fiscal 1997-2001.

The main results obtained may be summarised as follows (Figure 8):

- It is not possible to obtain a satisfactory explanation of the lending behaviour of both major and regional banks that holds true throughout the whole period.
- Changes in the price of land, which served as collateral for loans, affected the lending behaviour of both major and regional banks, in the sense that higher land prices acted to lower costs on loans and hence to increase them, in the bubble period.
- The constraint on banks' capital seems to have become binding, especially for major banks, since the second half of the 1990s. It was at this time that Japan experienced its banking crisis.

Thus it is difficult to describe the lending behaviour of Japanese banks precisely enough with a single optimisation model. However, the following possibilities can be pointed out. One is that bank behaviour might be crucially influenced by different factors in different periods. The other is that the same factor might have a different degree of impact depending on the period.

4. Simulations of bank write-offs

The second of the two possibilities introduced at the end of the previous section may apply to the capital constraint. When banks dispose of NPLs, they have to decide how much to write off. If they write off NPLs, they have to prepare for unexpected losses. However, future returns on loans should improve with the removal of unprofitable assets from their balance sheets. Capital constraints may affect this decision-making process. If the constraint is severely binding, banks may prefer to make provisions rather than to carry out write-offs since by doing so they would avoid unexpected losses and the resulting capital impairment. The extent to which the capital constraint is a binding factor in this decision-making process may vary depending on the period.

In order to check this point, we use a dynamic macro model to perform simulations. Figure 9 gives a brief description of the simulation algorithm. The bank's utility is assumed to be a function of its own expected future profits and the variance of this expectation. The bank is assumed to be facing uncertainty with regard to the macroeconomic condition in the future, about which it forms adaptive expectations. The bank goes bankrupt when its capital adequacy ratio falls below a certain minimum level.

Two time points, the beginning of fiscal 1997 and of fiscal 2001, are considered. The bank is assumed to have full information on the economic structure at the end of fiscal 1996 and fiscal 2000 respectively. Two hypothetical cases are considered: one where the bank is aggressive in carrying out write-offs, the other where it is not (Figure 10). The difference between the banks' respective utilities in these two cases can be obtained through simulations. Appendix 2 explains the details of the model and the way simulations are conducted.

The main simulation results can be summarised as follows:

- In fiscal 1997, the capital constraint proved a binding condition in determining the amount of write-offs carried out by the bank (Figure 11). According to the simulation, the probability at that time that the bank would go bankrupt was fairly high, especially in the "aggressive write-off" case. The bank was therefore cautious about being overly aggressive in its write-offs.
- This result is more or less the same even when the bank possesses perfect foresight about the future macroeconomic condition (Figure 12).
- In fiscal 2001, on the other hand, the incentive for the bank to be aggressive in its write-offs was stronger (Figure 13). This may reflect changes in the bank's situation, such as a gradual correction of the bank's once optimistic expectations about the future economic condition, as well as enhancement of the bank's capital via injections of public funds.

The above results coincide with the fact that banks have been more active in their writing-off of NPLs in recent years. In addition, major banks are trying to reduce their stock holdings, as stocks are regarded as assets which carry a relatively high price fluctuation risk given their current capital levels. Such a reduction allows them to ease their capital constraints and to achieve more effective use of their capital. The Bank of Japan launched a scheme to purchase stocks held by banks to support their efforts in this regard and to mitigate the negative effects of stock price fluctuations on their capital.

It is expected that the changes in the behaviour of Japanese banks reviewed in this note will become more firmly reinforced and this would contribute to improving their profitability over the coming years.

Appendix 1: Derivation of the optimal condition for bank behaviour

Model[2]

Consider the following representative bank value function (V):

$$V_t = \beta E_t \left[\sum_{\tau=0}^{\infty} CF_{t+\tau} \right] \tag{A1-1}$$

where β is the subjective discount factor, E_t is the expectations operator conditional on information available in period t, and CF denotes the cash flow earned in each period. We define the bank's cash flow as:

$$CF_t = r_{Lt}L_{t-1} + r_{St}S_{t-1} - r_{Ct}Call_{t-1} - r_{Dt}D_{t-1} - C_t \tag{A1-2}$$

where r_{Lt}, r_{St}, r_{Ct}, and r_{Dt} represent, respectively, the rates of return in period t on loans (L), securities (S), call money ($Call$), and deposits including debentures (D) outstanding at the end of period $t-1$.[3] C describes a cost function on loans which we specify as:

$$C_t(FL_t, L_{t-1}, D_{t-1}) = a_1 FL_t + \frac{a_2}{2}\frac{FL_t^2}{D_{t-1}} + a_3 L_{t-1} \tag{A1-3}$$

where FL_t is the net flow of loans in period t, assuming that, as new loans increase, credit exposure also increases to borrowers about whom available financial information is insufficient, resulting in higher monitoring costs for the bank ($a_2 > 0$).[4] The parameter a_3, on loans outstanding at the end of period $t-1$, is regarded as a proxy for the magnitude of non-performing loans (NPLs) generated in period t, and is thus supposed to enter positively in equation (A1-3) ($a_3 > 0$).[5] In short, costs on loans here include losses from NPL disposal as well as the implicit general and administrative expenses incurred in loan management. In the meantime, the larger the deposits, a proxy for bank scale, the more likely it is that loan portfolios will be diversified, and we therefore incorporate deposits as a scale variable acting to mitigate costs on loans.

We also give the impact of changes in land prices (P_L) on the parameter a_3, which is expressed as:

$$a_3 = a_4 - a_5 \frac{P_{Lt-1}}{P_{Lt-2}} \tag{A1-4}$$

What equation (A1-4) implies is that appreciation in the value of land helps the bank to secure loans (ie its collateral role on loans), which is empirically found in the US bank data by Berger and Udell (1995).[6] If this implication is true, the sign condition will be that $a_5 > 0$.

---

2 In building a model, we owe much to work by Elyasiani et al (1995) and Ogawa and Kitasaka (2000) aimed at capturing the bank's optimal behaviour.

3 When the bank takes out a net call loan, we interpret this to mean that it has a negative holding of call money. Since banks can generally control both holdings of and returns on negotiable certificates of deposits (NCDs) and straight bonds, they are not included in deposits.

4 Although FL_t should be new loans made in period t in this sense, we use the difference in loans outstanding from period $t-1$ to t due to the availability of such data.

5 If a_3 is properly estimated, it should not be substantially different from actual credit costs (the NPL ratio) at banks.

6 In Japan, movements in land values are almost perfectly negatively correlated with movements in the number of firm bankruptcies.

The balance sheet condition requires that the following identity holds:

$$L_t + S_t + R_t + OA_t = D_t + Call_t + K_t + OL_t \tag{A1-5}$$

where R is bank reserves defined such that $R_t > \rho D_t$ (ρ: required reserve ratio, assuming simply that $R_t = \rho D_t$ in the optimal representative bank case), K denotes capital, and OA_t and OL_t represent, respectively, other assets and liabilities at the end of period t.

Without restrictions on asset management and given *a* change in deposits that is exogenously determined via the consumer's optimal resource allocation, the bank's optimal strategy is to choose {L, S, $Call$} in each period in order to maximise the value function (A1-1) subject to equations (A1-2) to (A1-5). Solving this dynamic optimisation problem yields the following first-order condition:

$$\frac{\partial C_t}{\partial FL_t} = \beta E_t\left[r_{Lt+1} - r_{Ct+1} - \frac{\partial C_{t+1}}{\partial L_t} + \frac{\partial C_{t+1}}{\partial FL_{t+1}}\right] \tag{A1-6}$$

Decomposing conditional expectation terms into their certainty equivalent values and an expectation error under the assumption of rational expectations and rearranging them, we obtain the bank's optimal lending function:

$$\frac{FL_t}{D_{t-1}} - \beta\frac{FL_{t+1}}{D_t} = b_0 + b_1\left(r_{Lt+1} - r_{Ct+1}\right) + b_2\frac{P_{Lt}}{P_{Lt-1}} + u_{t+1} \tag{A1-7}$$

where:

$$b_0 = \frac{\beta(a_1 - a_4) - a_1}{a_2}$$

$$b_1 = \frac{\beta}{a_2} > 0$$

$$b_2 = \frac{\beta a_5}{a_2} > 0,$$

and u_{t+1} is an expectation error uncorrelated with any information in period t.

In fact, bank lending behaviour has been restricted by the Basel Accord formally introduced in 1993, which is defined simply:

$$K_t \geq \kappa L_t, \tag{A1-8}$$

where κ is the required capital adequacy ratio.[7, 8] Taking account of this restriction and applying the first-order Kuhn-Tucker condition to the optimisation problem, we obtain the Euler equation to be estimated as:

$$\frac{FL_t}{D_{t-1}} - \beta\frac{FL_{t+1}}{D_t} = b_0 + b_1\left(r_{Lt+1} - r_{Ct+1}\right) + b_2\frac{P_{Lt}}{P_{Lt-1}} - \frac{\kappa}{a_2}\lambda_t + u_{t+1} \tag{A1-9}$$

where λ_t is the non-negative Lagrange multiplier associated with the bank's capital requirement restriction.[9] Since λ_t is unobservable, the fourth term on the right-hand side of equation (A1-9) is set to be $b_3\kappa_t$ in later estimations, where $b_3 > 0$ and κ_t is the actual capital ratio.

7 Note that, strictly speaking, L_t on the right-hand side of inequality (A1-8) should be the weighted risk assets derived from the BIS formula. Ito and Sasaki (1998) estimate the impact of the Basel capital standard on Japanese banks' behaviour, and confirm its significance empirically.

8 We do not account here for the existence of the bank lending channel, used to refer to the quantitative effect whereby deposits on the liability side affect loans on the asset side. Although this effect is empirically observed in the US bank data and documented in Kashyap and Stein (1997), we simply assume here perfect substitutability between deposits and money in the short-term financial market.

Data

In estimating the Euler equation (A1-9), we employ annual settlement data from the accounts of 10 major and 113 regional Japanese banks.[10] Our sample data run from fiscal 1982 to fiscal 2002. Due to data availability, the capital ratio (κ_t) is defined as core capital (Tier 1) divided by loans outstanding at the end of each period. For simplicity, the subjective discount factor (β) is set to be the average of the reciprocal of real gross returns on 10-year government bonds (deflator: GDP deflator) in the corresponding estimation period, and this is assumed to be common across all banks. Land prices (P_L) are obtained from the Japan Research Institute of Real Estate, and we assume that banks face different land prices, depending on the location of their head offices. If a bank is located in one of the six largest cities, we use the "six largest cities" land price index for that bank. Otherwise, we use the "other cities" land price index (which excludes the six largest cities).

Estimation method

Under the assumption of rational expectations, the error term, u_{t+1}, is uncorrelated with any variables known in period t. However, the Euler equation (A1-9) includes variables in period $t + 1$, and thus we use the iterative weighted two-stage least squares (2SLS) to estimate it as a system with the time-series, cross-sectional data.[11] Instrumental variables are the constant, twice-lagged dependent variables, the twice-lagged loan-call rate spread ($r_{Lt+1} - r_{Ct+1}$), once-lagged growth in stock values listed in the first section of the Tokyo Stock Exchange, and once-lagged growth in nominal GDP.[12] The purpose of including stock values and nominal GDP in the set of instruments is to consider the impacts of hidden profits from banks' stock holdings and of demand for bank loans by firms on the model.

In the estimation, the constant term b_0 in the system is often regarded as a factor that is idiosyncratic for each agent. There are two well known cases: the "fixed effects" and "random effects" cases. Since, even if estimated, these effects are not significant, we do not consider them in the estimation here. This is equivalent to carrying out a pooling estimation in which it is not only the parameters b_1 to b_3 in the Euler equation (A1-9) that are all common across all sample banks, but also the parameter b_0.

The estimation period is split into several subperiods: from fiscal 1985 to fiscal 1989 (the bubble period); from fiscal 1990 to fiscal 1996 (the first half of the 1990s); and from fiscal 1997 to fiscal 2001 (the second half of the 1990s onwards). We also consider the entire period from fiscal 1985 to fiscal 2001. All parameters reported in Figure 8 are estimated using a simultaneous weighting matrix and coefficient control, where the convergence criterion is 1.0E-07.

---

9 A good example of deriving the Euler equation (A1-9) is found in Zeldes (1989), in which the impact of quantitative borrowing constraints on consumers' optimal resource allocation is evaluated.

10 Due to the fact that mergers and nationalisation cause non-adjustable data discontinuities during the sample period, Shinsei Bank and Aozora Bank are excluded from the major bank sample, while Tokyo Star Bank and Kansai Sawayaka Bank are excluded from the regional bank sample. Mizuho Bank and Mizuho Corporate Bank, likewise Risona Bank and Saitama Risona Bank, are regarded as single banking entities, thus yielding the samples of 10 major banks and 113 regional banks.

11 Estimation results using the 3SLS method are basically the same, and thus are not reported in this note. They are available from the authors on request.

12 Other candidates for instruments can be lagged values of other independent variables in the Euler equation (A1-9). Even if they are included in the set of instruments, however, we find no significant changes in the results.

Appendix 2:
Model for simulating the bank's decision regarding write-offs

Model

The bank utility is determined by the mean and standard deviation of the present value of future profits:

$$U(\mu(PV), \sigma(PV)) \tag{A2-1}$$

The present value depends on expected profits $E_t(\pi_{t+j})$ over the next six half-year periods and is given by:

$$PV(\pi) = \sum_{j=1}^{6} \beta^j E_t(\pi_{t+j}) \tag{A2-2}$$

where we assume a unit subjective discount factor and a zero discount factor for inflation in the nominal value of profits.

Bank capital, Cap_t, is the state variable, the path of which is determined by the transition equation:

$$Cap_t = Cap_{t-1} + \pi_t - Tax_t - Div_t - Others_t \tag{A2-3}$$

where profits (π_t) reflect credit costs such as write-offs and loan loss provisions. The profit surplus, after deducting taxes (Tax_t), dividends (Div_t) and other factors ($Others_t$), determines the path of bank capital over time, as described in the transition equation (A2-3). Tax_t includes government capital injections into banks, and $Others_t$ covers other factors that affect bank capital, such as the introduction of deferred tax assets and any surplus from the revaluation of the bank's land holdings.

Profits π_t are defined by:

$$\begin{aligned} \pi_t &= R_t^L \cdot Loans_t - R_t^D \cdot Deposits_t - AdCosts_t - CrCosts_t, \quad if\ Cap_t / Loans_t \geq AdequacyRatio \\ &= 0, \quad \text{otherwise} \end{aligned} \tag{A2-4}$$

where R_t^L corresponds to the average rate of return on loans (obtained as total revenue divided by outstanding loans), R_t^D captures the average cost of funding (hereafter the "average funding rate", obtained by dividing total funding costs by outstanding deposits), and $AdCosts_t$ measures administrative costs including payroll costs.

The credit costs ($CrCosts_t$) reflected in NPL disposals comprise four parts: write-offs of new bad loans ($NewWO_t$); write-offs of existing bad loans not covered by loan loss provisions ($WO2_t$); loan loss provisions for new bad loans ($NewLLP_t$); and additional loan loss provisions for bad loans which have been partly covered by past loan loss provisions ($dLLP_t$). The total of $WO2_t$ and $dLLP_t$ corresponds to secondary losses, that is, unexpected losses which could not be predicted at the time the bank made its decision regarding disposals. Write-offs of bad loans with loan loss provisions ($WO1_t$) impair neither current profits nor bank capital, because these credit costs regarding $WO1_t$ were reflected in previous profits as either $NewLLP_{t-j}$ or $dLLP_{t-j}$ ($t - j < t$). Write-offs of $WO1_t$ reduce possible losses via $dLLP_{t+j}$, but this obliges the bank to give up the "real option" value inherent in bad loans, ie the possibility that these loans may become performing again.

The bank balance sheet constraint is:

$$Loans_t = Deposits_t + Cap_t \tag{A2-5}$$

We assume that all assets take the form of loans and all liabilities are deposits. The average rate of return on loans R_t^L, therefore, represents a gross based ROA, covering revenue from securities, fees and commissions, in addition to income from lending. The average funding rate R_t^D also covers funding from money and bond markets in addition to deposits.

R_t^L and R_t^D are determined by imposing equilibrium on the bank loan and deposit markets. First, firms' demand for bank loans and the bank's supply of loans are assumed to be functions of the following variables:

$$Loans_t^D = f^{LD}(NGDP_t, R_t^L) \quad \text{and} \tag{A2-6}$$

$$Loans_t^S = f^{LS}(R_t^L, Deposits_t, BLratio_{t-i}, CapRatio_t), \quad i = 0,1 \tag{A2-7}$$

where $NGDP_t$ is nominal GDP, $BLratio_t$ is the ratio of bad loans to total loans, and $CapRatio_t$ is the capital adequacy ratio obtained simply by dividing bank capital by bank loans (total assets). A high $BLratio_t$ negatively affects the bank's supply of loans for a given R_t^L, because it requires a premium for taking on the higher credit risk. The equilibrium condition in the loan market provides a reduced form of R_t^L, which is estimated by:

$$R_t^L = l_0 + l_1 NGDPdot_t + l_2 Depodot_t + l_3 BLratio_t + l_4 BLratio_{t-1} + \varepsilon_{Lt} \tag{A2-8}$$

where $NGDPdot_t$ represents the growth rate of nominal GDP, and $Depodot_t$ the growth rate of deposits. The bank capital constraint on lending, $CapRatio_t$, is omitted from the regression because the term is insignificant.

Second, the bank's demand for deposits and households' supply of deposits are assumed to take the following shapes:

$$Deposits_t^D = f^{DD}(R_t^D, Loans_t, Call_t, ExR_t) \quad \text{and} \tag{A2-9}$$

$$Deposits_t^S = f^{DS}(R_t^D, NGDP_t) \tag{A2-10}$$

where $Call_t$ is the call rate and ExR_t is the bank's reserves in excess of requirements. A reduced form of R_t^D obtained from the deposit market equilibrium condition is estimated by:

$$R_t^D = d_0 + d_1 NGDPdot_t + d_2 Call_t + d_3 \ln(ExR_t) + \varepsilon_{Dt} \tag{A2-11}$$

$Loans_t$ is omitted due to its insignificance in the regression.

Bank loans comprise both bad loans and good loans: the former are given by the "Risk Management Loans" disclosed by government and the Japanese Bankers' Association, while the latter are made up of the remaining loans. This gives:

$$Loans_t = BadLoans_t + GoodLoans_t \tag{A2-12}$$

Bad loans are divided into two categories; bad loans fully covered by loan loss provisions and bad loans proving not to be covered. We denote the former as LLP_t and the latter as $Naked_t$. The transition of LLP_t is given by:

$$LLP_t = LLP_{t-1} + NewLLP_t + dLLP_t - WO1_t \tag{A2-13}$$

while the transition of $Naked_t$ is:

$$Naked_t = Naked_{t-1} + NewNaked_t - dLLP_t - WO2_t \tag{A2-14}$$

We find that the transition of bad loans:

$$LLP_t + Naked_t (= BadLoans_t) = LLP_{t-1} + NewLLP_t + Naked_{t-1} + NewNaked_t - WO1_t - WO2_t \tag{A2-15}$$

is independent of $dLLP_t$ and $NewWO_t$, and only $WO1_t$ and $WO2_t$ can effect reductions in the outstanding amount of bad loans. New bad loans during period t, $NewBL_t$, are assumed to depend on the nominal economic growth rate ($NGDPdot_t$):

$$NewBL_t = f^{NewBL}(NGDPdot_t, dummy_{t=i}) \quad i = \text{FY95:2, FY97:2, FY98:2} \tag{A2-16}$$

$NewBL_t$ is divided into three categories: $NewNaked_t$, $NewLLP_t$ and $NewWO_t$. The ratios among the two types of disposals ($NewLLP_t$ and $NewWO_t$) and the uncovered outstanding amount of bad loans ($NewNaked_t$) are determined by historical data on new bad loans and their disposal.

Equations (A2-3), (A2-5) and (A2-15) provide us with an expression describing the transition of good loans:

$$\begin{aligned} GoodLoans_t - GoodLoans_{t-1} = (Deposits_t - Deposits_{t-1}) + (NewWO_t + WO1_t + WO2_t) \\ - NewBL_t + (\pi_t - Tax_t - Div_t - Others_t) \end{aligned} \tag{A2-17}$$

What this equation implies is that, supposing deposits remain unchanged, the bank's balance sheet freedom to expand its good loans depends on (i) total write-offs, (ii) new bad loans, and (iii) profit surplus.

The decrease in bad loans through write-offs improves future profits via a recovery in the rate of return on loans, as described in equation (A2-8), and also via the freedom to extend new good loans. In contrast, the write-off impairs current capital and therefore increases the risk of coming up against the constraint imposed by capital adequacy regulation, as is seen in equation (A2-4). The trade-off between the improvement in future profits and the rise in the risk of bankruptcy determines the optimal choice of write-offs. This optimal choice is dependent on the different business conditions faced by the bank at each stage, for example: its capital adequacy, its expectations of future economic growth, and the extent of its bad loans.

Details of simulation

The bank we examined in the simulations is a representative agent endowed with the aggregate figures of the banking accounts of all banks in Japan. The data run from the second half of fiscal 1992 to fiscal 2002, because NPL-related data are available only for this period.

For the initial values of all simulation variables, we adopt the value observed at the end of fiscal 1996 and 2000 respectively. For the exogenous deterministic variables, we make use of static expectations (Figure 9). Taxes, dividends and other factors in equation (A2-3) are omitted from the simulation, because it is difficult to make use of static expectations for variables which fluctuated significantly as a result of government policies such as capital injections. The influence of these variables is reflected in the initial values for each simulation. For technical reasons, administrative costs, which we regard as a proxy of payroll costs, are added to profits in the simulation.

Since there is only one stochastic factor, the nominal GDP growth rate, the distributions of the present values depend on how the bank forms its expectations of the growth rate. We assume that the expectations are adaptive, that is, the bank expects the growth rate to follow an AR(1) process with the same mean and variance as in the last six half-year periods. The choice of six periods derives from a survey on firm expectations of the real economic growth rate that suggests it takes about three years for firms to correct mistaken expectations by observing actual growth rates. The AR coefficient is estimated to be 0.77 over the full sample period. The means and variances for the two simulations with different initial starting periods are shown in the appendix table.

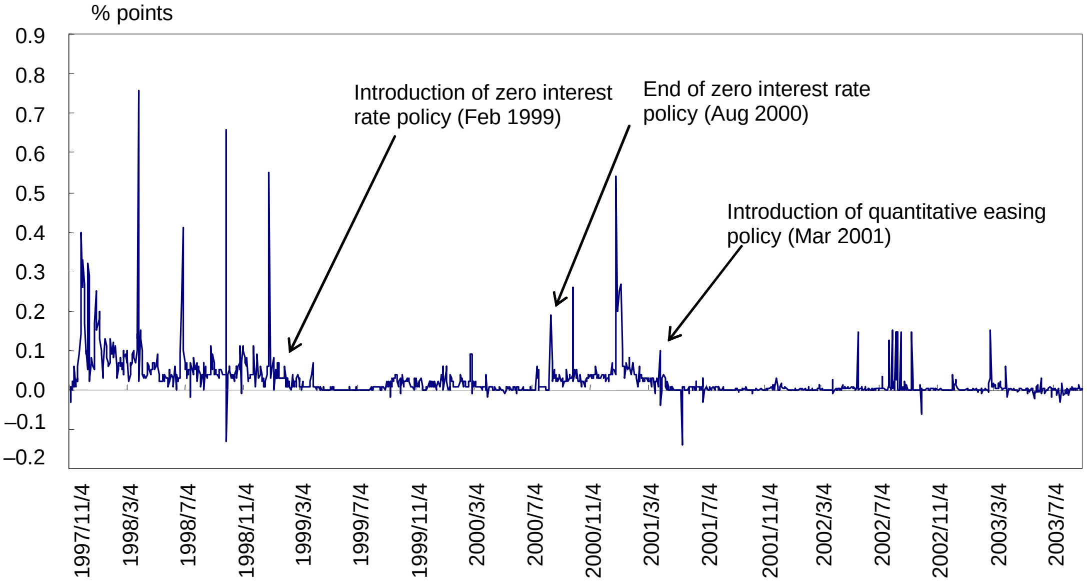
The simulation of "not aggressive" write-offs, starting in fiscal 1997, is based on actual figures for write-offs: the average of the first and second halves of fiscal 1996. The simulation of "aggressive" write-offs produces an amount some ¥2 trillion larger, almost the same as the average from fiscal 1997 to 1999 when the government adopted a strong initiative to push forward NPL disposal during and after the banking crisis. "Aggressive" write-offs in the simulation starting in fiscal 2001 are based on average write-offs in fiscal 2000. Write-offs in the "not aggressive" case are set to be ¥1 trillion less, almost the same as their average in fiscal 1996.

We carried out 100,000 simulations for each case. When a bank comes up against the minimum capital adequacy bound, this acts to terminate the loop in the updating process for the state variables: bank capital, and the bad loan components, LLP_t and $Naked_t$. Distributions of present values shown in Figures 11-13, where there are spikes at low levels of present values, suggest that some banks gain profits only at early stages of the simulation and then go bankrupt. Bell-shaped distributions at higher levels of present values in these figures correspond to the cases where banks are still alive at the end of the simulation period. These distributions show that banks can enjoy higher profits in the future through their aggressive write-offs only if they are able to survive the capital damages that accompany NPL write-offs. The appendix table illustrates how aggressive write-offs improve the average rate of return on loans (if banks remain alive with high probability), while at the same time impairing capital levels.

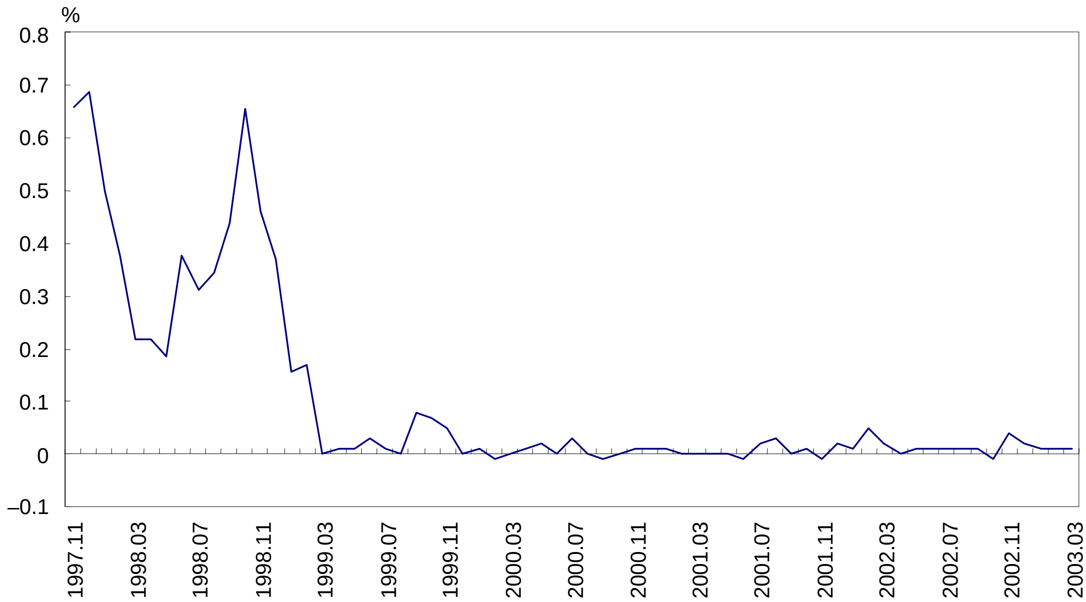
Figure 1

The liquidity premium
and the Japan premium

(1) The liquidity premium
(T/N call rate – O/N call rate, unsecured)



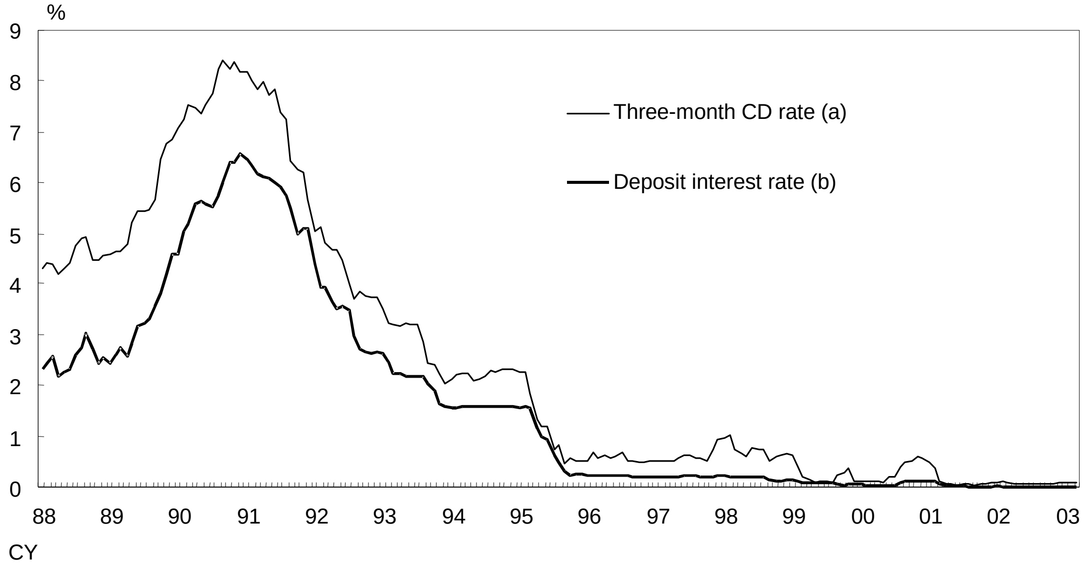
(2) The Japan premium



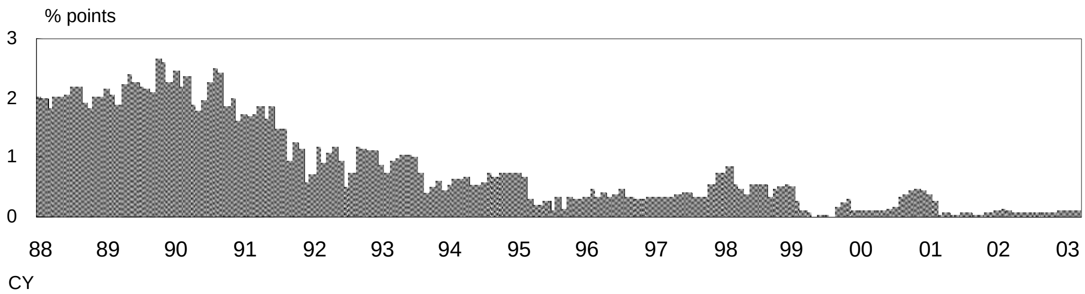
Note: The Japan premium is defined as the spread in Libor between Barclays and the Bank of Tokyo Mitsubishi.

Figure 2

Interest rates and deposit margin



Spread

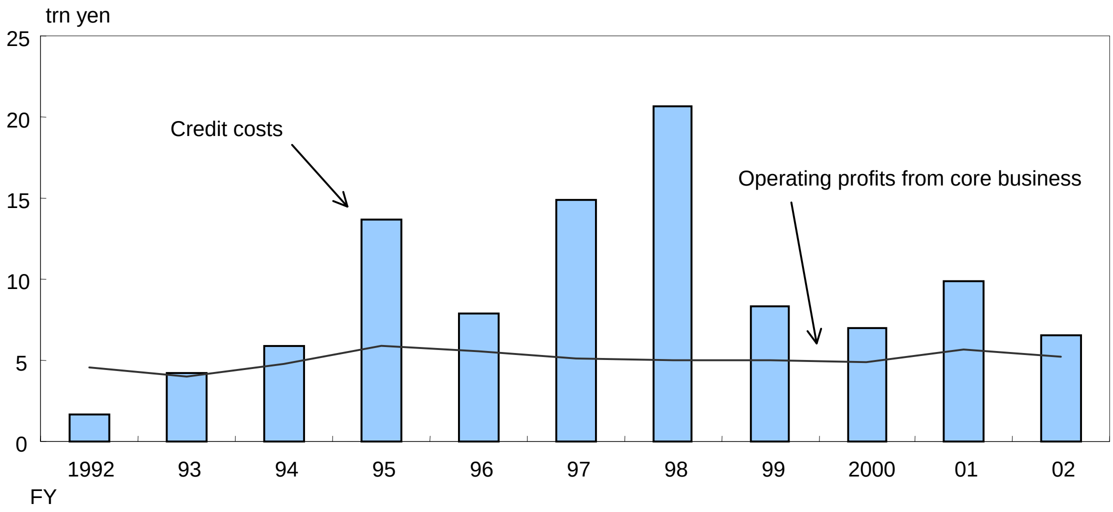


Note: Deposit interest rate = interest rate on three-month time deposits of less than ¥3 million.

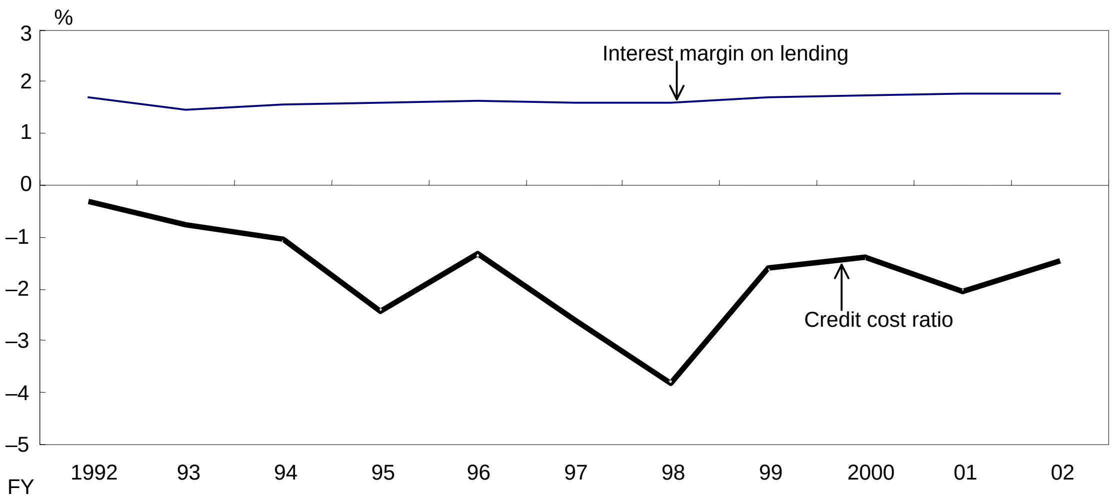
Figure 3

Credit costs and profits

(1) Credit costs and operating profit from core business (all banks)



(2) Interest margin on lending and credit cost ratio (all banks)

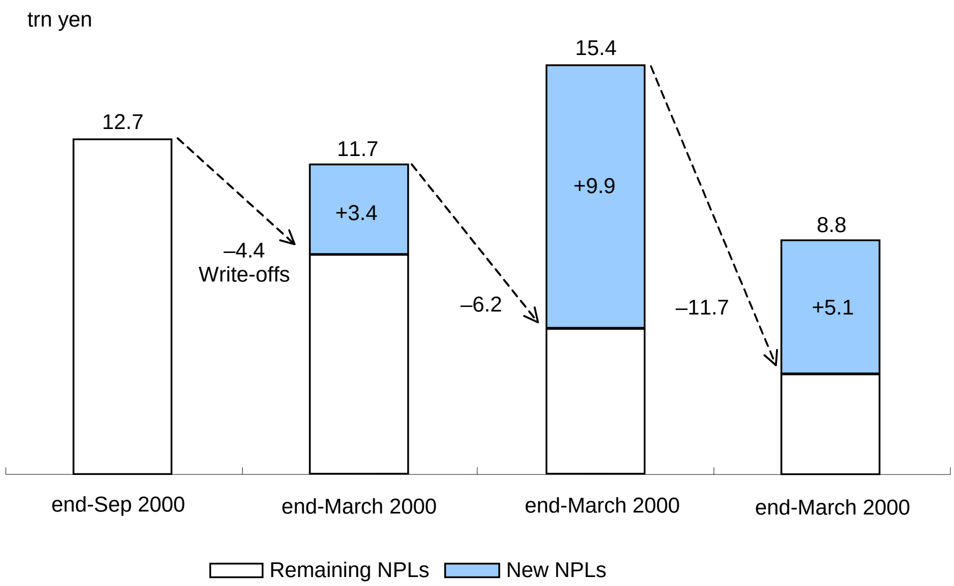


Note: Credit cost ratio = credit costs/total loans.

Figure 4

Progress in NPL disposal

(1) Removal of NPLs from balance sheets (major banks)



Notes: 1. NPLs here cover loans to borrowers classified as "bankrupt", "de facto bankrupt" and "in danger of bankruptcy". 2. Major banks here exclude Shinsei Bank and Aozora Bank.

(2) Loan loss provision ratio (provisions/total loans)

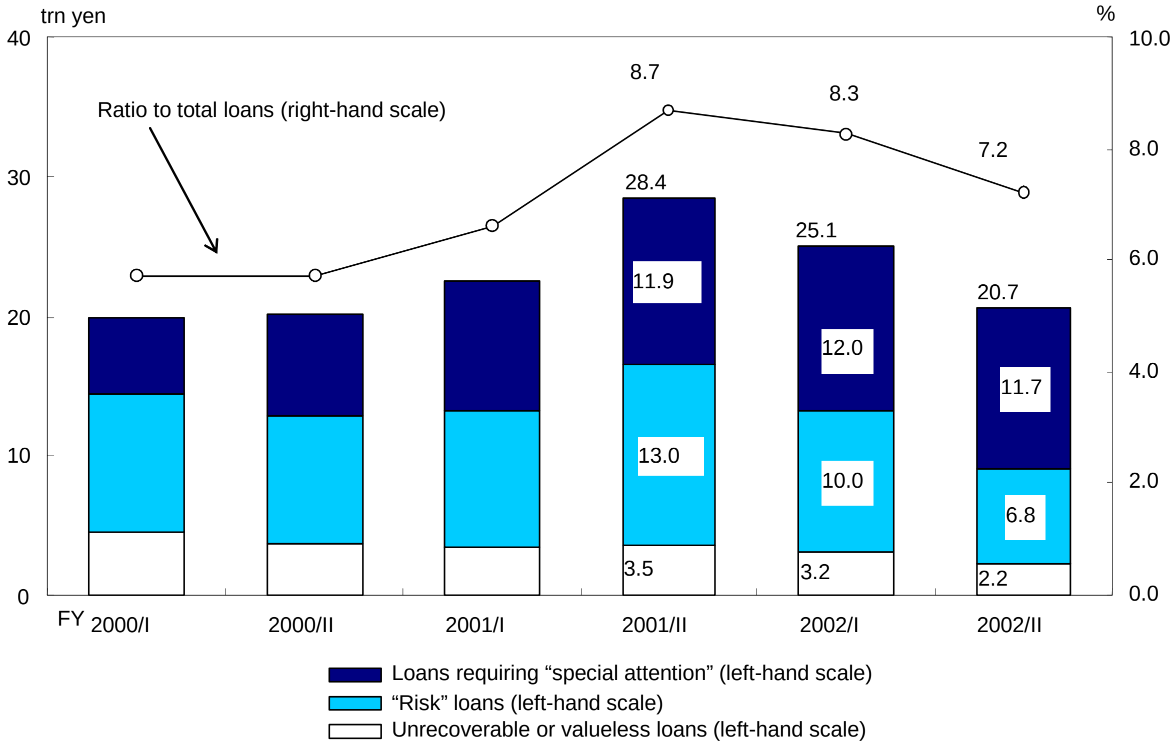
	All banks	Major banks (excluding Shinsei Bank and Aozora Bank)
Loans to "normal" borrowers and borrowers that "need attention"	1.4 (1.1)	1.7 (1.2)
Excluding loans requiring "special attention"	0.8 (na)	0.8 (0.7)
Loans requiring "special attention"	19.1 (na)	20.8 (14.2)
Loans to borrowers "in danger of bankruptcy"	33.6 (na)	39.4 (37.0)

Note: Percentage, at end-March 2003; figures in parentheses are at end-March 2002.

Figure 5

Non-performing loans

(1) Major banks



(2) Regional banks

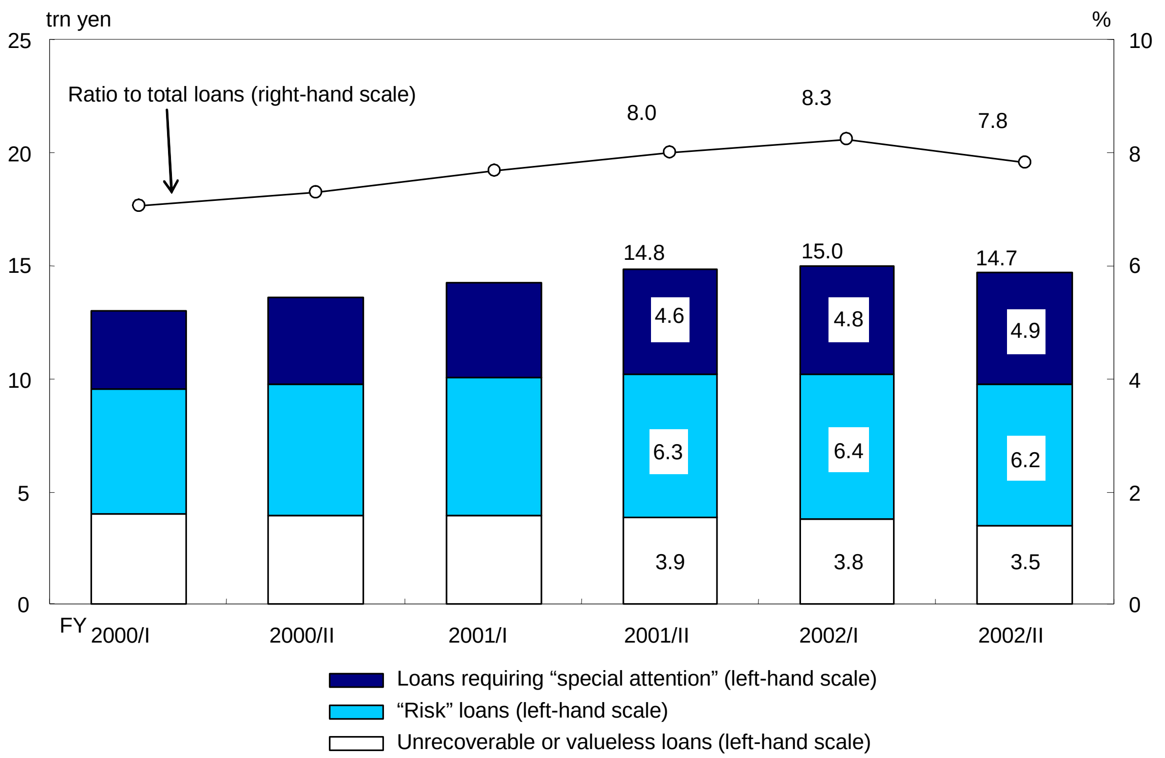
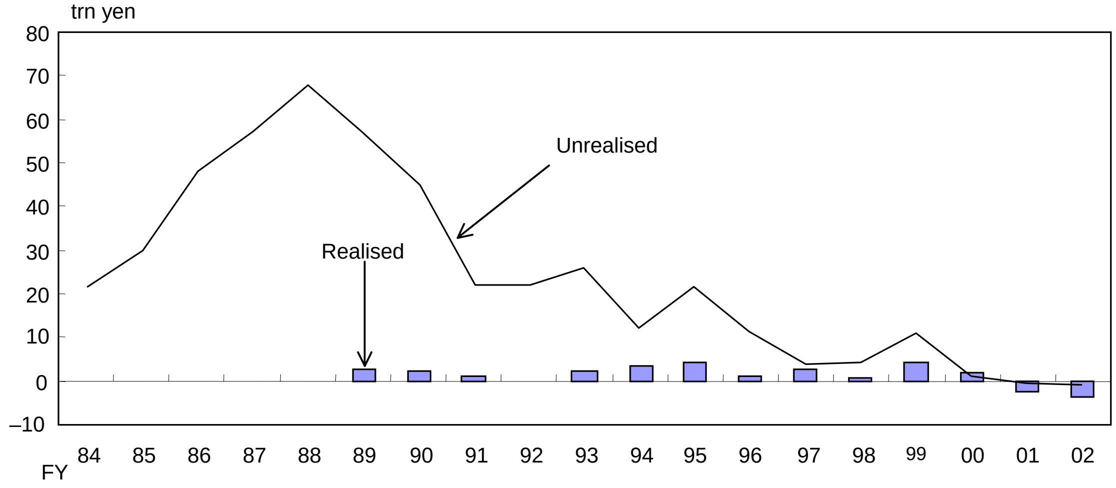


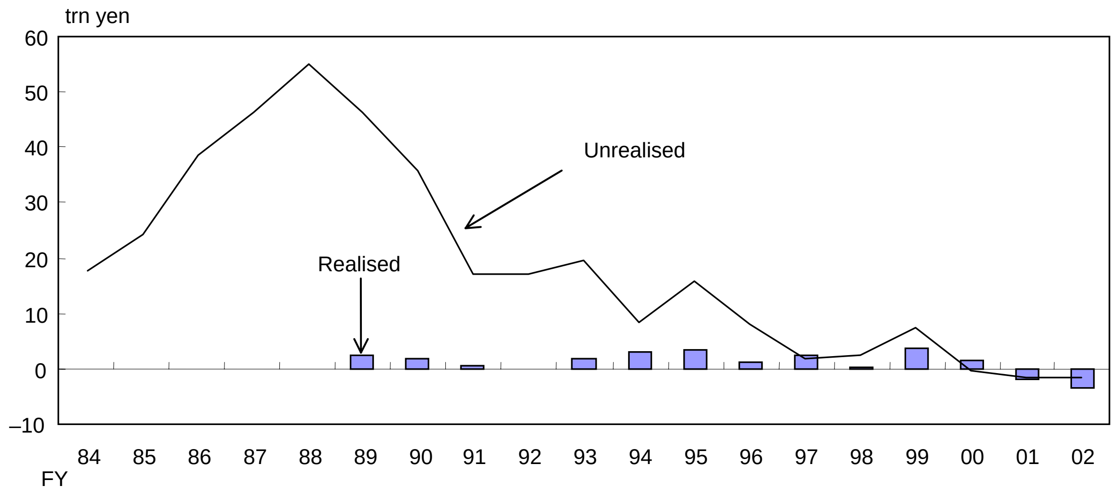
Figure 6

Stock-related gains/losses

(1) All banks



(2) Major banks



(3) Regional banks

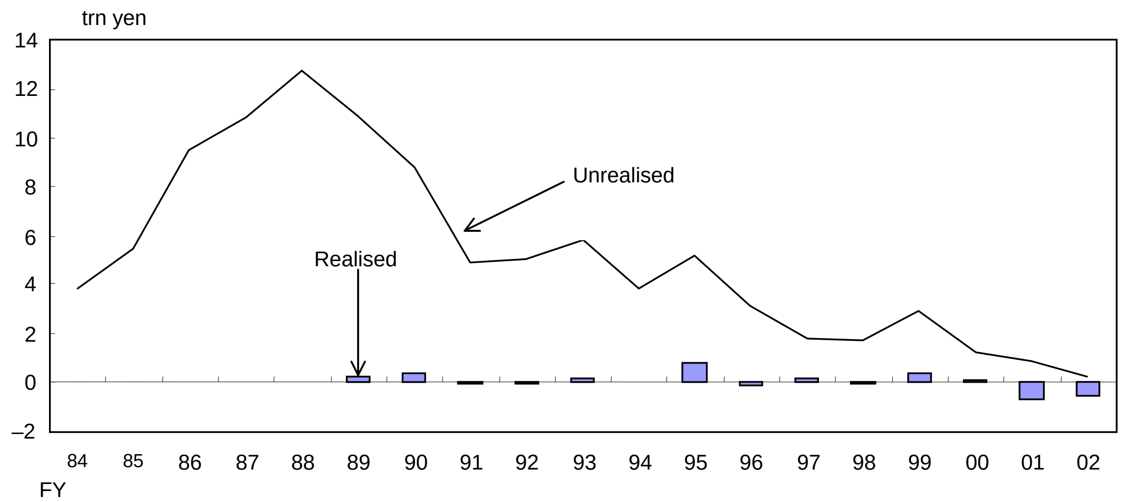
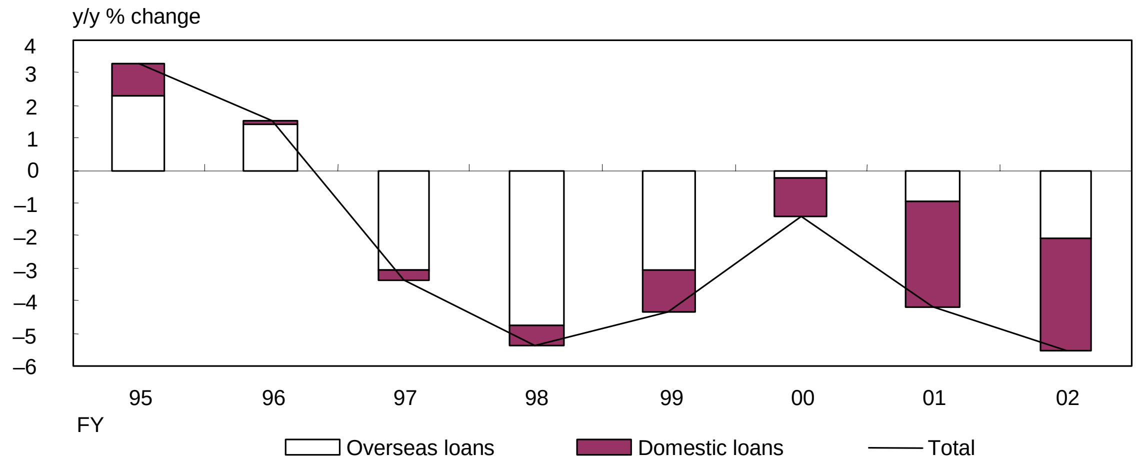


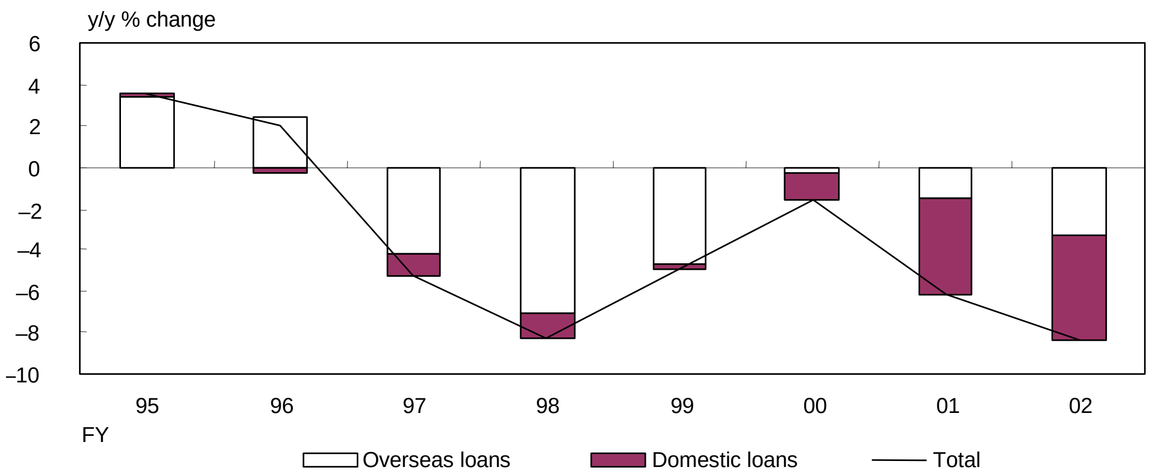
Figure 7

Changes in loans outstanding

(1) All banks



(2) Major banks



(3) Regional banks

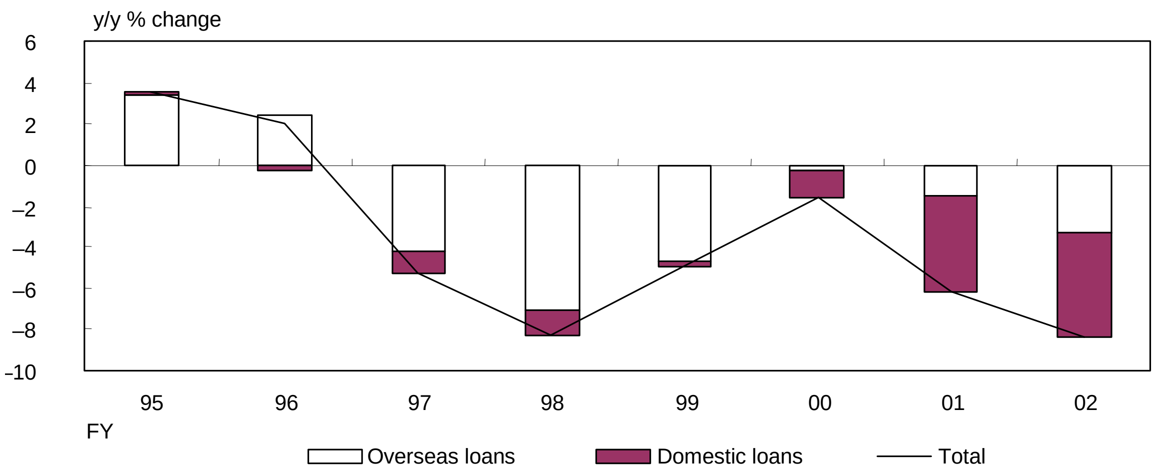


Figure 8-1

Estimation results for the optimisation model of bank behaviour

Equation: $\frac{FL_t}{D_{t-1}} - \beta \frac{FL_{t+1}}{D_t} = b_0 + b_1(r_{Lt+1} - r_{Ct+1}) + b_2 \frac{P_{Lt}}{P_{Lt-1}} - b_3 \kappa_t + u_{t+1}$

(1) Sample period: fiscal 1985-2001

Dependent variables	**Constant** b_0	**Loan-call rate spread** b_1	**Change in land prices** b_2	**Tier 1 ratio** b_3	**Discount factor (average** β**)**	**NPL ratio (average** α_3**)**	**Number of observations**
Major banks	–0.0122	0.5173	0.0085	0.2306	0.9656	0.0062	170
	(–0.602)	(0.980)	(0.198)	(0.652)			
Regional banks	–0.0106	0.0302	0.0100	0.0520	0.9656	0.0045	1,921
	(–0.807)	(1.539)	(0.989)	(0.987)			
(Reference) All banks	–0.0134	0.0225	0.0127	0.0625	0.9656	0.0064	2,091
	(–1.047)	(1.193)	(1.327)	(1.203)			

(2) Sample period: fiscal 1985-89

Dependent variables	**Constant** b_0	**Loan-call rate spread** b_1	**Change in land prices** b_2	**Tier 1 ratio** b_3	**Discount factor (average** β**)**	**NPL ratio (average** α_3**)**	**Number of observations**
Major banks	–0.5767 ***	0.1212	0.4482 ***	0.8022	0.9601	0.0033	50
	(–6.071)	(0.940)	(5.288)	(1.539)			
Regional banks	–0.1144 ***	0.0270	0.0889 ***	0.1074	0.9601	0.0023	565
	(–7.638)	(1.252)	(6.287)	(1.180)			
(Reference) All banks	–0.1285 ***	0.0224	0.0999 ***	0.1193	0.9601	0.0019	615
	(–8.675)	(1.077)	(7.295)	(1.309)			

Notes: *t*-values are in parentheses. *** 1% level. ** 5% level. * 10% level. NPL ratio (average α_3) is imputed with the estimated parameters, the average of discount factors and changes in land prices in the corresponding periods. See Appendix 1.

Figure 8-2

Estimation results for the optimisation model of bank behaviour (continued)

Equation: $\frac{FL_t}{D_{t-1}} - \beta \frac{FL_{t+1}}{D_t} = b_0 + b_1(r_{Lt+1} - r_{Ct+1}) + b_2 \frac{P_{Lt}}{P_{Lt-1}} - b_3 \kappa_t + u_{t+1}$

(3) Sample period: fiscal 1990-96

Dependent variables	**Constant** b_0	**Loan-call rate spread** b_1	**Change in land prices** b_2	**Tier 1 ratio** b_3	**Discount factor (average** β**)**	**NPL ratio (average** α_3**)**	**Number of observations**
Major banks	–0.0970 (–0.772)	1.5682 (1.340)	0.0891 (1.361)	0.8463 ** (2.111)	0.9633	0.0081	70
Regional banks	–0.0450 *** (–3.017)	0.0721 *** (3.304)	0.0473 *** (3.978)	0.0850 (1.606)	0.9633	0.0066	791
(Reference) All banks	–0.0445 *** (–2.927)	0.0632 *** (2.940)	0.0469 *** (4.025)	0.0829 (1.578)	0.9633	0.0070	861

(4) Sample period: fiscal 1997-2001

Dependent variables	**Constant** b_0	**Loan-call rate spread** b_1	**Change in land prices** b_2	**Tier 1 ratio** b_3	**Discount factor (average** β**)**	**NPL ratio (average** α_3**)**	**Number of observations**
Major banks	–0.6070 (–1.278)	2.6463 ** (2.499)	0.5959 (1.069)	1.6230 *** (3.386)	0.9743	0.0150	50
Regional banks	–0.4203 *** (–6.872)	0.0804 (1.540)	0.4413 *** (7.026)	0.1113 * (1.849)	0.9743	0.0116	565
(Reference) All banks	–0.4083 *** (–6.770)	0.0855 (1.640)	0.4284 *** (6.945)	0.1190 ** (1.993)	0.9743	0.0130	615

Notes: *t*-values are in parentheses. *** 1% level. ** 5% level. * 10% level. NPL ratio (average α_3) is imputed with the estimated parameters, the average of discount factors and changes in land prices in the corresponding periods. See Appendix 1.

Figure 9

Algorithm of simulation

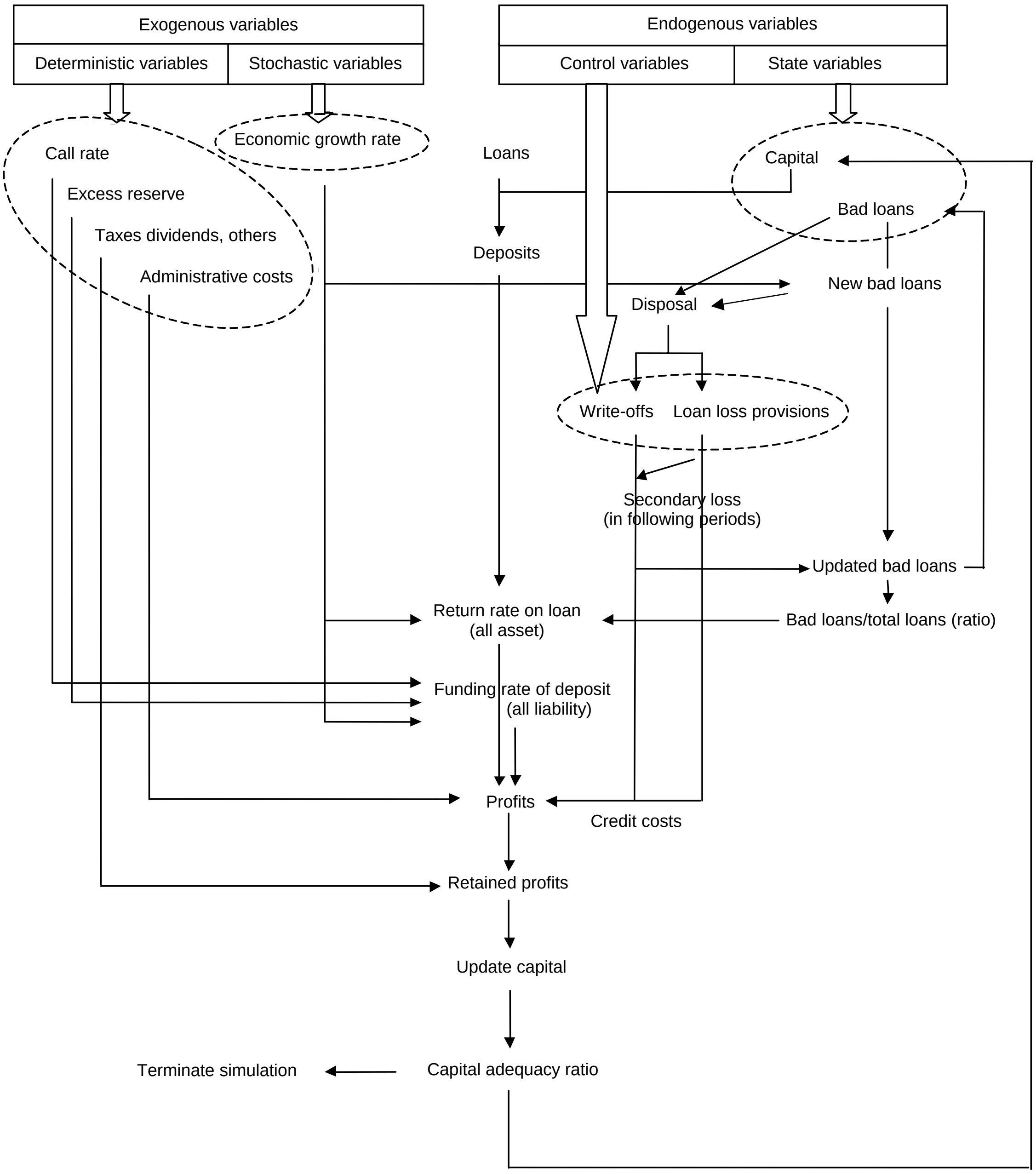


Figure 10

Four simulation cases

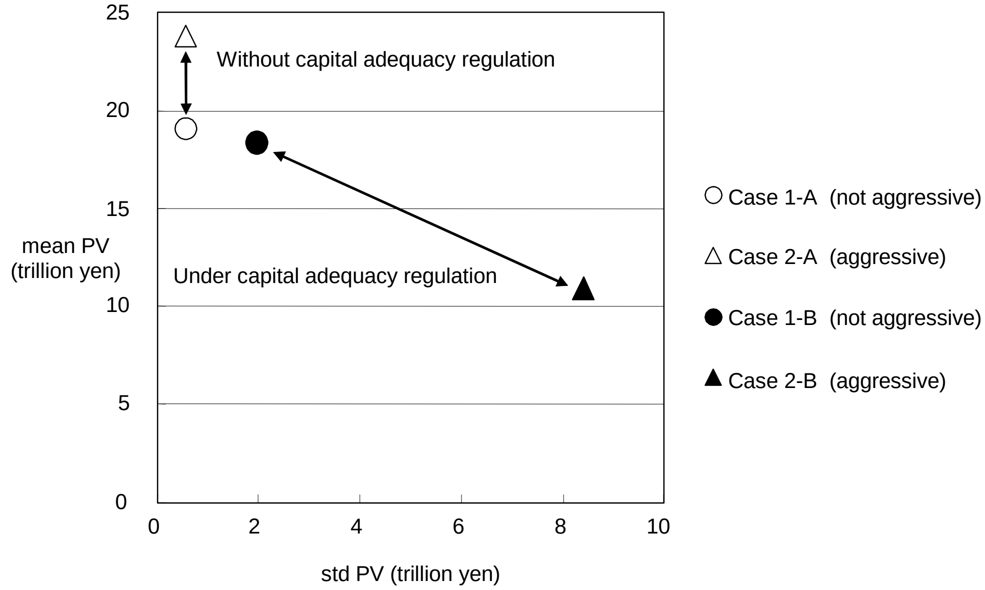
		Expectation of economic growth	
		Adaptive expectation	**Perfect forecast**
Aggressiveness in write-offs of bad loans	Not aggressive	Case 1	Case 3
	Aggressive	Case 2	Case 4

Note: Economic growth is measured by nominal GDP growth rate. Adaptive expectation is based on mean and standard deviation of the growth rates for the preceding six half-year periods. Perfect forecast case does not mean the bank's perfect forecast of the future *path* of the growth rate, but assumes that the bank can correctly predict its mean and standard deviation. See Appendix 2 for details on the amount of write-offs in the four cases.

Figure 11

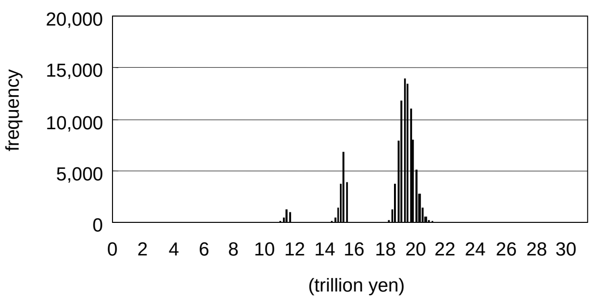
Mean and standard deviation of the bank's present value (1)

Beginning of fiscal 1997 as initial period

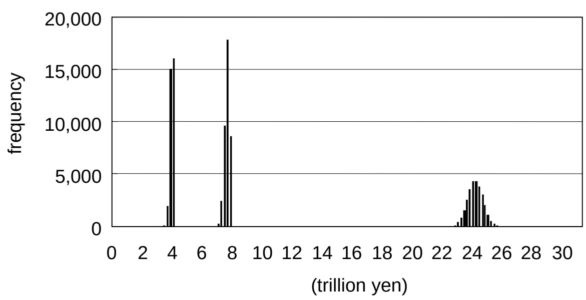


Distributions of present values

Case 1-B: ratio of termination = 49%



Case 2-B: ratio of termination = 72%

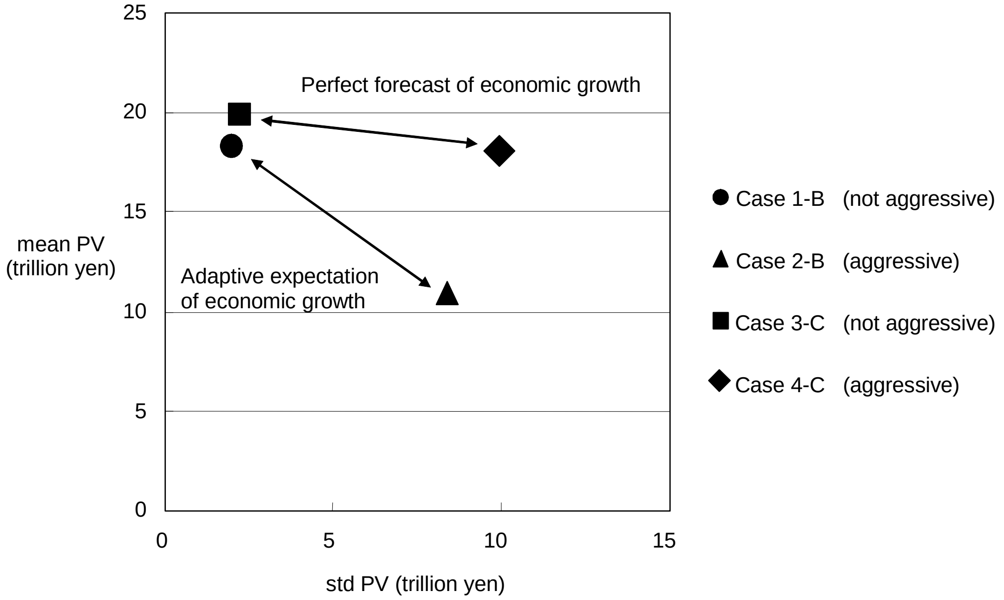


Note: Present values of profits are measured for next six half-year periods using unit subjective discount factor. See Appendix 2 for details.

Figure 12

Mean and standard deviation of the bank's present value (2)

Beginning of fiscal 1997 as initial period



Distributions of present values

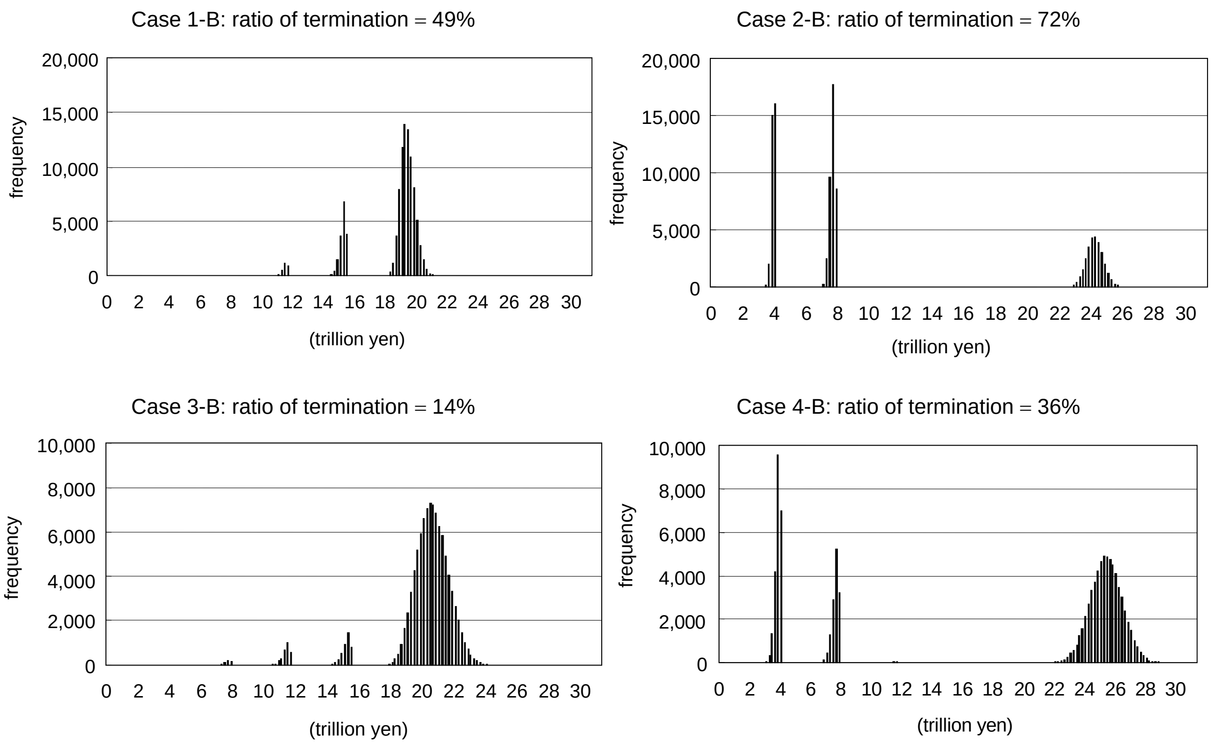
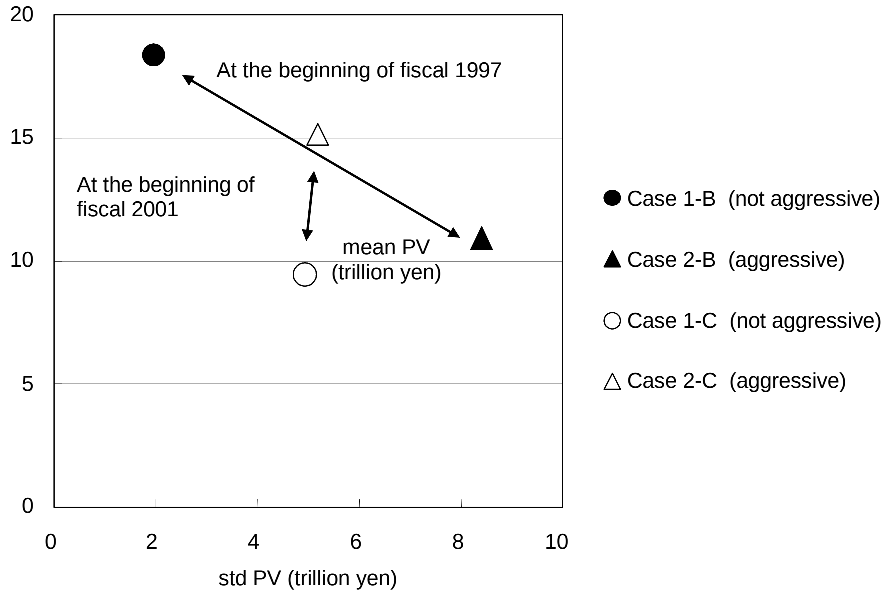
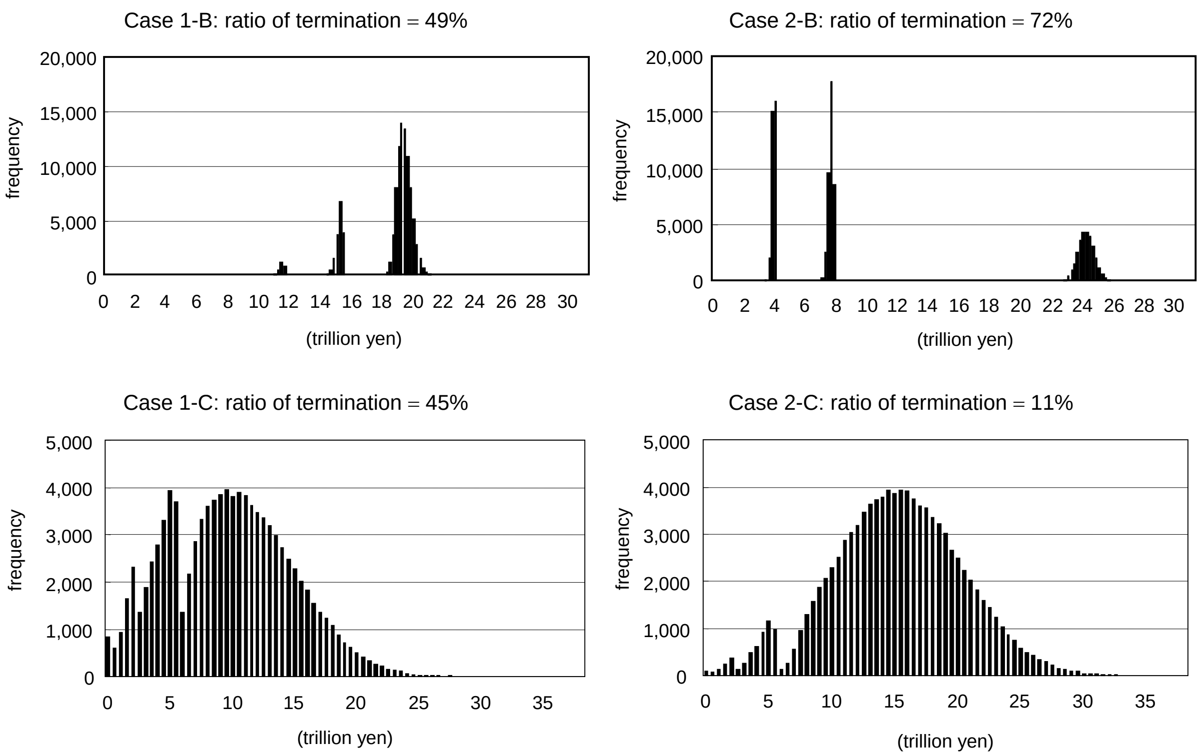


Figure 13

Mean and standard deviation of the bank's present value (3)



Distributions of present values



Appendix table

Simulation results: means through simulation periods (in trillions of yen)

(Termination ratio, %)	Growth rate (%) mean/std		Total return rate (%)	Total funding rate (%)	Capital	Credit cost	New bad loan	Bad loan	Out-standing LLP
Fiscal 1997, without capital adequacy regulation									
Case A1 (0)	1.0	0.5	1.9	1.1	27	3.5	4.1	24	11
Case A2 (0)	1.0	0.5	2.1	1.1	27	4.5	4.1	19	8
Fiscal 1997, under capital adequacy regulation									
Case B1 (49)	1.0	0.5	1.8	1.1	26	3.4	4.0	24	11
Case B2 (72)	1.0	0.5	1.2	0.7	16	2.7	2.4	13	6
Case B3 (14)	–0.1	1.0	1.7	0.9	27	3.6	4.5	25	11
Case B4 (36)	–0.1	1.0	1.5	0.7	22	0.4	3.6	16	7
Fiscal 2001, under capital adequacy regulation									
Case C1 (45)	–0.2	0.9	0.9	0.6	29	2.4	4.4	32	8
Case C2 (11)	–0.2	0.9	1.2	0.6	31	2.6	4.6	31	5

Note: Since “high termination ratio” cases produce zero values for all variables after termination, average values in the table tend to be lower than in “low termination ratio” cases.

References

Bank of Japan (2002): "Japan's nonperforming loan problem", *Bank of Japan Quarterly Bulletin*, 10 (4), pp 97-108.

——— (2003): "Review of Japanese banks' activities: profits and balance-sheets in fiscal 2002", *Bank of Japan Research Papers*, available from http://www.boj.or.jp.

Berger, A N and G F Udell (1995): "Relationship lending and lines of credit in small firm finance", *Journal of Business*, 68 (3), pp 351-81.

Elyasiani, E, K J Kopecky and D Van Hoose (1995): "Costs of adjustment, portfolio separation, and the dynamic behaviour of bank loans and deposits", *Journal of Money, Credit and Banking*, 27 (4), pp 955-74.

Ito, T and Y N Sasaki (1998): "Impact of the Basel capital standard on Japanese banks' behaviour", Institute of Economic Research, Hitotsubashi University, Discussion Paper Series, A-356.

Kashyap, A K and J C Stein (1997): "What do a million banks have to say about the transmission of monetary policy?", *NBER Working Papers*, no 6056.

Ogawa, K and S Kitasaka (2000): "Bank lending in Japan: its determinants and macroeconomic implications", The Institute of Social and Economic Research, Osaka University, Discussion Paper, no 505.

Oyama, T and T Shiratori (2001): "Insights into the low profitability of Japanese banks: some lessons from the analysis of trends in banks' margins", *Bank of Japan Research Papers*, available from http://www.boj.or.jp.

Zeldes, S P (1989): "Consumption and liquidity constraints: an empirical investigation", *Journal of Political Economy*, 97 (2), pp 305-46.

Non-performing loans and the real economy: Japan's experience

Nobuo Inaba, Takashi Kozu and Toshitaka Sekine,[1] Bank of Japan
Takashi Nagahata, London School of Economics

1. Introduction

Taking stock of a number of related studies[2] conducted within the Bank of Japan, our intention in this paper is to discuss the interrelationship between the increase in non-performing loans (NPLs) and the performance of the real economy in Japan since the 1990s.

Since the bursting of the asset price bubble in the early 1990s, NPL problems have been a central issue for researchers and policymakers in Japan. It is an issue that includes a whole range of topics, such as the extent of the NPLs residing on balance sheets in the financial sector; whether or not there was any credit crunch; how bank health should be restored, and whether this should involve injections of public funds; and the severity of the adjustment process - say, how far the already high unemployment rate would go up - over the course of restructuring.

Given this wide range of issues (and the limitations of space), we focus our attention on issues that relate directly to the interaction between NPLs and the real side of the economy.

Even within this narrower scope, our coverage in this paper is selective. We do not discuss, for instance, the increase in precautionary saving after the 1997-98 banking crises. This is not because the negative impact of these was negligibly small. Rather, it is because there is general agreement among economists concerning the huge cost associated with these banking crises. Our interest here, therefore, is in the more contentious issue of whether, when we abstract from these banking crises, there remains a significant link between NPLs and the real economy.

The remainder of the paper is organised as follows. Section 2 considers how the performance of the real economy affected the emergence of NPLs. Sections 3 to 5 then discuss how the increase in NPLs, in turn, distorted the performance of the real economy via malfunctioning in the banking sector. Finally, Section 6 concludes the paper.

2. Emergence of NPLs

2.1 Definitions of NPLs

First of all, we briefly review the definitions and recent status of NPLs.

For those who are not familiar with the NPL problems in Japan, definitions of NPLs have often been the source of confusion. This is because there are at least three definitions that are referred to, and these definitions have been changed over time (Figures 1 and 2).

- *Risk management loans* and *loans disclosed under the Financial Reconstruction Law* (*FRL*) *classification* are officially published NPLs in the sense that they are based on the criteria specified by a law or bylaw. Although they have different breakdowns, their two definitions

[1] We are grateful to Yumi Saita for her assistance and for allowing us to make use of the results of her research. The views expressed in this paper do not necessarily reflect those of the Bank of Japan.

[2] We leave technical details to background papers (Nagahata and Sekine (2002), Sekine et al (2003), Saita and Sekine (2001)).

broadly coincide, and hence produce similar figures for outstanding loans (¥34.8 trillion and ¥35.3 trillion, respectively, at end-March 2003).

- *Loans subject to self-assessment* are classified, depending upon borrower creditworthiness, in line with guidelines (the "Inspection Manual") produced by the Financial Services Agency (FSA):
  - Loans that, according to the terms of the *self-assessment*, are to "bankrupt" and "de facto bankrupt" borrowers correspond to "unrecoverable or valueless" loans under the *FRL classification*, while those to borrowers that the *self-assessment* classifies as "in danger of bankruptcy" correspond to "risk" loans under the *FRL classification*.
  - Loans to borrowers classified in the self-assessment as needing "attention" include a subcategory of loans to borrowers needing "special attention". Loans to borrowers that "need special attention" roughly correspond to loans requiring "special attention" under the *FRL classification*.[3] Since the figure for loans to borrowers that "need attention" but not "special attention" is substantial, outstanding loans to borrowers whom the *self-assessment* categorises as of or below the standard of needing "attention" (¥90.1 trillion at the end of March 2003) far exceed the apparently comparable figures for *risk management loans* and *FRL classified loans*.
- These definitions have substantially changed over time. As summarised in Figure 3, the criteria became tougher and their coverage became wider in response to public demand for better disclosure.

In this paper, in order to avoid ambiguity, when we refer to NPLs we are talking about *risk management loans* and *FRL classified loans*. As explained above, these broadly correspond to loans to borrowers classified in the *self-assessment* as being of or below the standard of needing "special attention". We consider loans to borrowers that "need attention" but not "special attention" to be quasi-NPLs.

In what follows, we define borrower **firms' ratings** by reclassifying the *self-assessment* ratings to get (i) "normal" borrowers (these remain the same as in the *self-assessment*); (ii) "doubtful" borrowers (those classified within the *self-assessment* as needing "attention" but not "special attention"); and (iii) "bad" borrowers (those who "need special attention", or are "in danger of bankruptcy", "de facto bankrupt", or "bankrupt" according to the *self-assessment* ratings). As described above, this category of "bad" borrowers basically captures NPLs, while "doubtful" borrowers correspond to quasi-NPLs.

Although declining, NPLs remain high. Under the current government initiatives, banks are required to dispose of loans that fall into or below the category "in danger of bankruptcy" within three years of their emergence.[4] By active sales of their NPLs (including to the Resolution and Collection Corporation) and debt forgiveness at times of corporate restructuring, banks had decreased their *risk management loans* in March 2003 by more than ¥8 trillion from a year earlier (Bank of Japan (2003a)). However, the NPL ratio (*FRL classified loans* divided by total loans outstanding) of major banks in March 2003 was 7.2%, which was still significantly higher than 4%, the target ratio to be achieved by March 2005 (FSA, "Program for Financial Revival", October 2002).

2.2 Link from the real economy to NPLs

So far, despite active debate in the media, there is little empirical research available relating the performance of the real economy to the emergence of NPLs. Some argue that it is the long-lasting recession that has been responsible for the increase in NPLs. Others appeal to the debt-deflation theory of Irving Fisher (1933) and insist on deflation (in the sense of a decline in general prices) as the

---

[3] To be precise, the figure for loans to borrowers whom the *self-assessment* determines as requiring special attention is larger than the comparable figure for *FRL classified* special attention loans. This is because the former counts loans to borrowers in their entirety, even if only part of these borrowings requires special attention.

[4] "Emergency Economic Package", April 2001. Banks are also required to dispose of 50% of these loans within one year of their emergence and about 80% within two years (FSA, "Measures for a Stronger Financial System", April 2002).

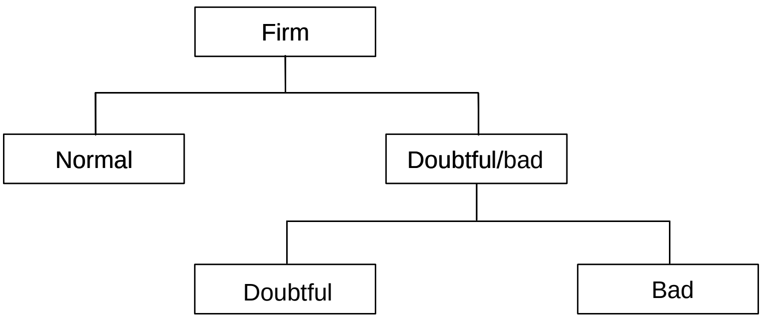
prime cause. The lack of adequate empirical research prevents economists from reaching any consensus on this issue.

However, it seems obvious that the sharp fall in asset prices, especially land prices,[5] is one of the dominant causes of NPLs. *Risk management loans* are heavily concentrated in real estate related industries, ie in construction and real estate, as well as among retailers and wholesalers (Figure 4).[6] During the bubble era of the late 1980s, firms in these industries were aggressive in their purchases of real estate properties, including countryside forests in order to develop then-lucrative resort areas such as golf courses (Figure 5).[7] The collapse of land prices after the bursting of the bubble severely impaired their balance sheets and made some of them insolvent.

In order to further investigate this issue, we believe that we need to exploit cross-sectional information on individual firms such as borrower firms' ratings (good/doubtful/bad) and their financial condition. For instance, the following calibration would create consistent data for NPLs and enable us to see the effects of the real economy. As seen in the previous section, due to frequent changes in the definitions, there is no such time series, which at least partly explains the scarcity of empirical research on this issue.

(i) First, a cross section model of individual borrower firms' current ratings is estimated by regressing them on various financial indicators obtained from their income statements and balance sheets.

More specifically, as such a cross section model, we believe it promising to use a nested logit model whose tree structure is as described below. A nested logit model is desirable because it is expected to fit the actual ratings better than an ordered probit model, which is the alternative often used in the related literature. Improved fit is likely to be achieved courtesy of one of the advantages of the nested logit model, namely that we can use different sets of explanatory variables for each nest, ie the explanatory variables for the choice between "normal" and "doubtful/bad" could differ from those used when looking at the choice between "doubtful" and "bad".



(ii) Then, individual borrower firms' ratings in the past are calibrated, using the coefficients obtained in the above step and historical data on selected explanatory variables.

Provided that the estimated nested logit model offers a reasonable fit, the calibration gives us an insight into the ratings firms would have received had they been subject to the recent borrower classification criteria. In providing such ratings, the calibration creates consistent data for NPLs, where "consistent" means NPLs are classified according to the same criteria.

[5] Land price indices in the Tokyo metropolitan area have fallen to 40-50% of their 1992 levels see Figure 13. In fact, hedonic estimation of judicial auction prices reveals that the price of land used as collateral for NPLs has fallen even more sharply (Saita (2003)).

[6] In one of the few pieces of empirical research on this subject, Ueda (2000) finds a significant correlation between the NPL ratios of individual banks and the fluctuation of land prices in the capital cities of prefectures where banks' headquarters are located.

[7] Trading houses belong to the retail and wholesale industries. In addition to various goods and services, they are also known to deal actively in real estate properties. See Tachibana and Sekine (2003) on how to estimate land investment carried out by these industries.

Note that such data in the longer term do not exist in reality, because, as seen in Figure 3, aggregated figures for self-assessments are only available from 1997, and there have even been changes since then, with the criteria for self-assessment said to have changed when the "Inspection Manual" was introduced.

As a very preliminary stage of research, we have estimated the cross section model using the most recently available data and calibrated borrower firms' ratings in the 1990s in the way described above. We find that, in the nested logit model, the choice between "normal" and "doubtful/bad" mostly depends on procyclical variables obtained from income statements (eg sales growth, the interest coverage ratio), while the choice between "doubtful" and "bad" mostly depends on non-cyclical variables obtained from balance sheets (eg the debt/asset ratio, which mainly reflects land price developments because the asset values that constitute its denominator are revalued at market prices). As a result, the share of the calibrated numbers of "doubtful" is characterised by cyclical fluctuation resembling business cycles in Japan, while the share of "bad" borrowers steadily increases somewhat. Calculating transition matrices with the calibrated borrower ratings, it turns out that the matrix in recessionary periods significantly differs from that in expansionary periods.[8]

In sum, we tentatively conclude that two different real factors are responsible for the increase in NPLs. One is a trend factor, which directly affected the numbers of "bad" borrowers. The other is a cyclical factor, which acted to increase NPLs indirectly, by increasing quasi-NPLs. Given that (i) NPLs have been concentrated in real estate related industries, and (ii) the choice between "doubtful" and "bad" in the nested logit model is dependent upon balance sheet variables like the debt/asset ratio, the trend factor is thought to be associated with the deterioration in firms' balance sheets that accompanied the fall in land prices. Meanwhile, the fall in land prices is thought to have reflected the bursting of the bubble as well as ongoing structural changes.

3. Firms' balance sheet condition vs banks' balance sheet condition

In Sections 3 and 4, we discuss how the increase in NPLs affected real economic activity in Japan. First, in Section 3, we examine the respective roles played by firms' and banks' balance sheet condition in determining firm investment and bank lending. Then, in Section 4, we consider another problem associated with NPLs, namely forbearance lending.

It is an issue of some contention among economists whether or not banks, faced with a deterioration in their balance sheet condition, restrained their lending and so hampered investment. Theoretically, as pointed out by Krugman (1998), banks with damaged balance sheets might have an incentive to favour risky projects - this is known as "gambling for resurrection". In opposition to this, Van den Heuvel (2001) shows how a bank with an impaired balance sheet might decrease its lending in order to satisfy the risk-based capital requirements of the Basel Accord. There is also an empirical difficulty in distinguishing the respective roles played by firms' and banks' balance sheet condition. Identification of distinct roles for each is problematic because, at the macro level, firms' balance sheets and banks' balance sheets are different sides of the same coin.

In order to overcome this empirical difficulty, we rely on micro panel data. At the level of a diversified micro data set we can distinguish between the roles of firms' and banks' balance sheets, provided that there is a sufficient number of firms whose own balance sheets are in good condition but whose main banks' balance sheets are not, and vice versa.

The basic strategy below is to augment conventional forms of firm investment and bank lending functions with variables that represent firms' and banks' balance sheet condition, and then to check whether coefficients on these variables are significant.[9]

---

8 Using Moody's and S&P's data, Nickell et al (2000) and Bangia et al (2002) find that the transition matrices differ depending on whether a given business cycle period is expansionary or recessionary.

9 See Chatelain et al (2001) and Ehrmann et al (2001) for recent examples of firm investment functions and bank lending functions using micro panel data.

3.1 Firm investment

We estimate the following error-correction specification of a firm investment function, using micro panel data on 1,078 listed firms:

$$\left(\frac{I_{it}}{K_{i,t-1}}\right) = \rho\left(\frac{I_{i,t-1}}{K_{i,t-2}}\right) + \sum_{h=0}^{1}\beta_h \Delta y_{i,t-h-1} + \sum_{h=0}^{1}\gamma_h \Delta j_{i,t-h-1} + \lambda_0 (k-y)_{i,t-2} + \lambda_1 y_{i,t-2} + \lambda_2 j_{i,t-2}$$

$$+ \theta CF_{it} + \phi\left(\frac{D}{A}\right)_{i,t-1} + \varphi Cap_{it} + u_{it}$$

where I_{it} is the real investment of firm i at time t; K_{it} is its real capital (small k denotes its logarithm); y_{it} is the log of its real output; j_{it} is the log of its user cost of capital; CF_{it} is its cash flow divided by its nominal capital; and u_{it} is an error term. Δ denotes the first difference operator.

Firms' and banks' balance sheet condition is represented by the following variables. First, each firm's balance sheet condition is captured by its debt/asset ratio, *D*/*A*, where assets *A* are revalued at market prices using the perpetual inventory method. Then, each bank's balance sheet condition is captured by an adjusted capital adequacy ratio, *Cap*, which takes into account NPLs, capital gains/losses and deferred tax assets.[10] For each firm, *Cap* is calculated as a weighted average of its main banks' *Cap*, where the weights represent the main banks' shares of long-term loans. Main banks are defined to be the three city/long-term credit banks whose long-term loans are the largest.

Following Gibson (1997), we split our sample into two subsamples according to whether or not firms have ever issued bonds. Non-bond-issuing firms are supposed to face tighter financial constraints than bond-issuing firms, because they have fewer external funding options and are hence more dependent on bank lending.

Figure 6 summarises the estimation results. Insignificance of the cash flow terms aside, signs and sizes of estimated coefficients are largely in line with prior expectations.[11] The points to be noted are:

- Firms' balance sheet condition, *D*/*A*, is negative and significant for both bond-issuing and non-bond-issuing firms; while
- Banks' balance sheet condition, *Cap*, is positive and significant only for non-bond-issuing firms.

What this implies is that, after the collapse of the asset price bubble, firms restrained their investment in order to reduce the burden of existing debts. Moreover, it indicates that, faced with erosion of their capital adequacy, banks restrained their lending and hence hampered the investment of firms without access to the capital market. This finding is consistent with the story of a "credit crunch".

In Figure 7, contributions to changes in I/K_{-1} are calculated from the sample averages of the variables of interest (*D*/*A* and *Cap*) and their coefficients. Firms' balance sheet condition is found to have had a relatively large negative impact throughout the 1990s. Meanwhile, the negative impact of banks' balance sheet condition is particularly large for non-bond-issuing firms during the FY1996-98 subperiod, which spans the occurrence of the banking crises in Japan. However, even prior to that subperiod, a non-negligible negative impact is observed for non-bond-issuing firms.

In short, NPLs hampered firm investment via a deterioration in both firms' and banks' balance sheet condition.[12] In a sense, the deterioration in banks' balance sheet condition may be said to have had a propagation effect, because it distorted the investment of bank-dependent firms, even when the balance sheets of the latter were in good condition.

---

[10] (Shareholders' equity + capital gains/losses from securities + loan-loss provisioning – risk management assets – deferred tax assets)/assets.

[11] See Nagahata and Sekine (2002) for a discussion of the insignificant cash flow terms. Also, see Bank of Japan, Research and Statistics Department (2003) for a more general exposition of weak business fixed investment in the 1990s.

[12] Ogawa (2001) and Sekine (1999) also find both firms' and banks' balance sheet condition mattered for firm investment.

3.2 Bank lending

In order to check the robustness of the above findings, it would be desirable to see whether a similar story holds for bank lending. In the context of a putative credit crunch, it is important to check whether deteriorating bank balance sheets acted to reduce bank lending, something which is not directly observed in the above estimation.

The role of bank balance sheets can be checked by estimating the following reduced-form bank lending function using micro panel data on banks:

$$\Delta l_{it} = \sum_{j=1}^{3} \gamma_j \Delta l_{i,t-j} + \mu \Delta d_{it} + \sum_{j=0}^{3} \theta_j \Delta r_{t-j} + \sum_{j=0}^{3} \rho_j DI_{i,t-j} + \sum_{j=0}^{4} \lambda_j \left(\frac{D}{A}\right)_{i,t-j} + \varphi Liq_{i,t-1} + \phi Cap_{i,t-1} + u_{it}$$

where l_{it} is the log of outstanding loans of bank i at time t; d_{it} is the log of outstanding deposits; r_t is the short-term interest rate; DI_{it} is the diffusion index of business conditions in the survey conducted on firms; and Liq_{it} is the bank's liquidity ratio.[13] Again, to capture firms' and banks' balance sheet condition, firms' debt/asset ratios $(D/A)_{it}$ and banks' adjusted capital adequacy ratios Cap_{it} are included. Both DI_{it} and $(D/A)_{it}$ for each bank are obtained as weighted averages of D/A and DI at the industry level, where the weights are the industry shares of outstanding loans at each bank.

We have preliminarily estimated the above function using data on individual banks in the 1990s. All the long-run coefficients are found to be significant and have expected signs - thus that on D/A is negative and significant and that on Cap is positive and significant. Calculating contributions to annual growth in bank lending, using the sample averages of variables, movements in both D/A and Cap make large negative contributions to bank lending, just as in the investment function above.[14]

These results are in line with the findings for firm investment, in that both firms' and banks' balance sheet condition matters.

3.3 Implications for monetary policy

In the above firm investment function, the coefficients on the interest rates are negative and significant. For instance, in Figure 6, most of the user costs of capital (j is its level and Δj is its first difference), which are calculated from the yield on 10-year JGBs, are negative and significant. Also in the above bank lending function, the change in short-term interest rate Δr_t is found to be negative and significant.

These findings imply that a conventional transmission channel was working even after the bubble burst. This is contrary to the widely held belief that monetary policy was largely ineffective - a belief borne out, for example, by the simple correlation between changes in loans outstanding and the call rate, which turned out to be positive after the bubble burst (Figure 8). Our finding suggests, however, that the positive impact of lowering the interest rate was obscured by the negative impact of the deterioration in firms' and banks' balance sheet condition.

In order to further investigate the issue, we can re-estimate the bank lending function by splitting the sample period into two subperiods, say at 1997 Q4. The latter subperiod contains the introduction of both the zero interest rate policy (1999 Q1) and the quantitative easing policy (2001 Q1).

We preliminarily find that the coefficient on the short-term interest rate is negative and significant in the former subperiod but it turns out to be positive and significant in the latter subperiod (ie interest rate cuts acted to decrease bank lending). Even when we replace the short-term interest rate with a quantitative measure such as base money, the coefficient on base money is negative and significant in the latter subperiod (ie increases in base money act to decrease bank lending).

---

[13] (Cash and deposits + call loans + government securities)/debts outstanding.

[14] Explanatory variables in levels (*DI*, *D/A*, *Liq* and *Cap*) are subject to normalisation. This makes use of either their historical averages or of constant terms obtained from regressions on other macro variables such as the real growth rate.

Although we cannot dismiss the possibility that the wrong signs are due to some misspecification, this result coincides with Kimura et al (2003) and Fujiwara (2003), who also fail to find theoretically consistent monetary policy effects in recent years.

4. Forbearance lending

Recently economists have been paying more attention to another phenomenon associated with NPLs, namely "forbearance lending" (or what Peek and Rosengren (2003) term "ever-greening policy" and Caballero et al (2003) term "zombie lending"). Japanese banks are said to have been reluctant to write off NPLs and to have rolled over their lending, even in cases where there was little prospect of the borrower firm being able to repay the loans extended.

There are several theoretical models which try to reveal why or under what conditions banks have an incentive to engage in forbearance lending. In reality, some or all of these models may well be thought to hold at the same time. This is because, as seen below, they are not mutually inconsistent.

- Kobayashi and Kato (2001), along somewhat similar lines to Krugman (1998), argue that a change in banks' risk preferences makes them softer about providing additional loans. Once a bank increases its exposure to a firm, the bank becomes risk-loving and begins to control that firm as if it were a dominant shareholder.
- Sakuragawa (2002) develops a model in which a bank without sufficient loan loss provisioning has an incentive to disguise its true balance sheet so as to satisfy the minimum capital requirement.
- Berglöf and Roland (1997), applying a soft budget constraint model, consider a game between a bank and a firm in which the bank continues to provide loans to the firm even after the latter's liquidation value plunges following a decrease in asset prices.
- Baba (2001), using real option theory, shows that uncertainties associated with the write-off of NPLs - such as the reinvestment return from freeing up funds by write-off, the liquidation loss, and the possible implementation of a government subsidy scheme, etc - induce banks to delay writing off NPLs.

In order to see whether banks have been engaging in forbearance lending, we investigate the relationship between firms' debt/asset ratios *D*/*A* and their outstanding loans. In a preliminary estimation of a cross section model in Section 2, we find that loans to firms with higher debt/asset ratios tend to become NPLs. If banks have indeed been engaging in forbearance lending, loans would have been apt to increase to firms whose debt/asset ratios were above a certain level.

More specifically, using micro panel data on 580 firms, we test the above inference by estimating the following function:

$$l_{it} = \alpha_0 l_{i,t-1} + \alpha_1 i_{it} + \alpha_2 \left(\frac{D}{A}\right)_{i,t-1} + \alpha_3 \left(\frac{D}{A}\right)^2_{i,t-1} + \alpha_4 ROA_{i,t-1} + \alpha_5 + u_{it}$$

where i_{it} is the loan/deposit interest rate spread for firm *i* at time *t*; and ROA_{it} is the return on assets, which controls the firm's profitability.

If banks have been engaging in forbearance lending, α_2 would be negative and α_3 positive. That is, when *D*/*A* is small, banks would squeeze loans as *D*/*A* increases. However, when *D*/*A* exceeds a certain level, banks would start to squeeze loans less hard (or would conceivably even increase loans, if *D*/*A* were sufficiently large).

This turns out to have been the case for the construction and real estate industries after the bubble burst (Figure 9). In the subperiod from FY1993-99, the coefficient on the squared debt/asset ratio is positive and significant for the construction and real estate industries, which make up a large share of NPLs (Figure 4). This supports the view that banks provided forbearance loans to firms in these industries.

Forbearance lending is supposed to suppress the profitability of Japan's economy by bailing out inefficient firms producing poor returns. Moreover, the theory suggests that not only do inefficient firms survive, but they also reduce their levels of effort since they anticipate that banks will bail them out (Berglöf and Roland (1997)).

In the construction and real estate industries, firms with higher debt/asset ratios or faster loan growth are likely to have lower *ROA*. In Figure 10, *ROA* is regressed on a cross term comprising loan growth and the debt/asset ratio as follows:

$$ROA_{it} = \gamma_1 ROA_{i,t-1} + \gamma_2 \Delta l \cdot \left(\frac{D}{A}\right)_{i,t-1} + \gamma_3 \Delta share_{it} + \gamma_4 + u_{it}$$

where $share_{it}$ denotes firm i's share of its industry sales. The coefficient on the cross term is negative and significant for the construction and real estate industries, to which banks are supposed to have provided forbearance loans. This seems to indicate the presence of moral hazard among these firms, in the sense of Berglöf and Roland (1997).

As long as banks continue to provide forbearance loans and do not dispose of their NPLs, the quality of their loan portfolios will decline and they themselves will remain vulnerable.

5. Inefficient resource allocation

So far, we have observed the reluctance of banks to extend credit to potentially profitable firms, thus hindering the emergence of more efficient firms (Section 3); and also their reluctance to write off bad loans to non-profitable firms, thus securing the survival of inefficient firms (Section 4). Although at first sight these two phenomena look quite different, in that one involves failing to *expand* credit whereas the other involves failure to *shrink* credit, both have the same effect: they prevent credit from shifting to relatively efficient sectors. In other words, both the credit crunch and forbearance lending are symptoms of the malfunctioning Japanese banking sector.

In what follows, we provide evidence which supports the view that financial intermediation has indeed been weakened since the bubble burst.

5.1 *Tankan* survey

Figure 11 offers evidence from the *Tankan*'s Diffusion Index of lending attitudes at financial institutions. The horizontal axis describes the share of firms replying that lending attitudes are "severe", while the vertical axis gives the share of firms replying that they are "accommodative". Under normal circumstances, we expect the trade-off between the two shares to trace out a curve running from southeast to northwest.

Weakening financial intermediation should be captured in this setting by a northward shift of the curve, since the share of firms replying "severe" would not decline even in the face of monetary easing. In an analogy with the Beveridge curve for the labour market, an outward shift of the curve implies less efficient financial intermediation.

In fact, there was an apparent northward shift in the curve in the early 1990s. Since then, the curve has not shifted back. This indicates a weakening of financial intermediation around the middle of the 1990s.

5.2 Sectoral credit shifts

In order to confirm the above result, we use the following measure to capture credit shifts across sectors:[15]

[15] In fact, the idea of the sectoral shift measure comes from Lilien (1982), who calculates a measure of sectoral labour shifts. Lilien uses this measure as a proxy for the size of sectoral shocks.

$$\sigma_t^L = \left[\sum_{i=1}^{n} \left(\frac{L_{it}}{L_t} \right) (\Delta_4 \ln L_{it} - \Delta_4 \ln L_t)^2 \right]^{1/2}$$

where L_{it} is outstanding loans to industry i at time t and L_t denotes aggregate outstanding loans at time t $(L_t = \sum_i L_{it})$, and Δ_4 is the fourth-order difference operator.

When a large amount of credit is reallocated from one industry to another, σ^L is expected to increase. This is because such a reallocation would be expected to increase the dispersion of credit growth across sectors, implying a greater difference between $\Delta_4 \ln L_{it}$ and $\Delta_4 \ln L_t$.

In fact, σ^L declined significantly from the 1980s to the 1990s (Figure 12). Given that sectoral shocks increased during the 1990s, as illustrated by another Lilien-type measure based on sectoral job vacancies (Osawa et al (2002)), this decline in the sectoral credit shift measure indicates the inefficiency of resource allocation through financial intermediation.

6. Conclusion

In this paper, we have taken stock of related research carried out within the Bank of Japan in order to discuss the interrelationship between the increase in NPLs and real economic performance in Japan since the 1990s. The main points can be summarised as follows:

- The deterioration in firms' balance sheets due to the collapse of land prices was responsible for the increase in NPLs. Cyclical downturns seemed to be also responsible, albeit indirectly, in that they increased quasi-NPLs.
- The increase in NPLs, in its turn, distorted real economic performance via malfunctioning in the banking sector. Both a "credit crunch" and "forbearance lending" took place, and these caused a decline, through the banking sector, in the efficiency of its resource allocation.

In tandem with the government, the Bank of Japan has endeavoured to restore bank health through bank supervision. Recent measures include its advocacy of the discounted cash flow) methodology for provisioning (Bank of Japan (2002, 2003b)), as well as the purchases of equities from the banking sector aimed at reducing banks' equity exposure and keeping it down at the level of their Tier 1 capital. The Bank has also made efforts to strengthen the monetary transmission mechanism. As part of its efforts in this direction, the Bank decided to purchase asset-backed securities.

As a next step, we believe that more research investigating the process of asset price deflation is warranted. The research reviewed in this paper gives us to understand that the fall in land prices was responsible for the increase in NPLs that ended up suppressing the real growth of Japan's economy. However, we do not know why land prices fell so far. Although the fall in land prices is generally thought to have reflected the bursting of the bubble as well as ongoing structural changes (eg rapid ageing, hollowing out, etc), we do not have any quantitative sense of the extent of each factor's contribution.

We also believe that more work is needed on banks' profitability, since bank health cannot be restored unless banks become reasonably profitable. Uncovering the causes of banks' currently low profitability is vital. The weakness of the real economy, excessive competition due to overbanking, competition from government financial institutions and problems in bank management are often cited as reasons for low profits, and sensible policymaking requires a clear ranking of the degree to which each of these is responsible.

Figure 1

Non-performing loan classifications in Japan

At end-March 2003, in trillions of yen

Risk management loans	
Loans to borrowers in bankruptcy	2.2
Past due loans	15.9
Loans in arrears by three months or more	0.5
Restructured loans	16.2
Total	34.8

Loans disclosed under the Financial Reconstruction Law	
Unrecoverable or valueless loans	5.7
Risk loans	13.0
Loans requiring special attention	16.6
Total	35.3

Loans subject to self-assessment	
Bankrupt, de facto bankrupt	5.7
In danger of bankruptcy	13.0
Need attention Need special attention	71.4
Normal	371.7
Sub total	90.1

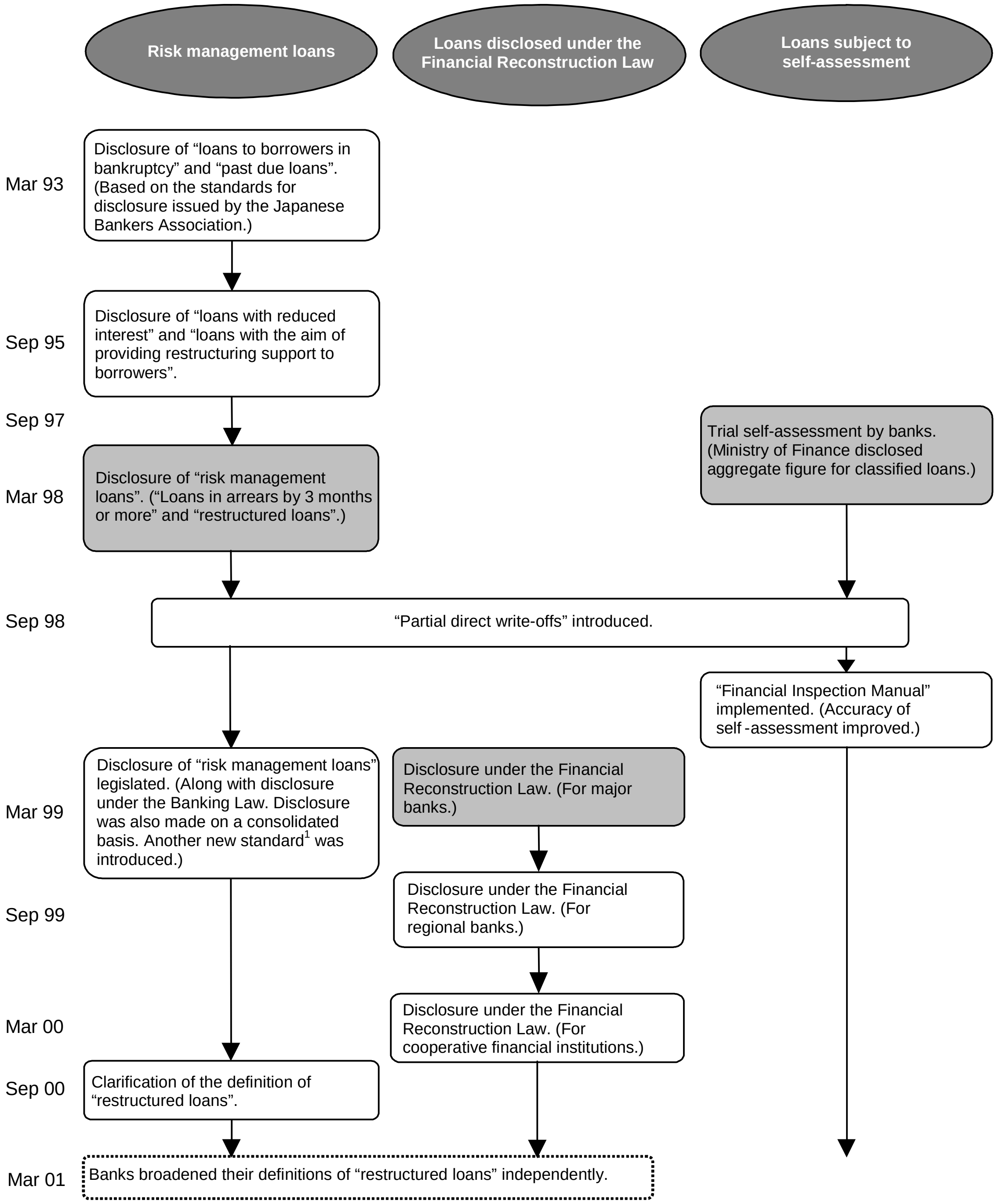
Figure 2

Definitions of non-performing loan classifications

(1) Risk management loans	
Loans to borrowers in bankruptcy	Loans where interest is not collected because borrowers are in bankruptcy.
Past due loans	Loans where interest is not collected, excluding those categorised above.
Loans in arrears by three months or more	Loans where principal or interest is in arrears by three months or more from the due date specified in the related loan agreement.
Restructured loans	Loans for which the bank has provided more favourable terms and conditions to the borrower than those in the original agreement, with the aim of providing restructuring support. These include reducing interest rates, rescheduling interest and principal payments, or waiving claims on the borrower.
(2) Loans disclosed under the Financial Reconstruction Law	
Bankrupt	Loans to borrowers who are legally or formally bankrupt, or virtually bankrupt borrowers with no prospects of resuscitation. (These correspond to loans categorised in the self-assessment as “bankrupt” and “de facto bankrupt”.)
De facto bankrupt	Loans to borrowers who have not gone bankrupt but are in financial difficulties, and thus whose lenders are unlikely to receive the principal and interest concerned on due dates. (They correspond to loans categorised in the self-assessment as “in danger of bankruptcy”.)
In danger of bankruptcy	“Loans in arrears by three months or more” and “restructured loans”. (The definitions of these are the same as under risk management loans.)
(3) Loans subject to self-assessment	
Bankrupt	Legally or formally bankrupt borrowers who are in the bankruptcy/liquidation process; who have filed for bankruptcy under the Commercial Law, the Corporation Reorganization Law or the Civil Rehabilitation Law; or whose deals are suspended at the clearing house.
De facto bankrupt	Borrowers who have serious financial difficulties with no prospect of resuscitation. Typically, they are seriously undercapitalised or have debt overdue for a long time. Although they are not legally or formally bankrupt, they are deemed bankrupt in practice.
In danger of bankruptcy	Borrowers who have financial difficulties and are likely to go bankrupt in the future. Typically, they are undercapitalised.
Need attention	Borrowers who have problems with interest payments or amortisation; or borrowers who record losses.
Need special attention	Borrowers all or part of whose debts are categorised as “loans requiring special attention” under FRL classified loans.
Normal	Borrowers who do not have particular problems.

Figure 3

Development of the NPL disclosure principles

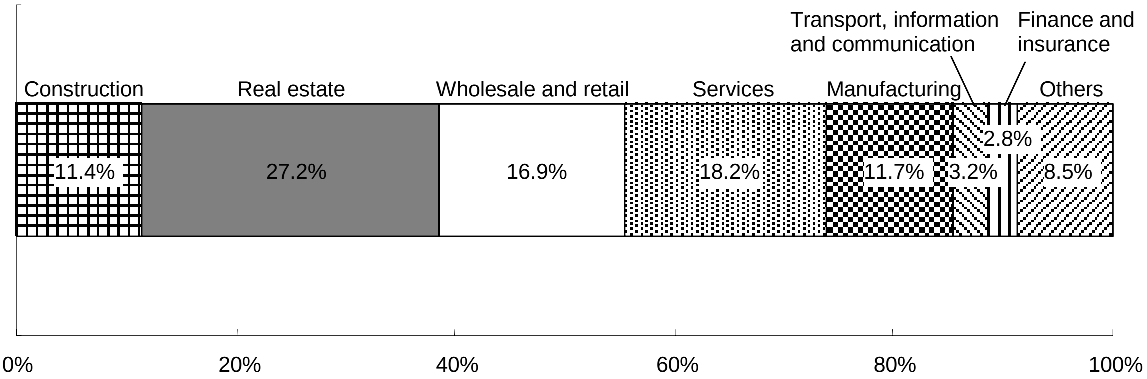


[1] Coverage of "past due loans" was extended, ie loans to borrowers "in danger of bankruptcy" must be included within "past due loans" even if they are not overdue.

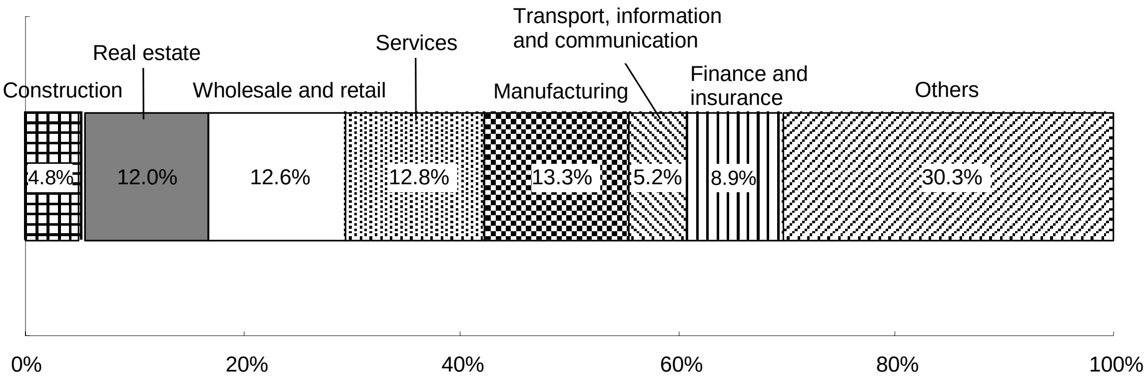
Figure 4

Breakdown of NPLs by industry

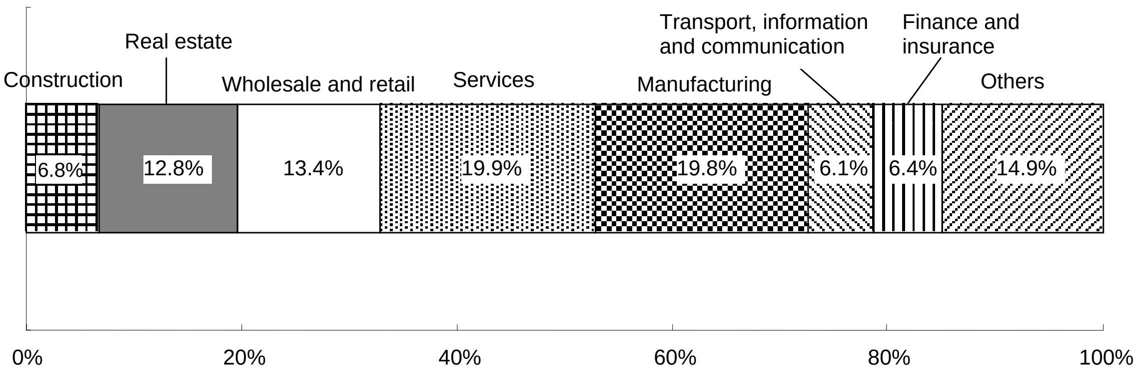
(1) Risk management loans



(2) Overall loans and discounts outstanding



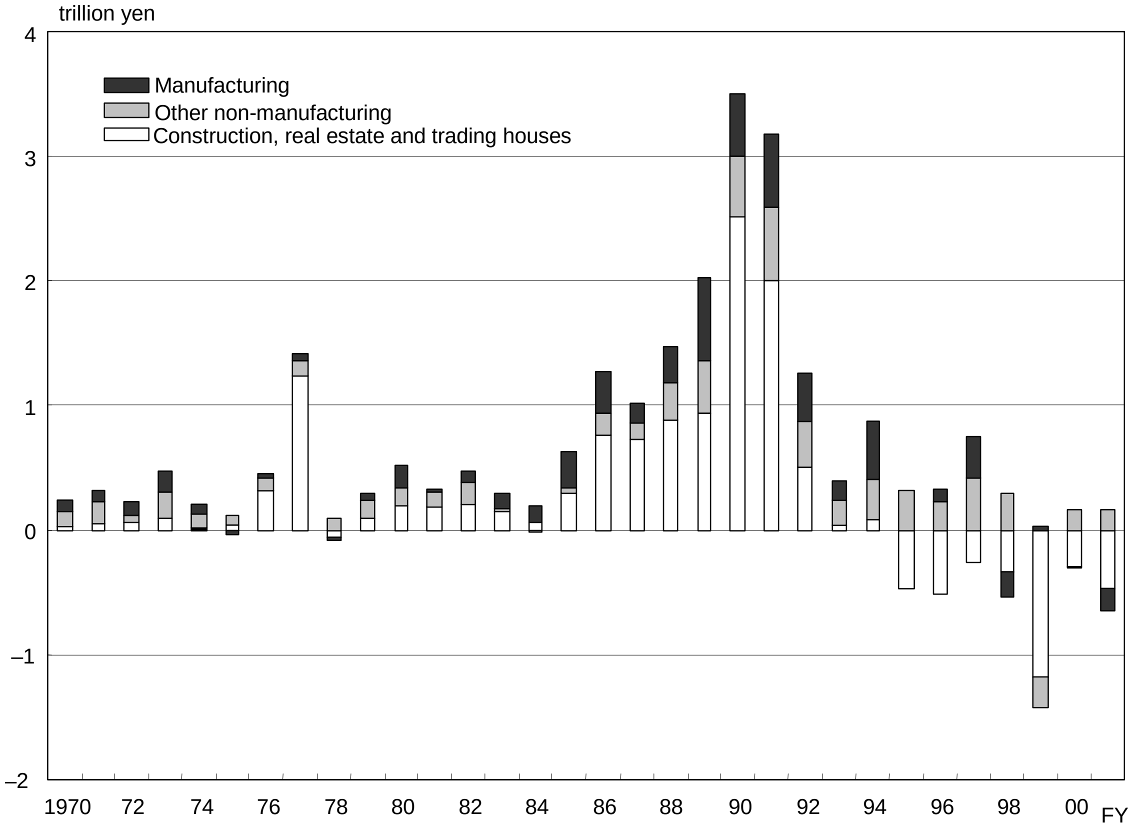
(3) Gross domestic product



Notes: 1. Risk management loans and overall loans and discounts outstanding are as of March 2003. Gross domestic product is as of FY2001. 2. Risk management loans are those disclosed by 13 major banks, ie city banks, long-term credit banks and trust banks, and 73 regional banks. They are based on banking accounts and trust accounts of domestic branches; unconsolidated data with some exceptions using consolidated data.

Figure 5

Land investment by industry



Source: Tachibana and Sekine (2003).

Figure 6

Estimation results for investment function

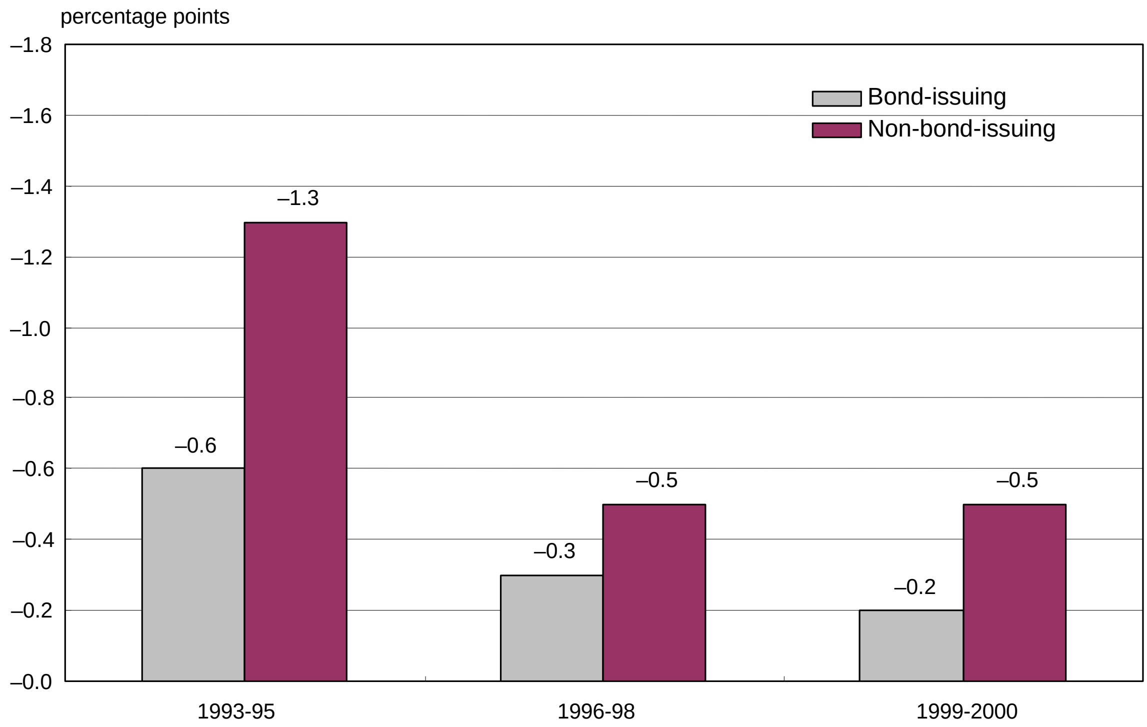
	(1)		(2)	
Dependent	I/K_{-1}		I/K_{-1}	
Bond issue	**Yes**		**No**	
Bank info	**Yes**		**Yes**	
I_{-1}/K_{-2}	–0.01	(0.04)	0.001	(0.04)
Δy	0.04	(0.04)	0.03	(0.05)
Δy_{-1}	0.09	(0.04)**	0.01	(0.04)
$(k-y)_{-2}$	–0.08	(0.04)**	–0.07	(0.04)*
y_{-2}	0.00	(0.01)	–0.05	(0.03)
Δj	–0.06	(0.02)***	–0.10	(0.03)***
Δj_{-1}	–0.07	(0.03)***	–0.08	(0.03)**
j_{-2}	–0.07	(0.04)*	–0.11	(0.06)
$CF/(p^k K)_{-1}$	–0.05	(0.07)	0.11	(0.07)
$(D/A)_{-1}$	–0.16	(0.05)***	–0.25	(0.09)***
Cap	0.07	(0.15)	0.56	(0.26)**
Sample period	1993-2000		1993-2000	
Observations	6,871		1,617	
Firms	856		222	
σ	0.086		0.096	
Sargan	123.9	[0.10]	141.1	[0.28]
AR(2)	–0.33	[0.74]	–0.51	[0.61]

Notes: 1. System GMM estimation (unbalanced panel). Coefficients on constants and time dummies are omitted. 2. Estimated coefficients are obtained from two-step estimators. Figures in parentheses are standard errors from two-step estimators with the Windmeijer small sample corrections. "***", "**" and "*" denote statistical significance at the 1%, 5% and 10% level, respectively. 3. AR(2) is a test for second-order residual serial correlation (the null hypothesis is no serial correlation). Sargan is a test for over-identifying restrictions (the null hypothesis is to satisfy over-identification). Figures in squared brackets are p-values. 4. Instruments for first-differenced equations are $(I_{t-2}/K_{t-3}),\ldots,(I_{t-9}/K_{t-10}),\ldots,\Delta y_{t-2}, \Delta y_{t-3}, (D/A)_{t-1}, (D/A)_{t-2}, \Delta j_t, \Delta j_{t-1}, Cap_t, Cap_{t-1}$. Those for level equations are $\Delta(I_{t-1}/K_{t-2})$. For column (2), $\Delta y_{t-4},\ldots,_{t-9}$ are added as instruments for the first-differenced equation.

Figure 7

Contribution of balance sheet condition to I/K_{-1}

(1) Contribution of firms' balance sheet condition



(2) Contribution of banks' balance sheet condition

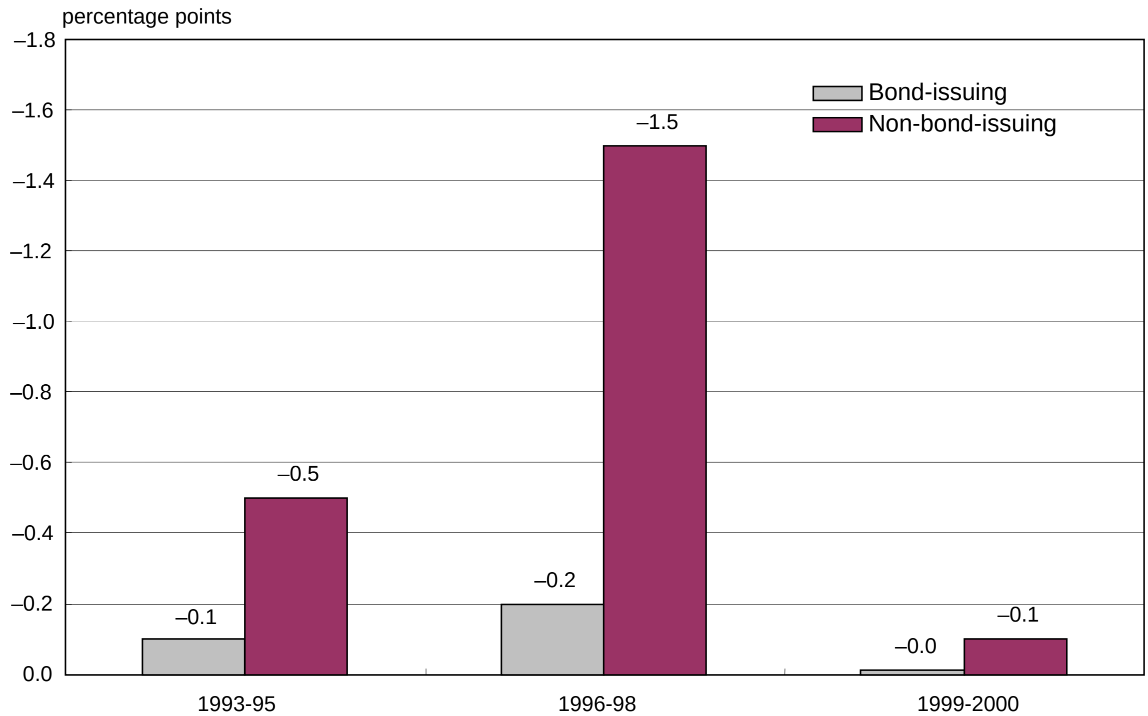


Figure 8

Correlations between loan growth and call rate

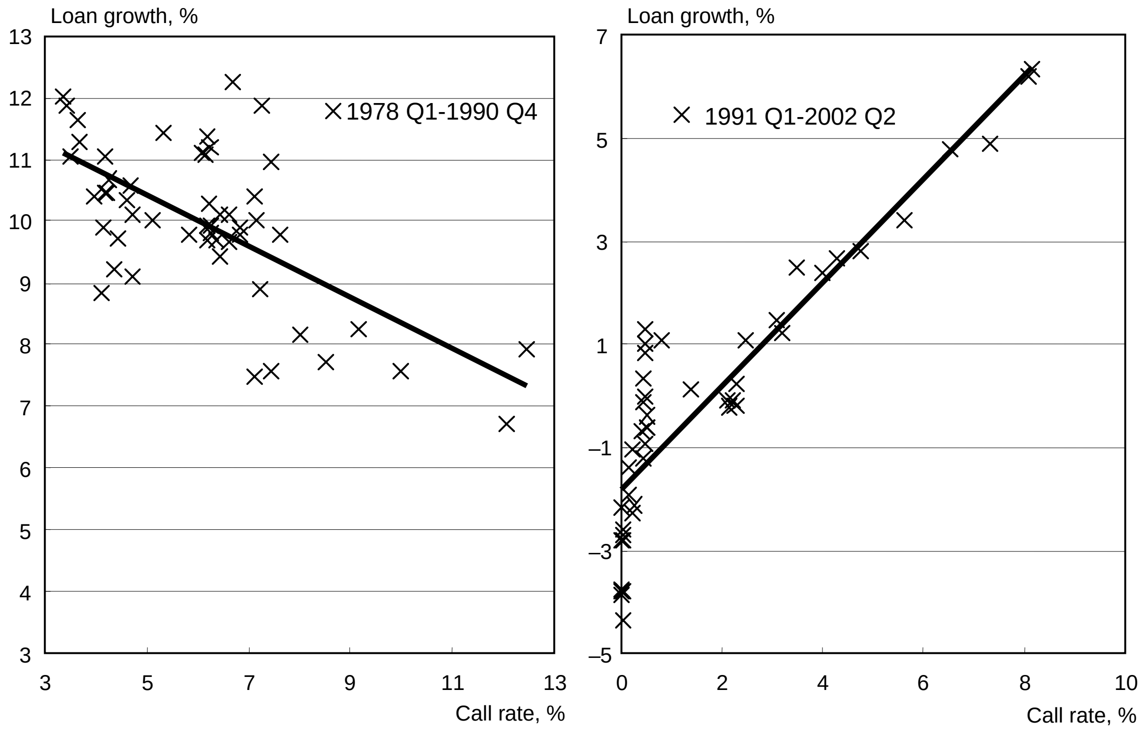


Figure 9

Estimation results for bank lending function

Industry Dependent	Manufacturing I		Construction and real estate I		Other non-manufacturing I	
			(A) Sample period: 1993-99			
I_{-1}	0.94	(0.02)***	0.97	(0.10)	0.97	(0.03)
r	0.12	(0.05)**	0.14	(0.08)*	0.04	(0.03)
$(D/A)_{-1}$	–0.12	(0.99)	–3.41	(1.76)*	–1.31	(1.12)
$(D/A)^2_{-1}$	–0.75	(2.11)	3.23	(1.94)*	2.02	(1.68)
ROA_{-1}	0.003	(0.01)	0.05	(0.03)	0.001	(0.02)
Observations	3,072		408		1,160	
Firms	384		51		145	
σ	0.06		0.09		0.05	
AR(2)	0.46	[0.65]	–1.37	[0.17]	–0.28	[0.78]
Sargan	112.7	[0.16]	37.4	[1.00]	116.3	[0.11]
			(B) Sample period: 1986-92			
I_{-1}	0.98	(0.02)	0.96	(0.05)***	0.98	(0.03)***
r	0.06	(0.02)	0.12	(0.03)***	0.10	(0.03)***
$(D/A)_{-1}$	–2.44	(1.49)	–3.51	(1.97)*	0.52	(1.20)
$(D/A)^2_{-1}$	4.07	(3.25)	4.40	(3.60)	–1.90	(1.82)
ROA_{-1}	–0.01	(0.01)	0.01	(0.02)	–0.01	(0.01)
Observations	3,072		408		1,160	
Firms	384		51		145	
σ	0.07		0.07		0.05	
AR(2)	0.10	[0.92]	–1.61	[0.11]	0.69	[0.49]
Sargan	125.2	[0.04]	36.64	[1.00]	111.2	[0.19]

Notes: 1. System GMM estimation (balanced panel). Coefficients on constants and time dummies are omitted. 2. Estimated coefficients are obtained from two-step estimators. Figures in parentheses are standard errors from two-step estimators with the Windmeijer small sample corrections. "***", "**", and "*" denote statistical significance at the 1%, 5%, and 10% level, respectively. 3. AR(2) is a test for second-order residual serial correlation, obtained from one-step estimators (the null hypothesis is no serial correlation). Sargan is a test for over-identifying restrictions (the null hypothesis is to satisfy over-identification). Figures in squared brackets are *p*-values. 4. Instruments for first-differenced equations are $I_{t-2},\ldots,_{t-5}$, $K_t,\ldots,_{t-5}$, $(D/A)_{t-2},\ldots,_{t-5}$, and $ROA_{t-2},\ldots,_{t-5}$. Those for level equations are ΔI_{t-1}, $\Delta(D/A)_{t-1}$, and ΔROA_{t-1}.

Figure 10

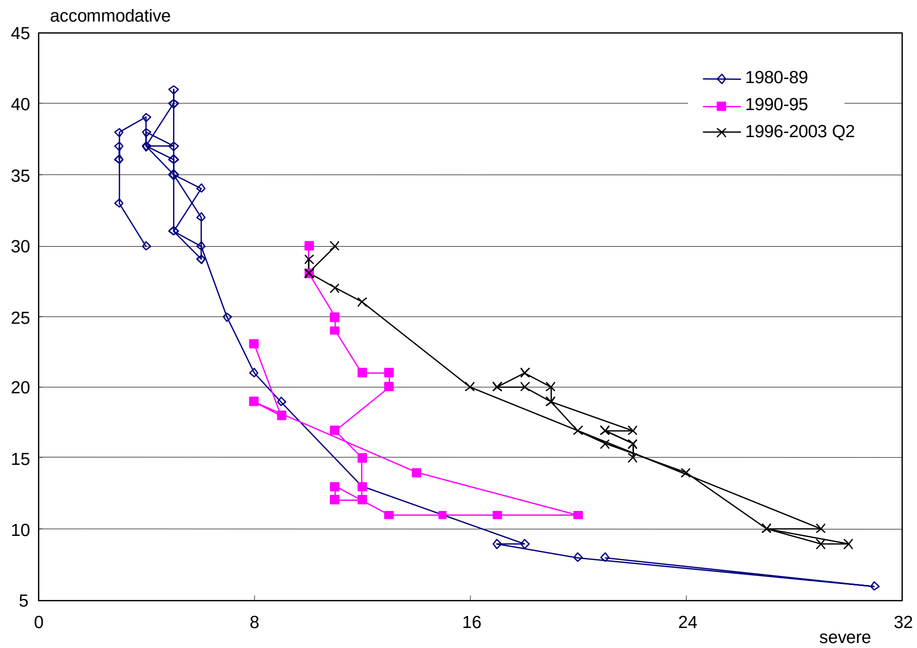
Firm profitability

Industry Dependent	Manufacturing *ROA*		Construction and real estate *ROA*		Other non-manufacturing *ROA*	
ROA_{-1}	0.54	(0.12)	0.73	(0.16)	0.83	(0.15)
$\Delta I\ (D/A)_{-1}$	–0.88	(1.23)	–2.56	(1.10)**	0.33	(0.92)
$\Delta Share$	3.49	(1.61)	–3.37	(1.91)*	0.70	(0.57)
Sample period	1993-99		1993-99		1993-99	
Observations	3,072		408		1,160	
Firms	384		51		145	
σ	4.18		1.45		1.67	
AR(2)	0.68	[0.50]	1.19	[0.23]	–0.32	[0.75]
Sargan	26.2	[0.07]	19.2	[0.32]	22.0	[0.19]

Notes: 1. System GMM estimation (balanced panel). Coefficients on constants and time dummies are omitted. 2. Estimated coefficients are obtained from two-step estimators. Figures in parentheses are standard errors from two-step estimators with the Windmeijer small sample corrections. "***", "**", and "*" denote statistical significance at the 1%, 5%, and 10%, respectively. 3. AR(2) is a test for second-order residual serial correlation, obtained from one-step estimators (the null hypothesis is no serial correlation). Sargan is a test for over-identifying restrictions (the null hypothesis is to satisfy over-identification). Figures in squared brackets are p-values. 4. Instruments for first-differenced equations are ROA_{t-2}, ROA_{t-3}, ΔI_{t-1}, ΔI_{t-2}, $(D/A)_{t-1}$, $(D/A)_{t-2}$, and $Share_t$. Those for level equations are ΔROA_{t-1}, ΔI_{t-1}, $(D/A)_{t-1}$, $(D/A)_{t-2}$, and $Share_t$.

Figure 11

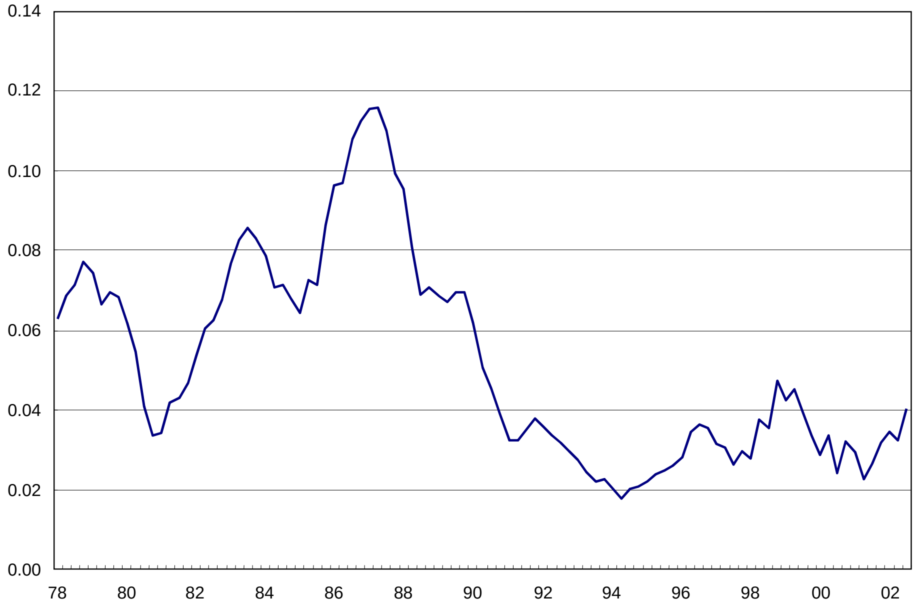
Tankan **survey on lending attitude of financial institutions**



Source: Bank of Japan, "*Tankan* short-term economic survey of enterprises in Japan".

Figure 12

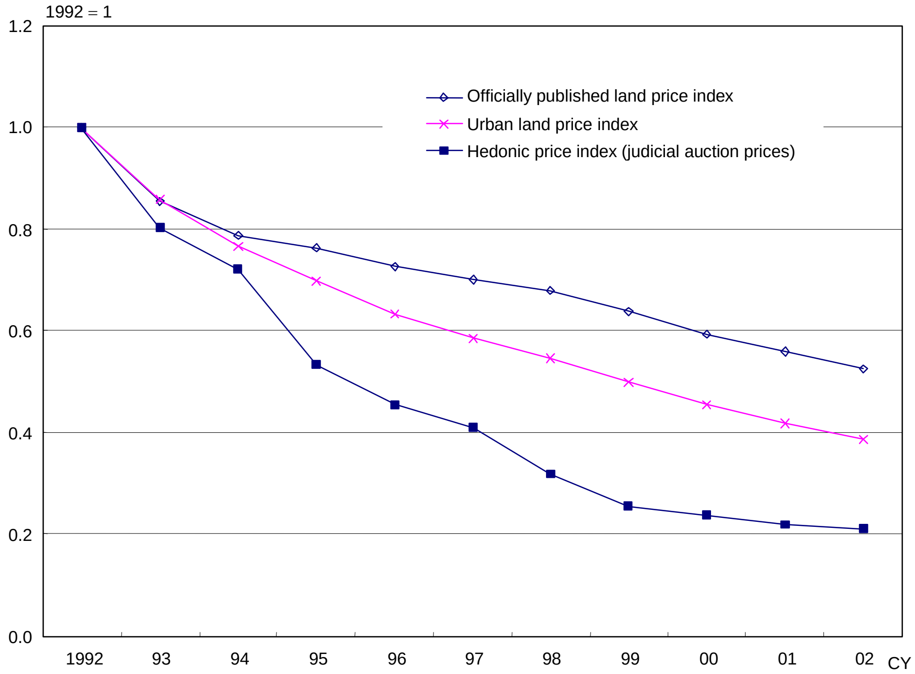
Sectoral credit shifts (σ^L)



Notes: 1. σ^L is calculated from 22 industries using data from Loans and Discounts Outstanding by Industry (Bank of Japan) from 1978 Q2 to 2002 Q4. 2. Current account overdrafts were not included in the series up to 1992 Q1, but have been included since then. 3. The figure for FY1993 is obtained from a linear interpolation of σ^L in 1992 Q1 and in 1993 Q2.

Figure 13

Land prices in Tokyo metropolitan area



Source: Saita (2003).

References

Baba, N (2001): “Optimal timing in banks’ write-off decisions under the possible implementation of a subsidy scheme: a real options approach”, *Bank of Japan Monetary Economic Studies*, 19 (3), 113-41.

Bangia, A, F X Diebold, A Kronimus, C Schagen and T Schuermann (2002): “Ratings migration and the business cycle, with application to credit portfolio stress testing”, *Journal of Banking and Finance*, 26, 445-74.

Bank of Japan (2002): “Japan’s nonperforming loan problem”, *Bank of Japan Quarterly Bulletin*, 10 (4), 97-108.

——— (2003a): “Review of Japanese banks’ activities: profits and balance sheets in fiscal 2002”, *Bank of Japan Research Papers*, available from http://www.boj.or.jp.

——— (2003b): “Evaluating the economic value of loans and the implication: toward transformation of the business model of banks and non-bank firms”, *Bank of Japan Research Papers*, available from http://www.boj.or.jp.

Bank of Japan, Research and Statistics Department (2003): “Recent trends in business fixed investment and the issues attending a full recovery: restoring firms’ capacity to generate capital investment”, *Bank of Japan Quarterly Bulletin*, forthcoming.

Berglöf, E and G Roland (1997): “Soft budget constraints and credit crunch in financial transition”, *European Economic Review*, 41, 807-17.

Caballero, R J, T Hoshi and A Kashyap (2003): *Zombie lending and depressed restructuring in Japan*, mimeo.

Chatelain, J B, A Generale, I Hernando, U von Karckreuth and P Vermeulen (2001): “Firm investment and monetary transmission in the euro area”, *European Central Bank Working Paper Series*, no 112.

Ehrmann, M, L Gambacorta, J Martines-Pages, P Sevestre and A Worms (2001): “Financial systems and the role of banks in monetary policy transmission in the euro area”, *European Central Bank Working Paper Series*, no 105.

Fisher, I (1933): “The debt-deflation theory of great depressions”, *Econometrica*, 1, 337-57.

Fujiwara, I (2003): “Has the effect of monetary policy changed during 1990s? An application of identified Markov switching vector autoregression to the impulse response analysis when the nominal interest rate is almost zero”, Osaka University, Discussion Papers in Economics and Business, no 03-07.

Gibson, M S (1997): “More evidence on the link between bank health and investment in Japan”, *Journal of the Japanese and International Economies*, 11, 296-310.

Kimura, T, H Kobayashi, J Muranaga and H Ugai (2003): “The effect of the increase in the monetary base on Japan’s economy at zero interest rates: an empirical analysis”, in *Monetary policy in a changing environment*, *BIS Papers*, no 19, 276-312.

Kobayashi, K and S Kato (2001): *Nihon Keizai no Wana (Trap of the Japanese economy)*, Nihon Keizai Shimbun-sha, Tokyo (in Japanese).

Krugman, P (1998): “It’s baaack: Japan’s slump and the return of the liquidity trap”, *Brookings Papers on Economic Activity*, 2, 137-205.

Lilien, D M (1982): “Sectoral shifts and cyclical unemployment”, *Journal of Political Economy*, 90, 777-93.

Nagahata, T and T Sekine (2002): “The effects of monetary policy on firm investment after the collapse of the asset price bubble: an investigation using Japanese micro data”, Bank of Japan, Research and Statistics Department, working paper, no 02-3.

Nickell, P, W Perraudin and S Varotto (2000): “Stability of rating transitions”, *Journal of Banking and Finance*, 24, 203-27.

Ogawa, K (2001): *Financial distress and corporate investment: the Japanese case in the 90s*, mimeo.

Osawa, N, K Kamiyama, K Nakamura, T Noguchi and E Maeda (2002): “An examination of structural changes in employment and wages in Japan”, *Bank of Japan Research Papers*, available from http://www.boj.or.jp.

Peek, J and E S *Rosengren (2003)*: “Unnatural selection: perverse incentives and the misallocation of credit in Japan”, NBER, working paper, no 9643.

Saita, Y (2003): “Land prices in the tokyo metropolitan area: a hedonic analysis of judicial auction”, Bank of Japan, working paper, no 03-E-4.

Saita, Y and T Sekine (2001): “Sectoral credit shifts in Japan: causes and consequences of their decline in the 1990s”, Bank of Japan, Research and Statistics Department, working paper, no 01-16.

Sakuragawa, M (2002): *Kinyuu Kiki no Keizai Bunseki* (*Economic analysis of financial crisis*), Tokyo University Press, Tokyo (in Japanese).

Sekine, T (1999): “Firm investment and balance sheet problems in Japan”, IMF, working paper, WP/99/111.

Sekine, T, K Kobayashi and Y Saita (2003): “Forbearance lending: the case of Japanese firms”, *Bank of Japan Monetary Economic Studies*, 21 (2), 69-92.

Tachibana, T and T Sekine (2003): “1990-nendai Ikou no Kigyou Bumon no Tochi Toushi ni tsuite (Land investment by firms after the 1990s)”, Bank of Japan, Research and Statistics Department, working paper, no 03-6, (in Japanese).

Ueda, K (2000): “Causes of Japan’s banking problems in the 1990s”, in T Hoshi and H Patrick (eds), *Crisis and change in the Japanese financial system*, Kluwer Academic Publishers, Norwell, pp 59-81.

Van den Heuvel, S J (2001): *The bank capital channel of monetary policy*, mimeo.

The bank lending channel in Chile[1]

Rodrigo Alfaro, Carlos García and Alejandro Jara, Central Bank of Chile
Helmut Franken, International Monetary Fund

1. Introduction

Modigliani and Miller (1958) undermined enthusiasm about the role of credit in the economy by suggesting that the capital structure of the firm was mostly irrelevant. Moreover, the strong and robust correlation between money and real variables found in the empirical literature of the 1960s provided strong support for the view that the main transmission mechanism for monetary policy operates through changes in the cost of capital and their impact on investment (the interest rate channel).[2] In that view, banks were important only because they created money. In the 1970s, however, the new field of the economics of information underscored the relevance of capital market imperfections and the uniqueness of bank loans against other forms of debt.[3] In this context, the "credit view" emerged as a new way of understanding the monetary policy transmission mechanism. This literature distinguishes between two subchannels, namely the broad credit channel and the bank lending channel, although more recent interpretations of the role that banks play in the transmission of monetary policy highlight the interaction between the two channels.[4]

This paper focuses on the bank lending channel, which emphasises the role played by banks in the transmission of monetary policy.[5] Thus, if the central bank follows a tight monetary policy, interbank lending is curtailed and the supply of funds for banks drops. Some banks might succeed in raising funds elsewhere, thus insulating their loan portfolios against monetary policy. Other banks, however, are forced to curtail their supply of credit, especially in the face of a strong negative monetary shock. Such a decrease in the bank loan supply is likely to be heterogeneous, as well, in the sense that heavily indebted households and small and medium-sized enterprises (SMEs), which are presumably bank-dependent, are crowded out of the market for bank loans and become severely financially constrained.[6] On the other hand, less binding adverse selection and moral hazard problems allow

---

1 This paper was published in Luis Antonio Ahumada and Rodrigo Fuentes (eds), *Banking Market Structure and Monetary Policy*, Santiago, Chile, © March 2004 Central Bank of Chile. This publication contains edited versions of some of the papers presented at the Sixth Annual Conference of the Central Bank of Chile that took place in Santiago on 12-13 December 2002. We thank Patricio Rojas, our conference discussant, as well as Simon Gilchrist and Anil Kashyap, for comments. The points of view expressed throughout this document are the authors' own, and are not necessarily shared by the institutions to which they are currently affiliated.

2 See, for example, Friedman and Schwartz (1963).

3 The seminal paper is Akerlof (1970). The author draws on the market for used cars to illustrate the problem caused by asymmetric information between dealers and buyers. Later references include Jaffee and Russell (1976), Townsend (1979), Stiglitz and Weiss (1981) and Diamond (1984), among many others.

4 See, for example, Huang (2003).

5 The broad credit channel (also known as the balance sheet channel) is related to the supply of credit by all financial intermediaries, emphasising the role of asymmetric information in the existence of an external financing premium. This premium is defined as the difference in the costs of external and internal financing. The external financing premium depends negatively on the net worth of a potential borrower and positively on the stance of monetary policy. Hence, it is a financial accelerator mechanism that amplifies the effects of monetary policy on investment and consumption decisions. See the Appendix for an overview of how both the broad credit channel and the bank lending channel are related to the whole set of monetary transmission mechanisms.

6 Because of their comparative advantages in information collection and processing, as well as their capacity to establish long-term relationships with their customers, banks are the only intermediaries able to offer credit to certain types of borrowers. However, banks that serve customers without any other market alternative have to deal with an asymmetric information problem, since it is difficult for the market to value their loan portfolios. Those banks will therefore experience difficulties in substituting their financial sources. For example, Goldberg et al (2002), using a survey on small firms conducted by the Federal Reserve, find that larger banks rely on standard techniques based on financial statements to make their commercial loan decisions. Smaller banks tend to deviate from these criteria, supporting their decisions with a much more personalised assessment of the entrepreneurs (of SMEs). In other words, the role played by asymmetric

large enterprises to maintain, if not increase, their access to domestic bank loans and other domestic financial sources.[7] As a result, the bank lending channel exacerbates the impact of a negative monetary policy shock in aggregate spending.

In distinguishing between movements in the demand and supply of bank credit - a key issue for interpreting the evidence on the bank lending channel - we follow a strategy of identification through heterogeneity, by comparing economic agents that are more likely to be affected by financial frictions with economic agents that are less likely to be so affected. In the words of Gilchrist and Zakrajsek (1995): “By observing and measuring the differential behavior of economic agents under consideration, one can potentially attribute some, if not all, of the difference in behavior to frictions caused by credit markets”.

Although we are well aware that the asymmetric nature of financial frictions also implies time-varying differences, that is, in and out of times of tight monetary policy, we concentrate on explaining cross-sectional differences by following a two-step approach. First, we follow a panel data approach to test how bank characteristics (size, liquidity and capitalisation) affect the response of loan supply after a change in monetary policy. Second, using the evidence gathered in the previous step regarding the main forces behind the bank lending channel, we construct an aggregate variable - the low-/high-quality ratio - aimed at capturing the availability of bank credit to households and SMEs vis-à-vis large enterprises. Using the low-/high-quality ratio, we test - within a vector autoregression (VAR) system - whether the bank lending channel exacerbates the effect of a monetary policy shock on macroeconomic activity.

Our panel data approach is closely related to Hernando and Martínez-Pagés (2001) and, to a lesser extent, to Kashyap and Stein (1995, 2000) and Kishan and Opiela (2000).[8] Our VAR approach is mainly related to Gilchrist and Zakrajsek (1995). Using this two-step approach, we conclude that the bank lending channel operated as a monetary policy transmission mechanism in Chile within the sample period, having a significant impact on macroeconomic activity.

The rest of the paper is organised as follows: Section 2 describes the data, Section 3 examines some methodological issues and presents the empirical results, and Section 4 concludes.

2. The data

The data used in this paper come mainly from financial statements of banks and publicly listed enterprises.[9] Our data set covers the period from the first quarter of 1990 to the second quarter of 2002. We also make use of several macroeconomic series, which are mostly taken from the Central Bank of Chile database.

When using bank statements, we consider only banks that are active participants in the credit market, excluding branches of foreign banks that are mainly engaged in cash and portfolio management activities.[10] This diminishes the problems associated with heterogeneous demand shocks, because

information is twofold: it affects the capacity of some banks to raise funds in situations of low market liquidity, and it generates a set of captive customers among banks.

[7] For example, if large firms are at the same time being directly affected by an external shock that is restricting their access to international financial markets, they will satisfy their financial needs domestically, thereby further crowding other agents out of financial markets. In addition to taking out bank loans, large Chilean enterprises have been actively issuing new domestic bonds in recent periods.

[8] See Cavieres (2002) for a study about the bank lending channel in Chile that closely follows Kishan and Opiela (2000).

[9] The bank statements are published in the statistical bulletin of the Superintendency of Banks and Financial Institutions (SBIF); the statements of publicly listed enterprises are taken from a data set assembled by the Santiago Stock Exchange containing all the information provided by the Fecu (ficha estadística codificada uniforme), a standardised statement that every listed company in Chile is required to file quarterly.

[10] When estimating the panel data, the original data set is adjusted slightly to take into account mergers that occurred during the sample period. We follow the intermediate strategy proposed by Hernando and Martínez-Pagés (2001), generating a new bank when a merger of banks of similar size takes place. If the merger is between banks of significantly different sizes, the data of the merged bank is considered as data of the largest merging institution and no new bank appears.

the share of different types of loans in the banks' portfolios does not differ significantly (Table 1). Even after this adjustment, our data set is quite representative of the credit market, accounting for more than 90% of total loans at any point in time (Graph 1).

Table 1

Characteristics of the banking system[1]

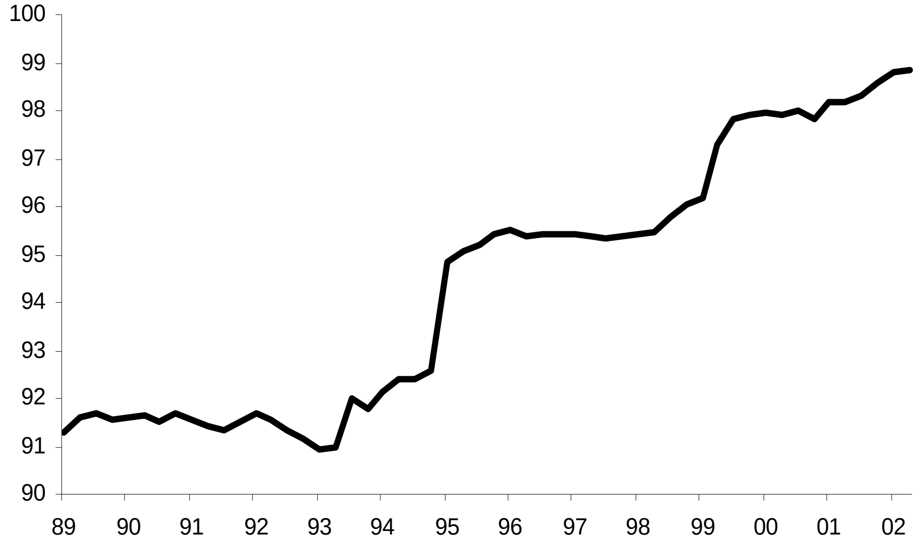
	Size				Capitalisation			
	<p25	**p25-50**	**p50-75**	**>p75**	**<p25**	**p25-50**	**p50-75**	**>p75**
Market share (%) of								
Total assets	3.9	10.0	23.1	63.0	29.5	40.0	22.8	7.7
Loans	1.1	4.5	20.5	73.9	36.6	46.2	16.5	0.8
Deposits	1.4	5.2	20.6	72.8	35.5	46.0	17.2	1.3
Size indicator								
Average number of bank branches	2.7	12.5	31.3	113.6	78.7	87.3	29.3	1.2
Average total assets[2]	12,134	32,117	71,944	205,512	122,428	180,964	97,110	34,403
Asset composition (%)	12.9	20.3	40.2	53.1	55.4	51.6	32.2	4.7
Loans	11.6	18.9	38.9	50.7	53.5	49.3	30.0	3.1
Commercial loans	44.3	44.7	57.4	57.0	59.4	58.9	53.4	48.3
Consumer loans	13.6	27.0	10.3	6.1	11.7	7.8	8.7	5.5
Mortgage loans	0.5	2.6	12.3	16.4	11.6	17.6	20.3	0.1
Other loans	41.7	25.7	19.9	20.5	17.3	15.8	17.6	46.1
Securities	6.8	7.8	9.6	14.7	8.8	12.8	10.6	4.6
Other assets	81.6	73.3	51.5	34.6	37.7	38.0	59.3	92.3
Liability composition (%)								
Deposits	51.2	68.4	63.9	62.5	66.3	64.3	61.1	52.0
Overnight deposits	7.5	4.8	8.6	14.1	11.4	12.7	13.4	7.2
Time deposits	43.8	63.6	55.3	48.4	54.9	51.6	47.7	44.8
Mortgage bonds	0.4	2.0	14.7	16.9	17.1	18.4	18.1	0.1
Foreign loans	8.0	9.5	6.7	7.7	4.6	4.2	5.7	2.8
Subordinate bonds	0.0	0.2	1.8	1.7	2.3	2.3	1.2	0.0
Stock of provisions	1.4	2.6	2.4	2.6	2.1	1.9	2.0	1.0
Capital and reserves	38.9	17.3	10.4	8.6	7.6	8.9	12.0	44.0

[1] This analysis is performed for the whole sample period (1990-2002). P*xx* refers to the corresponding percentile of the distribution of banks by asset size and capitalisation. The percentiles are calculated for each quarter separately. [2] In millions of pesos.

Sources: SBIF; authors' calculations.

Graph 1

Share of total loans for banks included in the sample, in per cent



Source: SBIF.

From these bank statements we collect total loans, consumer loans and commercial loans. The distinction between consumer loans and commercial loans also points towards a better identification of movements in the supply of credit.[11] Indeed, evidence indicates a differential behaviour of various types of loans during the business cycle (Graph 2), which suggests that diverse types of loans may be affected differently by demand shocks.

We also collect our proxies for bank characteristics - size, liquidity and capitalisation - which are based on how the existing empirical literature about the bank lending channel captures the potential problems associated with asymmetric information.[12] Size is defined as the bank's share of the total assets of the banking system; liquidity is defined as the ratio of liquid assets to total assets; and capitalisation is defined as the seasonally adjusted ratio of capital and reserves to total assets. Table 2 presents the main descriptive statistics on this set of bank characteristics.

Table 2

Descriptive statistics on bank characteristics[1]

	Mean	**Standard error**	**Minimum**	**Maximum**	**p25**	**p50**	**p75**
Size	4.21	4.01	0.03	19.04	0.87	3.24	5.92
Liquidity	20.69	9.01	4.48	53.92	13.41	19.58	27.26
Capitalisation	8.76	9.43	1.09	63.44	4.64	5.68	7.95

[1] P*xx* refers to the corresponding percentile of the distribution of banks by asset size, liquidity and capitalisation.

Source: Authors' calculations.

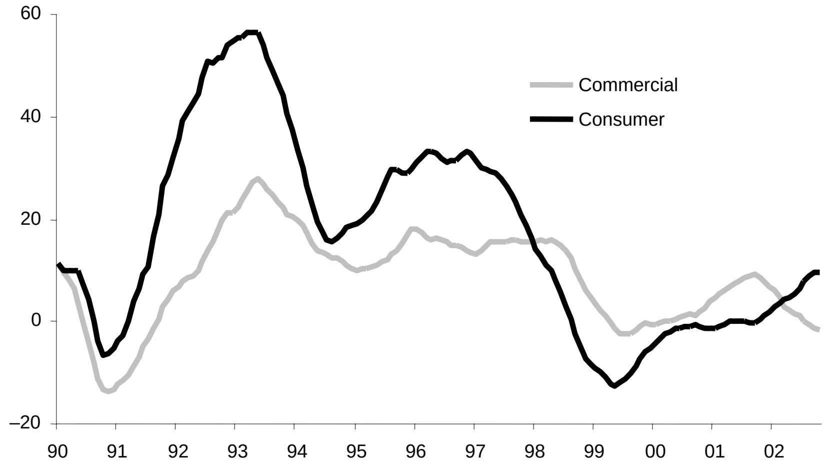
[11] As suggested by Hernando and Martínez-Pagés (2001).

[12] See, for example, Kashyap and Stein (1995, 2000) and Kishan and Opiela (2000).

Graph 2

Annual growth of consumer and commercial loans

All banks, moving average in per cent



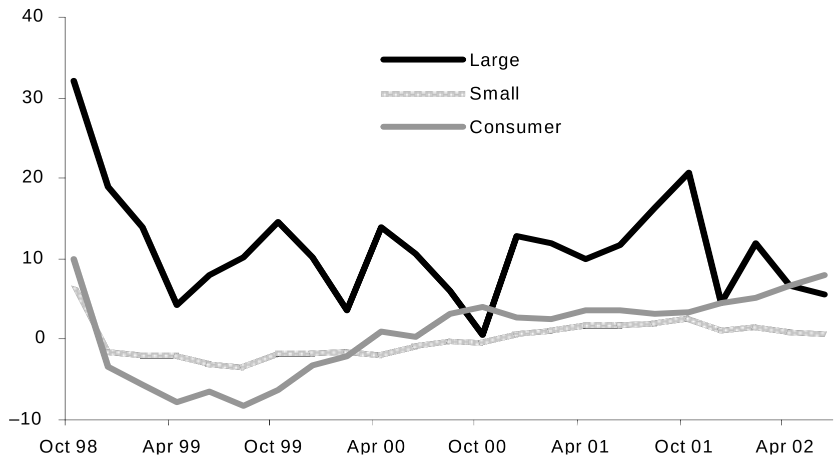
Source: SBIF.

From the statements of publicly listed enterprises, we take the total large corporate sector bank debt. Using this variable as the denominator and the consumer loans of the banking system as the numerator, we construct a variable that we call the low-/high-quality ratio, to capture the availability of bank credit to households and SMEs vis-à-vis large enterprises. Two features of this ratio deserve further explanation: the extent to which consumer loans capture not only household credit but also loans directed to SMEs; and the relation of this ratio to a flight to quality. With regard to the first feature, we could have measured credit to SMEs more directly using data that is available by loan size, but this series is only available as from 1996, and with less than quarterly frequency. However, when graphing the small business loans and consumer loans together (Graph 3), the two series follow a relatively similar path (the correlation is over 90%). Credit to SMEs is, in fact, known to usually take the form of a consumer credit in the Chilean banking industry, whereas credit to large enterprises follows a very different path.

Graph 3

Annual growth of small, large and consumer loans

In per cent



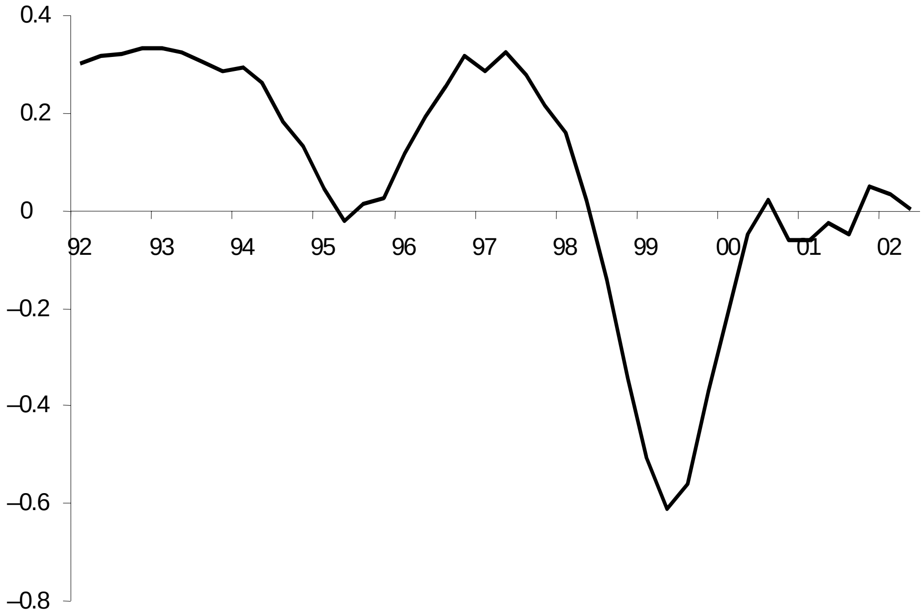
Source: SBIF.

With regard to the second feature, our low-/high-quality ratio is (inversely) related to the indicator of a flight to quality constructed by Caballero (2002) using precisely the share of large loans from the available data by loan size. Although our story is different from Caballero's, in the sense that we are trying to pin down the effect of a monetary policy shock instead of an external shock, the operative financial mechanism is basically the same: indebted consumers and especially SMEs are crowded out of the banking system by large firms, thus becoming severely financially constrained. Graph 4 shows a severe flight to quality effect in 1998-99, a period of extremely tight monetary policy.

Graph 4

Annual growth in low-/high-quality ratio

Moving average, in per cent



Sources: Central Bank of Chile; authors' calculations.

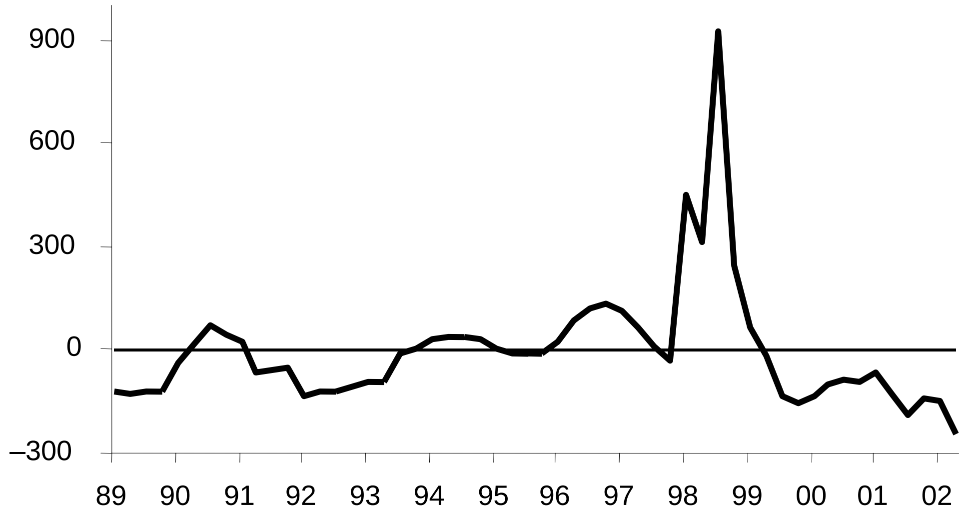
To identify the effect of a monetary policy shock on the supply of bank loans, we need an indicator that is closely tied to monetary policy. The international empirical literature offers several alternatives, but the set of choices in the case of Chile is limited by data availability. Within this limited choice set, we choose the term spread, defined as the difference between the monetary policy rate and the PRC8.[13] As explained in Gertler and Lown (2000), a positive movement in the term spread (so defined) simply reflects the fact that the monetary tightening is inducing a fall in long-term rates, because there are expectations of a drop in the short-term interest rate in the near future (Graph 5).

[13] The PRC8 are long-term indexed bonds issued by the Central Bank of Chile. See Estrella and Mishkin (1998) for a positive assessment of the predictive power of the term spread; see Gertler and Lown (2000) for an explanation of the close relationship between the term spread and monetary policy, particularly in periods of significant monetary tightening.

Graph 5

Term spread

In basis points



Source: Central Bank of Chile.

Finally, we use several macroeconomic series in the panel and the VAR system. Specifically, in the panel of banks we use the annual growth of real GDP to capture changes in income, and the annual depreciation of the real exchange rate to capture movements in relative prices. Both variables are intended to control for demand effects. In the VAR system, we use three additional endogenous variables (besides the low-/high-quality ratio and the term spread): namely, a proxy for macroeconomic activity (in logs and seasonally adjusted), the consumer price index (in logs and seasonally adjusted) and the real exchange rate (in logs). We use six different proxies for macroeconomic activity: real GDP, industrial production, business investment, durable goods consumption, unemployment rate and residential investment. In addition to these endogenous variables, every VAR model includes the following set of exogenous variables: terms of trade, inflation target, external output and a time trend.[14]

3. Methodological issues and empirical results

Our main goal in this section is to analyse whether the bank lending channel played any role as a transmission mechanism for monetary policy in the Chilean economy during the period 1990-2002 and, if so, whether this transmission mechanism plays any significant macroeconomic role. We follow a two-step approach. First, we use a panel of bank data to identify shifts in the loan supply curve in response to changes in monetary policy by exploiting the heterogeneity among banks. Such an exercise lets us gather evidence about where the bank lending channel has operated most strongly. Second, we use that knowledge to construct a variable that is likely to be a good proxy of how the

[14] This is justified on the grounds that Chile is a small open economy with an inflation target regime operating since the early 1990s. In particular, by including the terms of trade, we are controlling for external shocks. Hence, if we find that the low-/high-quality ratio influences economic activity following a monetary policy shock, we can interpret the flight to quality effect as being domestically driven.

bank lending channel exacerbates the monetary policy shock, thus having an independent and significant impact on aggregate spending. This variable is the low-/high-quality ratio, which captures the availability of bank credit to households and SMEs vis-à-vis large enterprises. Here again, we appeal to heterogeneity for identification purposes, this time among borrowers. Finally, we embed the low-/high-quality ratio within a VAR system to test whether the bank lending channel exacerbates the effect of a monetary policy shock on macroeconomic activity.

3.1 First step: a panel of bank data

As discussed in the introduction, a tight monetary policy reduces the amount of funds available to the banking system, and some banks are unable to offset the reduction in interbank funds owing to information problems. How do bank characteristics affect the response of loan supply following a monetary policy shock? To answer this question, we follow a panel data approach in which bank characteristics (size, liquidity and capitalisation) interact with the term spread (our indicator of monetary policy) to disentangle the differential behaviour of banks with regard to total loans, consumer loans and commercial loans.

In this panel model, the dynamic structure is adequately handled by introducing one lag for the endogenous variable and four lags for the term spread, the variables aimed at controlling for demand effects and the variables related to bank characteristics. Although including a lag of the dependent variable is trivial in the time-series context, the fixed-effects estimator is severely biased in a dynamic context. Instead of following the traditional approach to dealing with such a problem - namely, the Arellano and Bond generalised method of moments (GMM) procedure - we use the bias-corrected estimator proposed by Hahn and Kuersteiner (2002).[15]

The empirical specification within this panel data approach is the following:

$$y_{it} = \rho y_{it-1} + \sum_{j=0}^{4} x'_{it-j}\beta + z'_{it-1}\gamma + \sum_{j=1}^{4} x_{3it-j} z'_{it-1}\phi + \sum_{s=1}^{4} \sigma D_{st} + u_{it},$$

where y_{it} represents the annual growth of total loans, commercial loans and consumer loans, respectively; x_{it} is a vector of macroeconomic variables aimed at controlling demand side shocks (annual growth of GDP and annual depreciation of the real exchange rate) in addition to the monetary policy indicator (term spread); z_{it} denotes a vector of bank-specific variables (liquidity, size and capitalisation); D is a set of seasonal dummies; u_{it} is iid; $i = 1, ..., N$ represents the number of banks included in the data set; and $t = 1, ..., T$ is the time index from the first quarter of 1990 to the second quarter of 2002. Note that the bank-specific explanatory variables z_{it} are included with one lag to account for potential endogeneity.

We disentangle loan supply from loan demand effects by looking at cross-sectional differences in the response of bank loans to a monetary policy shock. Were these differences to be related to indicators of the degree of informational asymmetries (size, liquidity, or capitalisation), they would support the existence of the bank lending channel. More specifically, if the bank lending channel holds, we should expect a positive and significant cross-coefficient between the term spread and bank characteristics.

Table 3 shows the long-run coefficients for each of the explanatory variables. First, note that the long-run coefficient for the annual growth of real GDP, when statistically significant, is positive. Second, the long-run coefficient for annual real depreciation is always significant and negative. Third, the long-run coefficient of the term spread, which is positively related with a tighter monetary policy, is always significant and negative. Finally, regarding the interaction of bank characteristics with monetary policy, the results show that liquidity is always significant and positive, size is positive and significant only for total loans, and capitalisation is positive and significant only for consumer loans.

[15] The Arellano and Bond GMM procedure is subject to substantial finite sample bias, as shown by Alonso-Borrego and Arellano (1999) and Hahn et al (2002). For a more technical discussion of the methodological issues, see Brock and Franken (2003).

Table 3

Long-run coefficients and standard errors

Dependent variable	Coefficient	Standard error
1. Growth of total loans		
Real GDP growth	0.57*	0.19
Real exchange rate devaluation	–0.93*	0.11
Term spread	–4.31*	0.46
Bank characteristic and term spread:		
Liquidity	7.83*	1.56
Size	13.24*	2.83
Capitalisation	–1.43	3.85
2. Growth of consumer loans		
Real GDP growth	1.09*	0.19
Real exchange rate devaluation	–0.20**	0.10
Term spread	–2.65*	0.57
Bank characteristic and term spread:		
Liquidity	6.41*	1.66
Size	3.44	3.89
Capitalisation	5.39*	1.37
3. Growth of commercial loans		
Real GDP growth	–0.02	0.37
Real exchange rate devaluation	–1.71*	0.21
Term spread	–6.85*	0.99
Bank characteristic and term spread:		
Liquidity	13.59*	4.01
Size	2.22	4.21
Capitalisation	–3.94	6.28

Note: * 1% significance level; ** 5% significance level.

Table 4 shows the overall effects of a tight monetary policy in terms of the annual growth rate of total loans, consumer loans and commercial loans.[16] As can be seen from the table, tightening monetary policy results in a larger drop in the growth rate of total loans for small banks than for large banks.[17] In addition, the drop in the growth rate of all types of loans is larger for less liquid banks than for their

[16] The overall effects include the direct effect of monetary policy plus the interactive effects of bank characteristics with monetary policy. If the parameter is non-significant, it is computed as being equal to zero. Bank characteristics are evaluated at three representative levels for each category.

[17] A 1 percentage point increase in the term spread accounts for an annual reduction of 4.2% in total loans when the bank is small, but only 3.5% when the bank is large.

more liquid counterparts.[18] In the case of consumer loans, the bank lending channel operates through less capitalised banks.[19]

Table 4

Overall effect of a monetary policy shock on the rate of growth of loans[1]

	Size			Capitalisation			Liquidity		
	p25	**p50**	**p75**	**p25**	**p50**	**p75**	**p25**	**p50**	**p75**
Total	–4.2	–3.9	–3.5	–4.3	–4.3	–4.3	–3.3	–2.8	–2.2
Consumer	–2.6	–2.6	–2.6	–2.4	–2.3	–2.2	–1.8	–1.4	–0.9
Commercial	–6.9	–6.9	–6.9	–6.9	–6.9	–6.9	–5.0	–4.2	–3.1

[1] P*xx* refers to the corresponding percentile of the distribution of banks by asset size, capitalisation and liquidity.

Our preliminary results thus support the idea that the bank lending channel has operated in Chile. Furthermore, consumer loans seem to better capture the role played by informational asymmetries in the response of bank lending to monetary policy shocks. Indeed, both liquidity and capitalisation have played a restrictive role for consumer loans, while commercial loans have only been affected by liquidity. We argued above that consumer loans are a reasonably good proxy for bank credit directed to both households and SMEs. Hence, our results in this first step suggest that the decrease in banks' loan supply may have actually been heterogeneous, affecting more SMEs and, to a lesser extent, highly indebted households, than large enterprises. The next step concentrates on providing more solid evidence along these lines.

3.2 Second step: a VAR system including an aggregate proxy for the bank lending channel

The fact that banks' loan supply affects borrowers heterogeneously can be exploited to identify how the bank lending channel magnifies a monetary policy shock. We therefore construct the low-/high-quality ratio to capture the availability of bank credit to households and SMEs vis-à-vis large enterprises.[20] More specifically, we ask the following question regarding the impact of monetary policy on the real sector of the economy: does the bank lending channel play any significant macroeconomic role as a monetary transmission mechanism? To answer it, we analyse whether the low-/high-quality ratio has marginal predictive power over a set of macroeconomic variables.

We expect a negative monetary policy shock to reduce the low-/high-quality ratio (flight to quality), which would strongly affect bank-dependent households and SMEs by eliminating their only source of external funding.[21] For example, casual evidence for the Chilean economy shows that SMEs have

[18] A 1 percentage point increase in the term spread accounts for an annual reduction of 3.3% in total loans, 1.8% in consumer loans and 5.0% in commercial loans for a less liquid bank. On the other hand, a 1 percentage point increase in the term spread accounts for an annual reduction of only 2.2% in total loans, 0.9% in consumer loans and 3.1% in commercial loans for a highly liquid bank.

[19] A 1 percentage point increase in the term spread accounts for an annual reduction of 2.4% in consumer loans when the bank is less capitalised, but only 2.2% when the bank is more capitalised.

[20] See Section 2 for a more detailed explanation of this particular variable.

[21] See footnote 5.

quite limited access, if any, to bond-issuing or capital-raising on the stock market.[22] In other words, the decline in the low-high-quality ratio represents a decrease in the portion of banks' loan supply directed to those economic agents (households and SMEs) which bear the largest share of the costs associated with information problems. This may, in turn, have a significant effect on economic activity.[23]

The empirical approach used in this section consists in estimating a set of VAR models in levels, each of which includes the low-/high-quality ratio that accounts for the existence of the bank lending channel. Four endogenous variables are also included, namely the term spread as the indicator of the monetary policy stance, a proxy for macroeconomic activity (with six different alternatives), the real exchange rate and the price level. Finally, every model includes a set of exogenous variables: terms of trade, inflation target, external output and a time trend.[24]

To assess the macroeconomic importance of the bank lending channel, we test for the marginal predictive power of the credit variable (low-/high-quality ratio) by carrying out Granger causality tests and reporting the corresponding p-values. A rejection of the null hypothesis that the credit variable is irrelevant for predicting macroeconomic activity is one piece of evidence in favour of the bank lending channel. This evidence has to be complemented with two simultaneous conditions, however: rejection of the null hypothesis that the term spread is irrelevant for predicting the credit variable, and failure to reject the null hypothesis that the proxy for macroeconomic activity is useless in predicting the credit variable. In other words, the bank lending channel requires that lagged values of the term spread be significant in predicting the credit variable, which in turn must be significant in predicting either macroeconomic activity or other macroeconomic variables.

Table 5 shows the Granger causality test for each VAR model. The results support the hypothesis that the low-/high-quality ratio predicts macroeconomic variables in all cases. These results also indicate that the lags of the term spread are significant for predicting macroeconomic variables in just three out of six cases.[25] On the other hand, macroeconomic variables are not helpful for predicting the low-/high-quality ratio in each case, whereas the term spread is helpful for predicting the low-/high-quality ratio in all cases. The empirical evidence thus strongly supports a causality running from monetary policy to credit and from credit to macroeconomic activity.

---

[22] This is consistent with the international empirical evidence, which shows that finding alternative sources of credit is quite difficult for SMEs.

[23] The drop in the supply of bank credit pushes SMEs to curtail their productive activities, which are usually labour-intensive. This has a strong impact in terms of job destruction, since the affected workers are generally unskilled and thus difficult to absorb into other sectors of the economy. Because increasing unemployment rates are strongly correlated with consumer confidence (in the United States and elsewhere), aggregate demand falls. Hancock and Wilcox (1998) find that small banks engage in "high power" credit activities, with a drop in their credit supply having a large impact on economic activity measured in terms of unemployment, real wages, GDP and number of bankruptcies.

[24] We use a two-step procedure to define the optimal lag structure (Johansen (1995)): the first step uses the Schwarz Bayesian criterion; the second step adds additional lags for eliminating any evidence of serial correlation detected by the multivariate LM test statistics for residual serial correlation.

[25] At the 5% level of significance.

Table 5

VAR pairwise Granger causality/block exogeneity Wald tests
p-values from exclusion tests

Models classified according to proxies for macroeconomic activity	Variables exclude from:			
	Macroeconomic activity equation	**p-values**	**Low-/high-quality ratio equation**	**p-values**
GDP[1]	Monetary policy shock Low-/high-quality ratio	95.6% 0.0%	GDP Monetary policy shock	73.6% 0.2%
Industrial production[1]	Monetary policy shock Low-/high-quality ratio	4.5% 0.5%	Industrial production Monetary policy shock	90.7% 0.6%
Business investment[1]	Monetary policy shock Low-/high-quality ratio	68.7% 0.0%	Business investment Monetary policy shock	66.5% 0.2%
Durable consumption[2]	Monetary policy shock Low-/high-quality ratio	0.2% 1.9%	Durable consumption Monetary policy shock	52.6% 3.3%
Unemployment rate[1]	Monetary policy shock Low-/high-quality ratio	44.7% 0.0%	Unemployment Monetary policy shock	95.6% 0.8%
Residential investment[2]	Monetary policy shock Low-/high-quality ratio	3.1% 1.9%	Residential investment Monetary policy shock	55.4% 2.5%

Note: This table shows the results obtained from six VAR models. Each one uses a different option for measuring macroeconomic activity: real GDP, industrial production, business investment, durable consumption, unemployment rate and residential investment, respectively. Each proxy is added one at a time to the base VAR. The base model is comprised of five variables: real GDP, CPI, term spread, low-/high-quality ratio and real exchange rate. The exogenous variables are terms of trade, inflation target, external output and a time trend.

The numbers in the table are the p-values for the null hypothesis that some variables do not contain information for the dependent variable. For each model, we pick the equations representing both the proxy for macroeconomic activity and the credit variable (low-/high-quality ratio). Then we perform the following tests:

(i) Term spread and the credit variable do not Granger cause macroeconomic activity; and

(ii) Macroeconomic activity and monetary policy do not Granger cause the credit variable.

If p-values are lower than 5% we can reject the null hypothesis.

[1] Endogenous variables two lags, exogenous variables two lags. [2] Endogenous variables three lags, exogenous variables two lags.

To study the dynamics of the bank lending channel, we estimate a structural vector autoregression (SVAR) and report impulse responses to a monetary policy shock. The set of identifying assumptions is borrowed from a vast list of authors who use this type of identification scheme in VAR models.[26] Variables are thus divided into three recursive sets: non-policy variables that are not contemporaneously affected by policy variables; policy variables; and non-policy variables that are contemporaneously affected by policy variables.[27] In other words, the central bank's feedback rule is identified by dividing the set of non-policy variables into variables that cause a policy reaction and variables that are affected by the policy reaction. For the policy variables, we assume the following sequence of events: the central bank first sets an inflation target, which is an exogenous variable, and then sets the monetary policy stance.[28] For the non-policy variables, we assume a recursive causal relationship ordered as follows: price level, output and the credit variable.[29] Our positioning of the variable used as a proxy for the bank lending channel (low-/high-quality ratio) in last place is based on the assumption that the central bank is able to affect it contemporaneously through the monetary policy stance, since capital markets tend to respond faster than goods and labour markets.[30]

Graph 6 displays the estimated impulse responses (black lines). The low-/high-quality ratio decreases following the monetary policy shock, a result that is consistent with a flight to quality effect as described above. GDP declines about two quarters after a tightening in monetary policy. The maximum decline occurs about a year after the shock, and the effect gradually dies out thereafter. We observe a similar pattern when GDP is replaced by industrial production or unemployment rate, although the effect seems to be more persistent in the latter case.

When both investment and durable consumption replace GDP, these two components of aggregate output decline during the first year and a half. This result differs from the international empirical evidence. For example, Bernanke and Gertler (1995) find evidence that in the United States the decline of durable consumption and residential investment precedes that of business fixed investment. Their interpretation is against the conventional monetary policy transmission mechanism that operates through an earlier decline in investment. In the Chilean case, however, the impulse responses indicate that durable consumption and both types of investment decrease at approximately the same time. We interpret this as evidence that both transmission mechanisms are relevant for Chile.

---

26 See, for example, Bernanke and Blinder (1992), Gertler and Gilchrist (1994), Eichenbaum and Evans (1995), Strogin (1995), Christiano et al (1996, 1997, 1999) and Bernanke and Mihov (1998). For the case of Chile, see Bravo and García (2002).

27 In our particular case, we use an exactly identified VAR because additional identifying restrictions in the parameters do not change the results obtained in the impulse response functions.

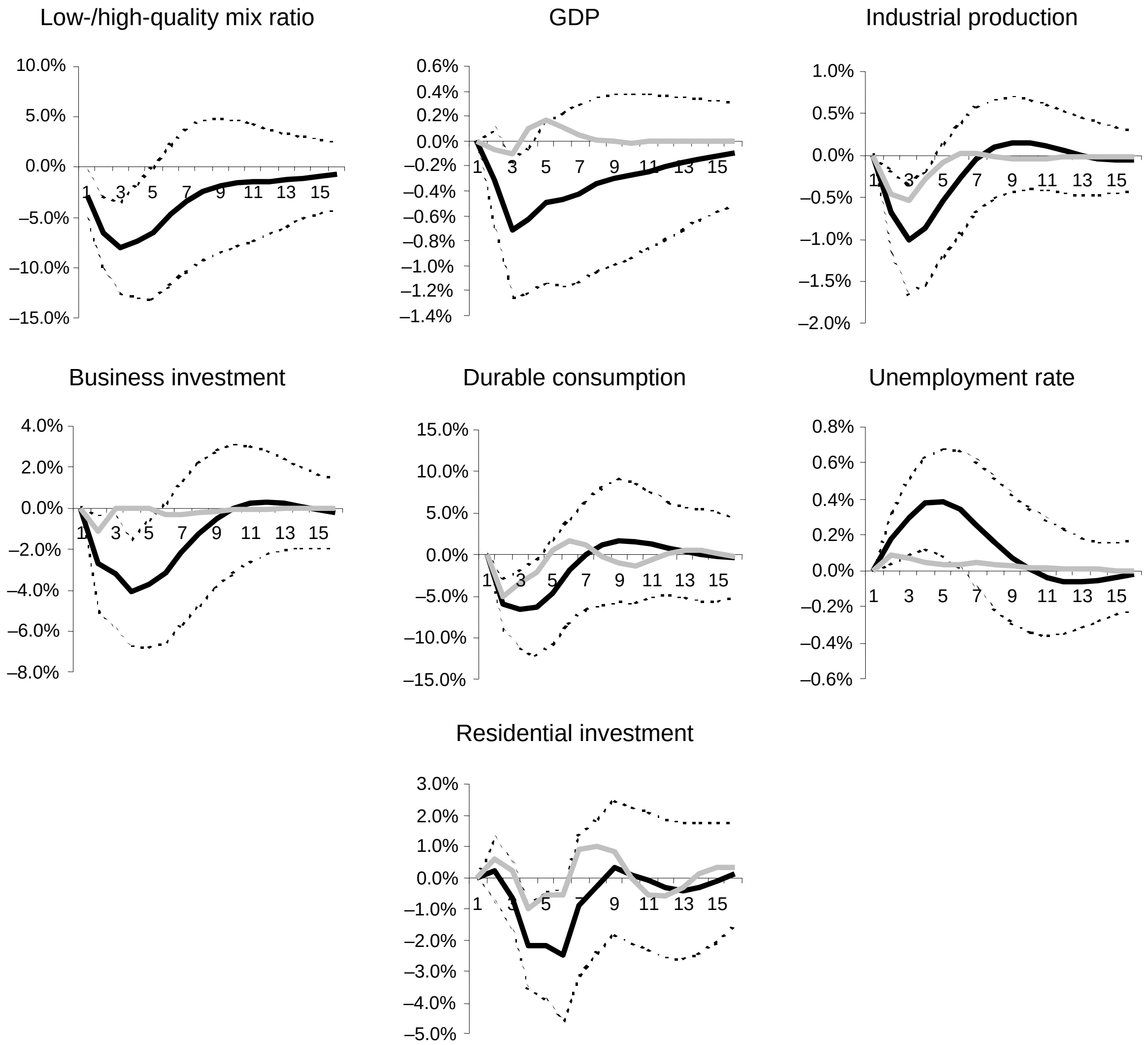
28 This assumption is consistent with the fact that the monetary policy rate is used as a fine-tuning policy, given a known inflation target.

29 The assumption behind this order is that the price level is stickier than output, a fact that is consistent with the high level of backward indexation in the Chilean economy (Jadresic (1996)).

30 To illustrate the identifying assumptions described above, assume that the central bank contemporaneously knows the evolution of the inflation rate but is not able to affect it. If the economy faces an inflationary shock (an oil shock, for instance), the central bank could respond with a change in the monetary policy rate. This, in turn, would have an immediate impact on other variables, such as the low-/high-quality ratio and the exchange rate. Only then might monetary policy affect variables such as GDP, investment, consumption and inflation.

Graph 6

Responses to structural one SD innovations ± 2 SE on monetary policy shock[1,2,3]



[1] For VAR specification see Table 5. [2] Black lines for the bank lending channel (low-/high-quality ratio) being endogenous. [3] Grey lines for the bank lending channel (low-/high-quality ratio) being exogenous.

The empirical strategy described above allows us to compare the impulse responses to a monetary policy shock in two different systems, in which the variable used as a proxy for the bank lending channel (ie the low-/high-quality ratio) is first defined as endogenous (black lines) and then as exogenous (grey lines). Shutting down the bank lending channel effect on other macroeconomic variables following a monetary policy shock establishes a measure of the macroeconomic relevance of the bank lending channel: namely, the difference between the two impulse responses.[31] To determine whether this difference is statistically significant, we display the dashed lines that represent a 95% confidence interval for each impulse response function when the bank lending channel is endogenous.

[31] From the Granger causality tests, we already know that the empirical evidence strongly supports a causality running from monetary policy to credit and from credit to macroeconomic activity. What we are doing here, therefore, is determining whether the flight to quality effect occurs as a result of a monetary policy shock or is driven by other factors.

If the impulse response functions calculated under the assumption that the credit variable is exogenous fall outside this confidence interval, we interpret this as evidence in favour of the macroeconomic relevance of the bank lending channel.

What do we find? The bank lending channel is unambiguously relevant in terms of GDP, business investment and the unemployment rate, since the responses of these variables are definitely much weaker if the proxy for the bank lending channel is exogenously included in the system. The other results also support the macroeconomic relevance of the bank lending channel to a degree, since durable consumption, residential investment and industrial production are on the brink of being statistically different from the case of an endogenous bank lending channel.[32]

4. Concluding remarks and directions for future research

We conclude that the bank lending channel operated as a monetary policy transmission mechanism in Chile during the period 1990-2002, with an independent and significant effect in terms of macroeconomic activity. The way that the bank lending channel seems to have operated in Chile is consistent with the international empirical evidence: first, some banks - less liquid banks and, to a lesser extent, smaller and less capitalised banks - are forced to curtail their supply of credit following a monetary policy shock; second, the access of households and SMEs to external financing is severely restricted following the drop in the supply of bank credit; third, the uneven distribution of the drop in the supply of bank credit, which can be associated with a flight to quality effect, has a significant influence in terms of macroeconomic activity. By pushing towards a better understanding of the way in which the bank lending channel operates as a transmission mechanism of monetary policy in Chile, our paper contributes to an improvement in the monetary policy decision framework.

Our focus in this paper is on explaining cross-sectional differences among economic agents (banks, firms and, to a lesser extent, households). The evidence gathered in this paper therefore points towards a bank lending channel operating across the sample period, abstracting from the asymmetries related to tightening versus easing of monetary policy and from the evolution of certain features in the economy that may affect the strength of the bank lending channel. For example, information problems are likely to be less binding in periods of relatively loose monetary policy, rendering the bank lending channel much less relevant as a transmission mechanism in comparison with periods of a tighter monetary stance. In particular, the large monetary policy shock in 1998-99 probably represents the bank lending channel operating at its maximum strength, although the counterfactual exercise of what would have happened had the exchange rate been allowed to depreciate sharply points to the possibility of a financial accelerator mechanism as well, through larger balance sheet effects. Another example is the role played by the increase in the capital base of banks during the 1990s, as well as the more widespread use of credit scoring. Both trends have probably strengthened the capacity of banks to deal with informational asymmetries.

This study underscores at least four avenues for future research that may deepen our knowledge of the functioning of the credit channel, in general, and the bank lending channel, in particular, as transmission mechanisms for monetary policy in the Chilean economy: (i) improvements in measuring monetary policy shocks; (ii) improvements in measuring the costs for bank-dependent borrowers associated with a drop in banks' credit supply; (iii) improvements in incorporating the effects of policy changes and financial sector developments; and (iv) improvements in assembling more comprehensive data sets at the microeconomic level.

---

[32] We are using a relatively small data set given the relatively large set of variables included in the VAR system, meaning that we are dealing with large sampling uncertainty. The 95% confidence interval is thus a rather strict test. For instance, researchers tend to use ±1 standard deviation when dealing with large sampling uncertainty, meaning that a 67% confidence interval for the true impulse response function is considered good enough for the purpose at hand (see, for example, Stock and Watson (2001)). If we use the latter benchmark, the macroeconomic relevance of the bank lending channel is unambiguously supported for all variables used as proxies for macroeconomic activity.

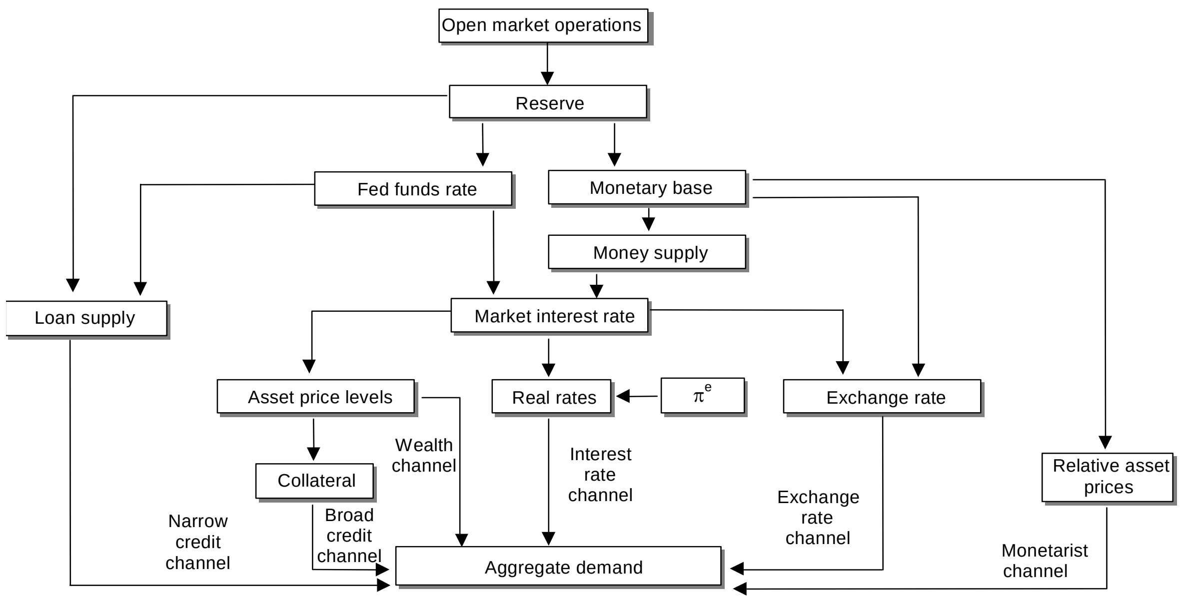
Appendix:
Subchannels of monetary transmission

The different transmission mechanisms of monetary policy can be illustrated by means of the diagram in Figure A1 (Kuttner and Mosser (2002)). The transmission mechanism process begins with the central bank's definition of a monetary policy rate. The interbank rate then converges to this objective through the regulation of the liquidity of the financial system. Once the liquidity is adjusted, different mechanisms start operating in the transmission channel. Four of these are activated by market interest rates moving in tandem with the interbank interest rate. These are the interest rate channel, in which an increase in the cost of capital reduces domestic aggregate demand through a fall in investment and in the consumption of durable goods; the exchange rate channel (in open economies), which operates through the effect of the uncovered interest rate parity on net imports; the asset price channel (stocks, bonds and real estate), which generates a wealth effect that has an impact on consumers' decisions; and the broad credit channel, which is also related to the market value of assets and which is described in the introduction. The transmission mechanism of monetary policy does not end there, however. It is possible to distinguish two additional channels; namely, the monetarist channel related to changes in relative asset prices and the bank lending channel, the main issue of our paper.

Figure A1

Channels of monetary policy transmission

Monetary policy transmission



References

Akerlof, G (1970): "The market for lemons: qualitative uncertainty and the market mechanism", *Quarterly Journal of Economics*, 84(3), pp 488-500.

Alonso-Borrego, C and M Arellano (1999): "Symmetrically normalized instrumental-variable estimation using panel data", *Journal of Business and Economic Statistics*, 17(1), pp 36-49.

Bernanke, B and A Blinder (1992): "The federal funds rate and the channels of monetary transmission", *American Economic Review*, 82(4), pp 901-21.

Bernanke, B and M Gertler (1995): "Inside the black box: the credit channel of monetary policy transmission", *Journal of Economic Perspectives*, 9(4), pp 27-48.

Bernanke, B and I Mihov (1998): "Measuring monetary policy", *Quarterly Journal of Economics*, 113(3), pp 869-902.

Bravo, H and C García (2002): "Medición de la política monetaria y el coeficiente de traspaso (pass-through) en Chile", *Revista Economía Chilena*, 5(3), pp 5-28.

Brock, P and H Franken (2003): "Sobre los determinantes de los spreads marginal y promedio de las tasas de interés bancarias: Chile 1994-2001", *Revista Economía Chilena*, 3(6), pp 445-63.

Caballero, R (2002): "Coping with Chile's external vulnerability: a financial problem", in N Loayza and R Soto (eds), *Economic growth: sources, trend, and cycles*, Santiago, Central Bank of Chile.

Cavieres, L (2002): *Transmisión monetaria: Un estudio del canal del crédito bancario en Chile*, Master's thesis, Universidad de Chile.

Christiano, L J, M Eichenbaum and C Evans (1996): "The effects of monetary policy shocks: evidence from the flow of funds", *Review of Economics and Statistics*, 78(1), pp 16-34.

——— (1997): "Sticky price and limited participation models: a comparison", *European Economic Review*, 41(6), pp 1201-49.

——— (1999): "Monetary policy shocks: what have we learned and to what end?", in J B Taylor and M Woodford (eds), *Handbook of Macroeconomics*, vol 1A, Chap 2, Amsterdam, Elsevier Science.

Diamond, D (1984): "Financial intermediation and delegated monitoring", *Review of Economic Studies*, 51(3), pp 393-414.

Eichenbaum, M and C Evans (1995): "Some empirical evidence on the effects of shocks to monetary policy on exchange rates", *Quarterly Journal of Economics*, 110(4), pp 1975-2010.

Estrella, A and F Mishkin (1998): "Predicting recessions: financial variables as leading indicators", *Review of Economics and Statistics*, 80(1), pp 45-61.

Friedman, M and A Schwartz (1963): *The monetary history of the United States, 1867-1960*, Princeton University Press.

Gertler, M and S Gilchrist (1994): "Monetary policy, business cycles and the behavior of small manufacturing firms", *Quarterly Journal of Economics*, 109(2), pp 309-40.

Gertler, M and C Lown (2000): "The information in the high yield bond spread for the business cycle: evidence and some implications", *NBER Working Papers*, 7549, Cambridge, Mass.

Gilchrist, S and E Zakrajsek (1995): "The importance of credit for macroeconomic activity: identification through heterogeneity", *Conference Series*, 39, Federal Reserve Bank of Boston, pp 129-58.

Goldberg, L, R Cole and L White (2002): "Cookie-cutter versus character: the micro structure of small business lending by large and small banks", *Banking and financial institutions working paper* 9(32), Rochester, NY, Social Science Research Network.

Hahn, J, J Hausman and G Kuersteiner (2002): *Biased-corrected instrumental variables estimation for dynamic panel models with fixed effects*, Massachusetts Institute of Technology.

Hahn, J and G Kuersteiner (2002): "Asymptotically unbiased inference for a dynamic panel model with fixed effects", *Econometrica*, 70(4), pp 1639-57.

Hancock, D and J Wilcox (1998): "The 'Credit Crunch' and the availability of credit to small businesses", *Journal of Banking and Finance*, 22(6-8), pp 983-1014.

Hernando, I and J Martínez-Pagés (2001): "Is there a bank lending channel of monetary policy in Spain?", *working paper* 99, Frankfurt, European Central Bank.

Huang, Z (2003): "Evidence of a bank lending channel in the U.K.", *Journal of Banking and Finance*, 27(3), pp 491-510.

Jadresic, E (1996): "Wage indexation and the cost of disinflation", *IMF Staff Papers*, 43(4), pp 796-825.

Jaffee, D and T Russell (1976): "Imperfect information, uncertainty, and credit rationing", *Quarterly Journal of Economics*, 90(4), pp 651-66.

Johansen, S (1995): *Likelihood-based inference in cointegrated vector autoregressive models*, Oxford University Press.

Kashyap, A and J Stein (1995): "The impact of monetary policy on bank balance sheets", *Carnegie-Rochester Conference Series on Public Policy*, 42, pp 151-95.

——— (2000): "What do a million observations on banks say about the transmission of monetary policy?", *American Economic Review*, 90(3), pp 407-28.

Kishan and Opiela (2000): "Bank size, bank capital, and the bank lending channel", *Journal of Money, Credit and Banking*, 32(1), pp 121-41.

Kuttner, K N and P C Mosser (2002): "The monetary transmission mechanism: some answers and further questions", *Federal Reserve Bank of New York Economic Policy Review*, 8(1), pp 15-26.

Modigliani, F and M Miller (1958): "The cost of capital, corporation finance, and the theory of investment", *American Economic Review*, 48(4), pp 261-97.

Stiglitz, J and A Weiss (1981): "Credit rationing in markets with imperfect information", *American Economic Review*, 71(3), pp 393-410.

Stock, J and M Watson (2001): "Vector autoregressions", *Journal of Economic Perspectives*, 15(4), pp 101-16.

Strogin, S (1995): "The identification of monetary policy disturbance: explaining the liquidity puzzle", *Journal of Monetary Economics*, 35(3), pp 463-97.

Townsend, R (1979): "Optimal contracts and competitive markets with costly state verification", *Journal of Economic Theory*, 21(2), pp 265-93.

Welfare analysis of non-fundamental asset price and investment shocks: implications for monetary policy

Frank Smets and Raf Wouters[1]

1. Introduction

The occurrence of large asset price fluctuations in the late 1980s and early 1990s raised a good deal of discussion among economic researchers and policymakers regarding whether and how central banks should respond to asset price fluctuations. One view (eg Bernanke and Gertler (2000)) suggests that central banks should take into account asset price movements only as far as these fluctuations have an impact on expected future inflation and output. This view also seems to describe fairly well the point of view of many policymakers (eg Greenspan (2002) or Goodfriend in BIS/CEPR (1998)).[2] An alternative view (eg Borio and Lowe (2002)) is that central banks should lean against large run-ups in asset prices, even if this risks undershooting the short-term inflation objective, because excessive asset price booms may lead to a sudden collapse, undermining the stability of the financial system and leading to large negative knock-on effects on output and prices. This view has recently received some support from policymakers (eg Issing (2003)), although a number of difficulties are typically identified. First, the policy-controlled interest rate may only be a very blunt instrument to control asset price bubbles and their inherent risks for future financial stability. Second, policymakers may have no comparative advantage in identifying whether asset prices are driven by fundamentals or not.

As the most recent downturn coincided with a sharp decline in investment expenditures and falling stock markets, the role of asset prices in monetary policy has again become very topical. The over-accumulation of capital in various sectors, associated with the preceding spectacular run-up in stock prices, led to a capital overhang and contributed to the size and the duration of the investment decline. Monetary policy has therefore been accused by some observers of not having paid enough attention to the asset price bubble that developed in the second half of the 1990s.

This paper analyses the costs and benefits of alternative monetary policy responses to non-fundamental asset price or investment shocks in a New Keynesian general equilibrium model. One advantage of using a micro-founded model is that the utility of the representative consumers can be used as a natural benchmark for analysing welfare. The model used is estimated and discussed in Smets and Wouters (2003a) and includes, amongst various other estimated structural shocks, both an investment-specific technology shock and a non-fundamental shock to equity prices. This paper, first, analyses the welfare costs of the non-fundamental equity price shocks when monetary policy is characterised by the estimated policy reaction function. It identifies various components of the welfare cost - inefficient inflation and wage dispersion, the cost of variability in consumption and employment, costs of adjusting investment plans and inefficiencies in the intra- and intertemporal allocation of resources - and discusses their relative importance. One major finding of this analysis is that the welfare cost of the non-fundamental shocks strongly depends on the steady state level around which the economy is fluctuating. If the steady state output level is below the first-best competitive output level, positive booms in economic activity driven by non-fundamental shocks to stock prices can be

[1] Frank Smets: European Central Bank and CEPR, e-mail: Frank.Smets@ecb.int, and Raf Wouters: National Bank of Belgium, e-mail: Rafael.Wouters@nbb.be. The views expressed are solely our own and do not necessarily reflect those of the European Central Bank or the National Bank of Belgium.

[2] This result is also confirmed by empirical research on the Fed's reaction function. Rigobon and Sack (2003) estimate the response of interest rates to equity price innovations, and find that this response seems to correspond with the impact that one can expect from these innovations on future output and inflation. Other policymakers have, however, mentioned that asset prices need some specific attention, for instance because of the imbalance between the time horizon of the typical forecast exercise for inflation and output on the one hand and the long-run implications of financial cycles on the real economy on the other hand (Issing (2003)).

welfare-improving, as they move the economy closer to the optimal output level. In contrast, recessions are extra-costly for the opposite reason.

In a second step, the paper then investigates the costs and benefits of alternative monetary policy rules. One finding is that the welfare costs of asset price shocks can be drastically reduced by a relatively strong response to inflation and the output gap. Another finding is that, in view of the asymmetry in the welfare costs of positive and negative asset price shocks, policymakers can improve welfare by responding less aggressively to booms than to busts. Such a policy will lead to a rise in average output, but at the cost of somewhat higher inflation.

Our analysis is most closely linked to Dupor (2001), who investigates the optimal monetary policy responses to asset price fluctuations under commitment from the perspective of the welfare of the representative household. He analyses the policy trade-off between goods price and asset price stability that arises when asset prices are influenced by inefficient shocks or bubbles and therefore cause inefficient real allocation decisions.[3] Overall, he shows that the optimal response to positive asset price shocks involves an undershooting of inflation in the short term.

A number of papers have analysed actual monetary policy behaviour during and following asset price booms (eg Borio and Lowe (2002) and Detken and Smets (2003)). Overall, asset price booms are characterised by a boom in output and investment and a more moderate increase in inflation. One interpretation of this evidence is that asset price booms tend to develop during periods with positive supply shocks that might increase expectations of future profits and productivity. Generally, periods of asset price booms also seem to be characterised by a relatively weak response of monetary policy (Detken and Smets (2003), Borio and Lowe (2002)). However, often the response to financial cycles is asymmetric: while monetary policy is rather reluctant to intervene in periods of booms, it intervenes much more aggressively in periods of financial crisis. During these periods, it is clear that an intervention of the monetary authorities is needed to stabilise the functioning of the financial markets and to avoid further disruptions in the financial system as a whole.

At the same time, the limitations of the current analysis for understanding the costs of financial volatility and imperfections need to be clearly spelled out. The model used does not contain a specified block for the financial sector. Moreover, the asset price shocks are introduced in an ad hoc and exogenous fashion. A full welfare analysis of the importance of non-fundamental asset price and investment cycles should be based on a model that can endogenously generate such asset price cycles. The optimal policy response may very well depend on the source of the financial market imperfections that lead to such non-fundamental financial and real volatility. One step in that direction has been taken by Bernanke and Gertler (2000). They develop a model in which information problems and capital market imperfections can explain why financial asset prices deviate from fundamentals and exert a specific influence on economic developments.[4] Bernanke and Gertler (2000) nevertheless conclude that a monetary policy that is concentrated on targeting inflation with a strong response on expected inflation and potentially the output gap is the appropriate monetary policy strategy. In their view there is no need to have a specific response to asset prices.[5] However, because the analysis is done in a linearised version of the model, they do not address the policy implications of the non-linear response of the external finance premium to various shocks. Indeed, one argument for a pre-emptive policy response to large asset price booms is that because of collateral constraints the output costs of an asset price collapse are larger than those of an asset price boom (eg Kent and Lowe (1997) and

---

3 Dupor (2002) extends this argument by noting that central banks are confronted with uncertainty and limited information on the nature of the asset price fluctuations. Such uncertainty makes the response of monetary policy to asset price shocks less aggressive. As discussed above, this is a traditional argument used by central bankers to motivate their non-response to rising asset price markets. Advocates for a more proactive policy argue that the uncertainty in evaluating financial markets and asset prices is perhaps not higher than that in interpreting output gaps. Some recent studies have established forecasting methods to evaluate different types of asset and credit market expansions (eg Borio and Lowe (2002)).

4 Bernanke and Gertler (2000) develop a financial accelerator model that generates an impact of financial asset prices mainly via wealth effects on consumption and via net worth or collateral effects on firms' investment decisions. They do not include, however, a direct impact on investment via the non-fundamental asset price. Investment decisions are based on the fundamental value of the projects. In our model the non-fundamental asset price directly influences the investment decision.

5 Cecchetti et al (2000), using a very similar model, draw less unambiguous conclusions. They observe that including a specific reaction to asset prices in the monetary policy rule will cause a higher inflation variability but a lower output variability and the final choice therefore depends on the policymakers' preferences.

Bordo and Jeanne (2002)). The model used in this paper does not capture such asymmetric costs and therefore cannot address the optimal policy response in such a context.

The rest of the paper is structured as follows. In Section 2, the model structure and its estimation are briefly discussed and the effects of a non-fundamental equity price shock are illustrated. Section 3 then presents the welfare costs of such shocks. Finally, Section 4 considers alternative monetary policies. Section 5 concludes.

2. Model structure and estimation results

The model used in this paper is a standard dynamic general equilibrium model with sticky prices and wages and with capital accumulation. The model contains several real and nominal frictions and is augmented with a complete set of structural shocks in order to fit the data. Two of those shocks directly influence investment spending. One captures the influence of technology shocks that affect the production of capital goods or the capital accumulation process. The second is related to shocks in the external financing conditions of the firms and is for simplicity labelled the equity price shock. This last shock should typically take up all the influences on investment expenditures that originate from non-fundamental fluctuations in financial markets or asset prices.

The model does not contain a financial sector and there are no financial frictions or capital market imperfections that might influence the behaviour of households or firms. In general, it is quite difficult to find evidence that financial variables provide significant additional explanatory power for investment expenditures. The type of financial variables that matter for investment seem to vary from country to country and over time. This indicates that the mechanisms at work are complicated and time-varying processes that are not easily modelled. For the time being, it seems acceptable therefore to consider the influence of financial markets and asset prices on the real sector as independent shocks that enter the model exogenously.[6]

In this section we briefly present the structure of the general equilibrium model and the parameter estimates of the model. For a more detailed discussion we refer to Smets and Wouters (2003a). The impulse response function following a non-fundamental investment shock is discussed in detail.

2.1 Model structure

In what follows we briefly explain the structure of the dynamic stochastic general equilibrium (DSGE) model, which is a standard New Keynesian general equilibrium model with monopolistic competition in the goods and labour market. Prices and wages are sticky and determined by a Calvo model that allows for indexation to past inflation levels for these price and wages that are not reset optimally. Nominal stickiness and indexation were estimated to be important. Capital accumulation is subject to adjustment costs that are expressed in terms of changes in the investment level. Household utility is characterised by habit persistence. These three features of the model will be important in the calculation and the evaluation of the welfare outcomes.

2.1.1 The household sector

Households maximise the following welfare function:

$$E_0 \sum_{t=0}^{\infty} \beta^t \varepsilon_t^b \left(\frac{1}{1-\sigma_c} \left(C_t - hC_{t-1} \right)^{1-\sigma_c} - \frac{\varepsilon_t^L}{1+\sigma_l} \left(\ell_t \right)^{1+\sigma_l} \right) \tag{1}$$

---

6 The ideal solution would be to have a model that is able to generate the bubble process endogenously. Gilchrist et al (2002) have recently developed a model where an increase in the dispersion of investors' beliefs under a short-selling constraint can result in a rise of the stock price above the fundamental value. The model predicts that managers will react to such an event by issuing new equity and increasing capital expenditures. Using the variance in the earnings forecasts to identify the bubble shocks in the asset price, they find that such orthogonalised bubble shocks have significant effects on Tobin's Q and real investment.

where β is the discount factor, ε_t^B and ε_t^L are the two preference shocks and the instantaneous utility function is separable in consumption, relative to the past consumption level reflecting the habit in preferences,[7] and labour effort. σ_c is the coefficient of relative risk aversion of households and σ_l represents the inverse of the elasticity of work effort with respect to the real wage.

Households maximise their objective function subject to the intertemporal budget constraint. Households' total income is given by the sum of wage income, rental returns on capital corrected for the costs related to the degree of capital utilisation and dividend payments. Total income is used for consumption or investment expenditures:

$$Y_t = w_t l_t + r_t^k z_t K_{t-1} - \Psi(z_t)K_{t-1} + Div_t = C_t + I_t \tag{2}$$

Utility maximisation results in first-order conditions for consumption:

$$E_t\left[\beta\frac{\lambda_{t+1}}{\lambda_t}\frac{1+R_t}{\pi_{t+1}}\right] = 1 \tag{3}$$

which states that the marginal rate of intertemporal subsitution should equal the real interest rate. The marginal utility of consumption λ_t is given by:

$$\lambda_t = \varepsilon_t^b\left(C_t - hC_{t-1}\right)^{-\sigma_c} - \beta\varepsilon_{t+1}^b h\left(C_{t+1} - hC_t\right)^{-\sigma_c} \tag{4}$$

Households own the capital stock that they rent out to the firm-producers of intermediate goods at a given rental rate of r_t^k. Households choose the capital stock, investment and the utilisation rate in order to maximise their intertemporal objective function subject to the intertemporal budget constraint and the capital accumulation equation, which is given by:

$$K_t = K_{t-1}[1-\tau] + \varepsilon_t^I\left[1 - S(I_t/I_{t-1})\right]I_t\,, \tag{5}$$

where I_t is gross investment, τ is the depreciation rate and $S(.)$ the adjustment cost function, which is a positive function of changes in investment level. Fluctuations in the investment level will result in a higher adjustment cost, leading to lower net investment accumulation. The process ε_t^I represents shifts in investment-specific technological progress. This fundamental shock to the investment decision process is assumed to follow a first-order autoregressive process with an iid normal error term: $\varepsilon_t^I = \rho_I\varepsilon_{t-1}^I + \eta_t^I$.

The first-order conditions for capital, investment and the utilisation rate are given by:

$$Q_t = E_t\left[\beta\frac{\lambda_{t+1}}{\lambda_t}\left(Q_{t+1}(1-\tau) + z_{t+1}r_{t+1}^k - \Psi(z_{t+1})\right)\right]\varepsilon_t^P\,, \tag{6}$$

$$Q_t\left(1 - S\left(\frac{I_t}{I_{t-1}}\right) - S'\left(\frac{I_t}{I_{t-1}}\right)\frac{\varepsilon_t^I I_t}{I_{t-1}}\right) + \beta E_t Q_{t+1}\frac{\lambda_{t+1}}{\lambda_t}S'\left(\frac{I_{t+1}}{I_t}\right)\left(\frac{\varepsilon_{t+1}^I I_{t+1}}{I_t}\right)\frac{I_{t+1}}{I_t} = 1 \tag{7}$$

$$r_t^k = \Psi'(z_t) \tag{8}$$

Equation (6) states that the value of installed capital Q is equal to the discounted value of the expected future returns as captured by the rental rate times the expected rate of capital utilisation minus the utilisation costs. The value of installed capital is also influenced by an exogenous iid shock which we label the equity premium shock. Equation (7) determines the optimal investment level given

[7] In the welfare calculations we assumed the habit persistence is expressed relative to the household-specific past consumption level. In the estimated model, the habit preference was expressed in terms of the aggregate wide past consumption level. For the empirical estimation of the model the difference between the two models is not important. In the welfare evaluation, the external habit persistence yields quite complicated results because of the externality effects. By retaining the internal habit specification we avoid these problems.

the value of installed capital and the investment adjustment cost function. Equation (8) relates the optimal degree of capital utilisation to the rental rate.

Finally, households also supply labour effort and set the wage rate. Wages are set according to the Calvo model allowing for a partial indexation to the previous period's inflation level.

This maximisation problem results in the following markup equation for the optimal wage:

$$\frac{\tilde{w}_t}{P_t} E_t \sum_{i=0}^{\infty} \beta^i \xi_w^i \left(\frac{\left(P_{t+i-1}/P_{t-1}\right)^{\gamma_w}}{P_{t+i}/P_t} \right) \frac{l_t^l +_i U_{t+i}^C}{1+\lambda_{w,t+i}} = E_t \sum_{i=0}^{\infty} \beta^i \xi_w^i l_{t+i}^l U_{t+i}^{\ell} \tag{9}$$

where U_{t+i}^{ℓ} is the marginal disutility of labour, U_{t+i}^{C} is the marginal utility of consumption, γ^w is the degree of indexation, ξ^w the Calvo probability and λ_w the markup included in wages. Equation (9) shows that in a flexible wage context, this equation would simplify to the traditional condition that wages equal a markup over the marginal disutility of work divided by the marginal utility of consumption. The aggregate wage process is described by:

$$\left(W_t\right)^{-1/\lambda_{w,t}} = \xi_w \left(W_{t-1} \left(\frac{P_{t-1}}{P_{t-2}} \right)^{\gamma_w} \right)^{-1/\lambda_{w,t}} + \left(1-\xi_w\right)\left(\tilde{w}_t\right)^{-1/\lambda_{w,t}} \tag{10}$$

reflecting the Dixit-Stiglitz aggregator function to define the aggregate labour supply index.

2.1.2 The firm sector

Output in the intermediate goods sector is produced by the following technology:

$$y_t^j = \varepsilon_t^{\alpha} \tilde{K}_{j,t}^{\alpha} L_{j,t}^{1-\alpha} - \Phi\,, \tag{11}$$

where ε_t^{α} is the productivity process, $\tilde{K}_{j,t}$ is the effective utilisation of the capital stock given by $\tilde{K}_{j,t} = z_t K_{j,t-1}$, $L_{j,t}$ is an index of different types of labour used by the firm and Φ is a fixed cost. Capital is assumed to be perfectly mobile between firms within each period. Cost minimisation implies that the income shares are constant:

$$\frac{W_t L_{j,t}}{r_t^k \tilde{K}_{j,t}} = \frac{1-\alpha}{\alpha} \tag{12}$$

Under these assumptions the firms' marginal cost is independent of the production level and only a function of the factor prices and productivity level:

$$MC_t = \frac{1}{\varepsilon_t^a} W_t^{1-\alpha} r_t^{k^{\alpha}} \left(\alpha^{-\alpha} \left(1-\alpha\right)^{-(1-\alpha)} \right) \tag{13}$$

Firms set prices according to the Calvo model with partial indexation:

$$E_t \sum_{i=0}^{\infty} \beta^i \xi_p^{\,i} \lambda_{t+i} y_{t+i}^j \left(\frac{\tilde{p}_t^j}{P_t} \left(\frac{\left(P_{t-1+i}/P_{t-1}\right)^{\gamma_p}}{P_{t+i}/P_t} \right) - \left(1+\lambda_{p,t+i}\right) mc_{t+i} \right) = 0 \tag{14}$$

where γ_p is the degree of indexation, ξ_p the Calvo probability and λ_p the markup incorporated in the price.

The law of motion of the aggregate price index is given by:

$$\left(P_t\right)^{-1/\lambda_{p,t}} = \xi_p \left(P_{t-1} \left(\frac{P_{t-1}}{P_{t-2}} \right)^{\gamma_p} \right)^{-1/\lambda_{p,t}} + \left(1-\xi_p\right)\left(\tilde{p}_t^j\right)^{-1/\lambda_{p,t}} \tag{15}$$

2.1.3 *The central bank*

The monetary authorities follow a generalised Taylor rule by gradually responding to deviations of lagged inflation from an inflation objective (normalised to be zero) and the lagged output gap defined as the difference between actual and potential output (Taylor (1993)). Consistently with the DSGE model, potential output is defined as the level of output that would prevail under flexible prices and wages.

$$R_t = \rho R_{t-1} + (1-\rho)\left\{\overline{\pi}_t + r_\pi(\pi_{t-1} - \overline{\pi}_t) + r_Y\left(Y_t - Y_t^p\right)\right\} + r_{\Delta\pi}(\pi_t - \pi_{t-1}) + r_{\Delta y}\left(Y_t - Y_t^p - \left(Y_{t-1} - Y_{t-1}^p\right)\right) + \eta_t^R \qquad (16)$$

The parameter ρ captures the degree of interest rate smoothing. In addition, there is also a short-run feedback from the current changes in inflation and the output gap. η_t^R and $\overline{\pi}_t$ are two monetary policy shocks: the first one represents the typical iid interest rate shocks, while the second one captures the long-run trends in the inflation objective of the central bank.

2.2 Estimation results and evidence on the non-fundamental investment shock

Smets and Wouters (2003a) estimate a linearised version of the model discussed above. The parameter estimates are summarised in Table 1. For estimation purposes, a linear approximation is sufficient, because the impact of the different identified shocks over a finite horizon is not significantly influenced by the higher-order terms. Of course, as discussed in Kim et al (2003), this argument does not apply for the welfare analysis performed in the next section.

The left-hand column of Table 1 contains the estimated parameters describing the behaviour of the stochastic shocks in the model. Smets and Wouters (2003a) estimate a whole series of shocks that can potentially influence the economy: a shock to total factor productivity, a shock to the intertemporal time preference of households, a shock to the relative weight of consumption and labour supply in the utility function, a government expenditures shock and a shock to the investment adjustment cost function (or to the capital good-specific technology). These five fundamental shocks to technology or preferences are assumed to follow a persistent first-order autoregressive process. In addition, Smets and Wouters (2003a) also allow for three markup shocks that affect the pricing in the goods market, the labour market and the market for existing capital goods. These three shocks produce inefficient price and allocation decisions and are assumed to be iid.[8]

The analysis in this paper concentrates on the latter of those three markup shocks, the inefficient equity price shock, which creates non-fundamental movements in investment expenditures. This iid shock, which can take a positive or negative sign, is of a somewhat different nature than the much more persistent asset price bubble shocks that are typically considered in the research on monetary policy and asset prices.[9] However, it has the same qualitative effects on output, investment and inflation as those shocks. As discussed in the introduction, a more sophisticated approach would model the underlying distortions that generate the bubble and the way firms react to such non-fundamental movements (see Gilchrist et al (2002) for such a model).

[8] We motivate this identification sheme in Smets and Wouters (2003b). Under uncertainty about the nature of the shocks, a robust discretionary monetary policy will favour interpreting persistent shocks as fundamental shocks that affect the natural output level and therefore need to be accommodated. Short-run fluctuations that do not seem to produce a persistent effect can be excluded in the estimation of the natural or efficient output level without creating risks of large errors. This implies that a persistent negative shock to the investment expenditures will be considered to have a negative effect on the natural output level the central bank is targeting. If the central bank were to consider it wrongly as an inefficient low investment level, and react by lowering the interest rate, this would lead to a rise in inflation and inflation expectations that would be very costly to overcome later. Under discretion, a more careful conservative monetary policy is beneficial. This argument is, however, less applicable for shocks that are less or not persistent. Therefore iid shocks can be classified as non-efficient shocks.

[9] Bernanke and Gertler (2000), Cecchetti et al (2000) and Dupor (2002) all consider persistent asset price bubbles, with or without a random duration. As in our case, the shocks are, however, introduced in an exogenous and ad-hoc fashion.

Table1

Estimated parameters of the DSGE model

Parameters defining shock processes		Parameter describing private agents	
Standard errors of the innovations:			
Productivity shock	0.59	Investment adjustment cost	5.91
Inflation objective shock	0.02	σ consumption utility	1.61
Consumption preference shock	0.25	h consumption habit	0.54
Government spending shock	0.32	σ labour utility	0.75
Labour supply shock	1.35	Fixed cost	1.49
Investment shock	0.10	Calvo employment	0.59
Interest rate shock	0.12	Capital utilisation adjustment cost	0.17
Equity premium shock	0.60		
Price markup shock	0.16	Calvo wages	0.76
Wage markup shock	0.28	Calvo prices	0.91
		Indexation wages	0.66
		Indexation prices	0.41
Persistence of the processes:		**Parameter describing monetary policy rule:**	
Productivity shock	0.83	r inflation	1.66
Inflation objective shock	0.92	r d(inflation)	0.20
Consumption preference shock	0.91	r lagged interest rate	0.94
Government spending shock	0.97	r output	0.15
Labour supply shock	0.96	r d(output)	0.17
Investment shock	0.94		

Source: Smets and Wouters (2003a).

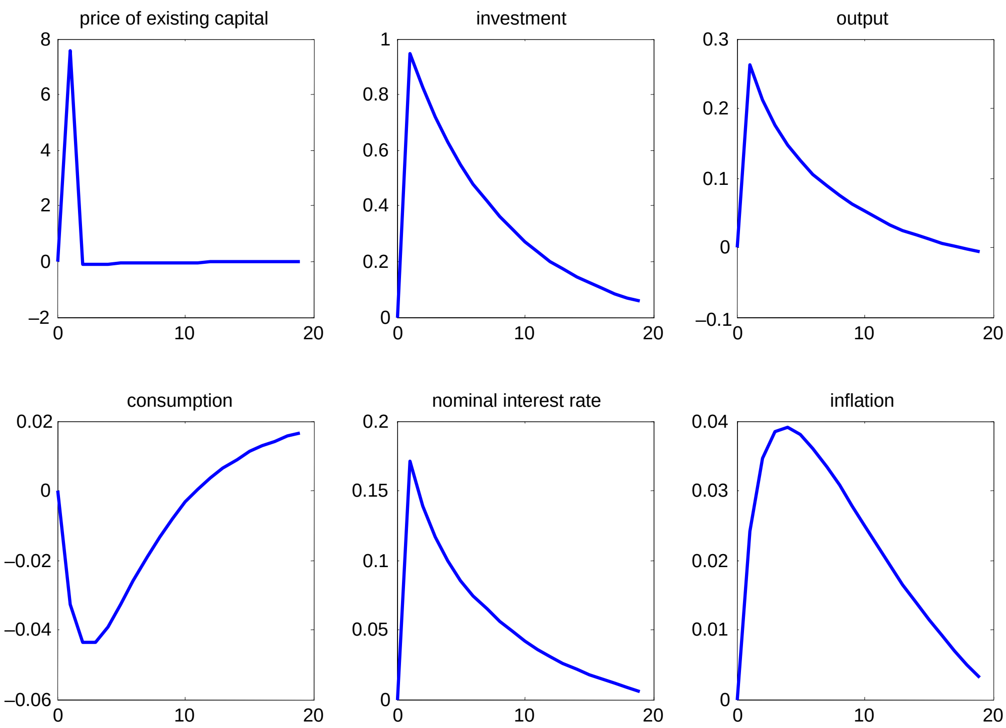
In Graph 1, we reproduce the impulse response of the non-fundamental investment shock using the non-linear model.[10] It is worth noting that this impulse response is very close to one in the estimated linear version of the model.

The shock immediately affects the price of installed capital, but due to its temporary nature only for one quarter. The price of existing capital increases by some 7% for a one standard error shock. Firms react immediately to the higher value of existing capital stock by increasing investment expenditures. The presence of capital accumulation costs in the form of changes in the level of investment implies that investment will only gradually return to its steady state level. Investment expenditures increase by 1% for the average shock and the shock dies out completely after four or five years.

[10] The non-linear model is solved under the assumption of perfect foresight using Dynare (Julliard (2003)). For the deterministic simulations Dynare uses a Newton-type algorithm.

Graph 1

Impulse response function following the non-fundamental investment shock in the non-linear model



Higher investment expenditures increase total aggregate demand by 0.2% and aggregate employment by 0.1%. The positive output gap will lead to an increase in the marginal cost as a consequence of rising wages and lower productivity. The impact on inflation is limited for several reasons. First, the estimated degree of nominal stickiness is relatively large. Second, monetary policy responds relatively strongly to the positive output gap. This restrictive policy reaction will create a crowding-out effect on private consumption, which lowers the overall aggregate demand expansion. Lower consumption also lowers the pressure on wage demands via the higher marginal utility of wages. Finally, the investment expansion also contributes to production capacity, increasing labour productivity. Summing up, the non-fundamental equity price shock increases investment and output significantly over a horizon of two to three years, but under the estimated monetary policy response the impact of the shock on inflation is very moderate. Although the size and the persistence of the effect of our shock on asset prices is not comparable to the much more persistent movements in asset prices during typical asset price booms, the qualitative effects are relatively similar to those of a standard asset price bubble as, for example, described in Borio and Lowe (2002) and Detken and Smets (2003).

Smets and Wouters (2003a) discuss the contribution of the various shocks to unconditional variance of the forecast errors in the observable variables. This variance decomposition indicates that the non-fundamental investment shocks explain around half of the forecast error of investment at the one quarter ahead horizon, but this contribution decreases very quickly for longer horizons. The contribution to the one quarter ahead forecast error in output is between 10 and 20% and also decreases quickly afterwards. The low persistence in the effects also explains why the contribution to the inflation process is very small. A historical decomposition (Smets and Wouters (2003c)) nevertheless shows that during specific periods the shocks have a significant impact on investment and output, but not on inflation. At longer forecast horizons, the fundamental investment shocks explain most of the fluctuations in investment and around 20% of output fluctuations. However, it is important to note that it is very difficult to distinguish the fundamental (persistent) from the non-fundamental (temporary) shocks, in particular because equity prices were not used in the estimation of the model. As the empirical identification is purely based on whether the shocks are persistent or not, one could also treat the persistent investment shock as non-fundamental. Obviously, this would increase the role of non-fundamental equity price shocks. Ultimately, a more realistic estimate of the importance of non-fundamental asset price shocks needs to be obtained by including information from asset prices in the estimation of the model.

3. The welfare implications of non-fundamental investment shocks

Non-fundamental equity price and investment shocks create several types of inefficiencies. First of all, they result in an inefficient intertemporal allocation of resources. An overestimation of the present value of the future returns from current investment expenditure leads to an over-accumulation of capital. The actual return on capital will not compensate for the forgone utility from present consumption. Second, positive demand effects from an asset price and investment shock lead to positive inflation in prices and wages. In our Calvo model this creates welfare costs through the dispersion in prices and wages and the resulting misallocation of resources among firms in the monopolistically competitive sector. Different prices and wages for otherwise similar products result in a lower consumption or labour bundle for a given nominal budget. Inflation also implies that prices deviate from the marginal cost plus markup. Finally, there are the costs of changing investment plans.

In general, these welfare costs will create a trade-off problem for optimal monetary policy. As shown in Dupor (2002), inflation stabilisation can more or less be obtained by setting the interest rate so as to stabilise total aggregate demand. However, stabilising the equity price and the resulting investment response will typically require a more restrictive policy and a larger crowding-out of other private expenditures. This will lead to an undershooting of the short-run inflation response. In deciding how strongly to respond to the non-fundamental investment shock, it is therefore important to have an idea of the relative size of the different costs that are involved.

The relative importance of these different costs is dependent on the steady state situation around which the fluctuations occur. If the steady state is around the optimal competitive output level, all non-fundamental fluctuations, both positive and negative, will be costly. However, if output is far below the efficient output level due to the markup distortion, higher demand can move the output level towards the first-best level and this generates welfare gains. These welfare gains have to be balanced against the rise in inflation that may result from an asymmetric response to the equity price shocks, further complicating the welfare analysis. Dupor (2001) studies the impact of a deterministic non-fundamental shock on welfare around the efficient steady state output level. He analyses the problem in a model with monopolistic competition and markup pricing, but he introduces an output subsidy financed by a lump sum tax, so that the steady state output equals the competitive level.

In the next section, we first calculate the welfare effects of a deterministic non-fundamental equity price shock. Given the identification problem discussed above, we analyse the effects of both the temporary and persistent investment shock. The latter type of shock compares well to the typical bubble shocks that are considered in Dupor (2002) and Bernanke and Gertler (2001). For comparison reasons, we also report the welfare effects of a fundamental investment shock that is caused by a change in the relative price of capital goods. For each of these three types of shocks, we study the welfare effects around the competitive equilibrium steady state output level and around the lower monopolistic competition equilibrium. We try to disentangle the different components of the welfare effects and show how the different frictions influence the relative size of the welfare effects. Next, we discuss the outcomes from a stochastic simulation exercise, based on a second-order approximate solution of the model. Also in this case, we calculate the different components of the welfare loss.[11]

[11] The welfare evaluation is based on the exact perfect foresight solution to the non-linear first-order equations for the deterministic shocks and on the second-order approximation solution of the model for stochastic simulations. These calculations were performed using Dynare (Julliard (2003), Schmitt-Grohe and Uribe (2002)).

3.1 Welfare analysis of a deterministic non-fundamental investment shock around the competitive equilibrium (CE) output level

Table 2 summarises the results for each of the three types of shocks around the CE output level. The first shock corresponds to the estimated temporary equity premium shock in Smets and Wouters (2003a) (illustrated in Graph 1). The shock has a standard error of 0.08. The effects of a positive and a negative shock are reported for later use when discussing issues of asymmetry.

Overall, the impact on welfare of this shock is small. This is not surprising as all the first-order conditions are fulfilled around the CE output level and therefore small disturbances do not create large inefficiencies. To assess the size of the impact on welfare, we follow the literature and express the change in welfare in terms of consumption equivalents. We calculate the change in certainty-equivalent consumption in percentage of its steady state level that yields exactly the same variation in the expected lifetime utility that follows from the shock. Since we consider one-time deterministic shocks in this exercise, we also express the consumption effect as a percentage of a one-period consumption level. The benchmark non-fundamental investment shock has an impact on welfare that is comparable to a 0.02% change in the consumption level.

Table 2

Welfare effects of a distortionary investment shock around the CE output level

	Iid shock		**Persistent shock**		**Fundamental shock**	
	+ shock	**– shock**	**+ shock**	**– shock**	**+ shock**	**– shock**
Total welfare effect	–0.0003	–0.0004	–0.0005	–0.0005	0.0725	–0.0721
In % of steady state consumption level	–0.0176 100.00%	–0.0239 100.00%	–0.0332 100.00%	–0.0312 100.00%	4.5317 100.00%	–4.4992 100.00%
Price dispersion cost	–0.0007 3.99%	–0.0010 4.12%	–0.0019 5.87%	–0.0019 6.18%	–0.0027 –0.06%	–0.0026 0.06%
Wage dispersion cost	–0.0011 6.19%	–0.0015 6.38%	–0.0037 11.14%	–0.0036 11.63%	–0.0015 –0.03%	–0.0014 0.03%
Capital adjustment cost	–0.0072 40.70%	–0.0098 40.83%	–0.0035 10.44%	–0.0035 11.21%	–0.0037 –0.08%	–0.0038 0.08%
Variance cost	–0.0042 24.09%	–0.0059 24.78%	–0.0149 44.77%	–0.0147 47.21%	–0.0161 –0.36%	–0.0162 0.36%
Intra-/intertemporal inefficiency	–0.0044 25.03%	–0.0057 23.89%	–0.0092 27.77%	–0.0074 23.77%	4.5555 100.53%	–4.4752 99.47%

The second column reports the welfare effects of the more persistent shock, which corresponds to the persistent investment shock in Smets and Wouters (2003a). This shock has a much more persistent and hump-shaped effect on investment and output and is very similar to the shock considered in Dupor (2002). Taking into account that the shock considered in Dupor (2002) is some five times bigger, the welfare effects of the shocks are somewhat smaller in our setup, but the size is of the same magnitude. Differences are partly due to differences in the modelling of the investment adjustment cost function and the habit persistence process.

Table 2 also decomposes the welfare effects into the most important elements. First of all, there is the cost of inflation measured by the degree of price and wage dispersion. This cost is estimated by using the index for price and wage dispersion (similar to the expression presented in Benigno and Woodford (2003)). The expression for wage dispersion is:

$$\Lambda_t^w = \xi_w \Lambda_{t-1}^w \pi_{w,t}^{\theta(1+\sigma l)} \pi_{p,t-1}^{-\gamma w \theta(1+\sigma l)} + (1-\xi_w) * \left(\frac{1-\xi_w \pi_{w,t}^{\theta-1} \pi_{p,t-1}^{-\gamma w(\theta-1)}}{1-\xi_w} \right)^{-\frac{\theta(1+\sigma l)}{1-\theta}} \tag{17}$$

where θ is the price elasticity of demand, which is itself related to the markup $1+\lambda w = \theta/(\theta-1)$.

The moderating impact of partial indexation on the dispersion measure is clear from this expression. The corresponding equation for price dispersion is:

$$\Lambda_t^p = \xi_p \Lambda_{t-1}^p \pi_{p,t}^{\theta(1+\sigma l)} \pi_{p,t-1}^{-\gamma p\theta} + (1-\xi_p) * \left(\frac{1-\xi_p \pi_{p,t}^{\theta-1} \pi_{p,t-1}^{-\gamma p(\theta-1)}}{1-\xi_p} \right)^{-\frac{\theta}{1-\theta}} \tag{18}$$

These dispersion measures appear in the aggregate utility function as a cost that augments the input of labour to produce the given aggregate output of consumption goods:

$$U_t = \varepsilon_t^b \left(\frac{1}{1-\sigma_c} (C_t - hC_{t-1})^{1-\sigma_c} - \frac{\varepsilon_t^L}{1+\sigma_l} (L_t)^{1+\sigma_l} * \Delta_t^w * \Delta_t^{p^{(1+\sigma l)/(1-\alpha)}} \right) \tag{19}$$

The size of these inflation dispersion costs taken together only makes up some 10% of the total welfare cost. This relatively small size is somewhat surprising especially within the framework of a Calvo model. Erceg and Levin (2002) have stressed that the Calvo model produces very large welfare effects of price stickiness, compared for instance to the Taylor-type stickiness with fixed duration contracts. Rotemberg and Woordford (1997) also find a very high coefficient on the inflation dispersion term in their second-order approximation of the welfare function. In our model, indexation to past inflation and habit persistence in the utility function reduce the relative weight of inflation dispersion in this approximation. The impact of partial indexation to past inflation on the inflation dispersion costs can easily be evaluated. Keeping all other parameters constant, the assumption of indexation reduces the welfare costs of price and wage dispersion in our model by half. A more important explanation for the small inflation costs is the very mild response of inflation following this type of non-fundamental investment shock. As explained above, this is due to the estimated monetary policy rule together with the flexible technology assumptions.

The second important component of the welfare loss refers to the adjustment costs that have to be incurred when firms change their investment plans. These costs take the form of a fraction of investment expenditures that does not result in an increase in the capital stock. The higher the volatility of the investment flows, the higher the fraction of investment that will be lost. These investment adjustment costs account for 40% of the total welfare cost following the temporary equity price shock and for about 10% following the more persistent investment shock.

A third component of the welfare cost that can be identified is the loss that results from the variance in the consumption and labour supply flow. We calculate this component from the second-order approximation to the utility function:

$$0.5 * (1-h)\overline{C} * \left((1-h)\overline{C}\right)^{-\sigma} (1+\sigma)\hat{C}_t^2 + cte * 0.5 * \overline{L} * \overline{L}^{\sigma_l} (1+\sigma_l)\hat{L}_t^2 \tag{20}$$

Finally, the remaining loss is due to inefficiencies in both the intra- and intertemporal allocation of resources. Intratemporal inefficiencies are caused by the frictions in prices and wages, which imply that prices and wages do not reflect the marginal cost of production or the marginal disutility from labour effort. The intertemporal inefficiencies are caused by the non-fundamental shock as discussed above.[12] The variance terms and the remaining first-order inefficiencies explain about 25% of the total welfare cost.

For the more persistent shock the composition of the welfare loss changes slightly. Inflation raises relatively more under the persistent shock and the contribution in the costs is therefore somewhat higher. The same applies for the responses in consumption and labour and this increases the variance term. The more persistent shock is better anticipated by definition and therefore creates less volatility in investment and less capital adjustment costs.

[12] Both components could be identified if we were to consider the impact of the shock in the flexible price-wage model. However, the overall impulse response function of the shock changes strongly in the flexible price model and this makes the comparison less interesting.

The fundamental investment shock, caused by a persistent shift in the relative price of the capital goods, produces a totally different picture. The welfare effects of such a shock depend of course on the sign of the shock: a positive shock implies a temporal increase in the productivity of the capital good producing sector and therefore leads to an expansion of the production potential of the economy. The size of the welfare effect is much higher compared to the costs discussed above. Of course, over time positive and negative shocks cancel each other out and therefore the welfare implications of these shocks have to be analysed in a stochastic simulation. This analysis will be performed in the next section.

3.2 Welfare analysis of a deterministic non-fundamental investment shock around an inefficiently low (MCE) output level

Now we turn to the discussion of the welfare effects of a non-fundamental investment shock around an inefficiently low steady state level of output caused by the markups in a monopolistic competitive world. The welfare effects of the non-fundamental shock are strongly asymmetric under this assumption and the effect of a positive shock on welfare even turns out to be positive. A positive shock increases the output level and employment. Nominal stickiness prevents prices and wages from adjusting quickly to the higher marginal costs and marginal disutility levels, so that the markups are temporally reduced. This will move the economy towards the efficient output level that would prevail in the absence of markup distortions. In the estimated model, these welfare gains turn out to be much higher in magnitude than the costs from inflation, capital adjustment or increased variances.

Table 3

Welfare effects of a distortionary investment shock around the lower MCE output level

	Iid shock		Persistent shock		Fundamental shock	
	+ shock	– shock	+ shock	– shock	+ shock	– shock
Total welfare effect	0.0095	–0.0117	0.0297	–0.0304	0.0932	–0.0927
In % of steady state consumption level	0.5921 100.00%	–0.7307 100.00%	1.8569 100.00%	–1.8974 100.00%	5.8208 100.00%	–5.7855 100.00%
Price dispersion cost	–0.0002 –0.04%	–0.0003 0.04%	–0.0006 –0.03%	–0.0006 0.03%	–0.0007 –0.01%	–0.0007 0.01%
Wage dispersion cost	–0.0003 –0.04%	–0.0004 0.05%	–0.0014 –0.08%	–0.0014 0.07%	–0.0007 –0.01%	–0.0006 0.01%
Capital adjustment cost	–0.0071 –1.21%	–0.0098 1.34%	–0.0033 –0.18%	–0.0033 0.17%	–0.0035 –0.06%	–0.0036 0.06%
Variance cost	–0.0012 –0.21%	–0.0017 0.24%	–0.0068 –0.37%	–0.0068 0.36%	–0.0076 –0.13%	–0.0076 0.13%
Intra-/intertemporal inefficiency	0.6010 101.49%	–0.7185 98.34%	1.8689 100.65%	–1.8854 99.37%	5.8333 100.22%	–5.7729 99.78%

The welfare gain from a positive non-fundamental shock in the benchmark case is similar to a 0.6% increase in the steady state consumption level. The cost of a negative shock is somewhat larger because all the welfare effects go in the same direction but also because of the concave relation between welfare and output, which implies that the welfare costs are increasing as one moves further and further away from the first-best output level.

Gali et al (2001) derive similar welfare effects from business cycle fluctuations that are driven by stochastic movements in the inefficient wage markup. If business cycle fluctuations are associated with variations in economic efficiency, they show that periods of booms imply lower inefficiency and therefore higher welfare, while recessions are leading to lower efficiency and welfare losses. These welfare losses of recessions are higher than the welfare gains of booms because of the concave relationship between welfare and their efficiency gap measure. They also indicate that these welfare costs are potentially important compared to the traditional costs from efficient fluctuations around the competitive steady state level. However, they do not discuss fully the implications for monetary policy that follow from these asymmetric welfare effects.

3.3 Welfare analysis: the stochastic case

In order to approximate the welfare effects in the stochastic case we use a second-order approximation to the model solution.[13] We compare again the welfare results around the CE efficient steady state output level and the lower MCE output level.

Table 4

Welfare effects of a distortionary iid investment shock in a stochastic simulation

	Steady state output CE	**Steady state output MCE**
Total welfare effect	–0.1020	–0.0847
In % of steady state consumption level	–6.3052 100.00%	–5.2347 100.00%
Price dispersion cost	–0.1013 1.61%	–0.0338 0.65%
Wage dispersion cost	–0.1442 2.29%	–0.0343 0.66%
Capital adjustment cost	–1.4341 22.74%	–1.4321 27.36%
Variance cost	–0.5205 8.25%	–0.1558 2.98%
Intra-/intertemporal inefficiency	–4.1051 65.11%	–3.5787 68.36%

The welfare effects of both exercises are very similar. The temporary non-fundamental shocks generate a welfare loss that is equivalent to around 5% of the steady state output level (one period). Price and wage dispersion and the variance term make up only a small fraction of this cost. Capital adjustment costs explain 25 to 30% of the cost and the linear inefficiency term explains the remaining 60-65%. This high proportion of the cost that is related to the inefficiencies caused by the investment shock suggests that a monetary policy that takes into account the non-symmetric welfare effects of the shock might have a substantial impact on these welfare costs. This point will be further analysed in the next section.

4. Welfare implications from alternative monetary policy responses to the non-fundamental investment shock

The previous welfare analysis assumed that monetary policymakers were following the estimated generalised Taylor rule. In this section, we perform stochastic simulations assuming alternative monetary policy rules in order to analyse the impact of monetary policy behaviour on the welfare effects of the shock.[14] Again we start by assuming, first, that the economy is fluctuating around the

[13] We performed these calculations with Dynare (Julliard (2003)) using the Schmitt-Grohe and Uribe (2001) algorithm for the second-order approximation solution.

[14] We leave an analysis of the optimal monetary policy response for future research.

efficient competitive economy output level. This exercise will allow us to compare our results with the discussion in the literature on how monetary policy should react to asset price shocks. Next, we consider the same exercise around the lower monopolistic competitive equilibrium (MCE) output level and discuss how this affects the implications for monetary policy.

4.1 Monetary policy and non-fundamental investment shocks around the CE output level

Under these assumptions, optimal monetary policy from a welfare perspective is faced with a trade-off between stabilising inflation and stabilising investment. Stabilising investment will imply a stronger reaction to the non-fundamental shock, so that other private expenditures are crowded out further and inflation will become negative. In order to illustrate the impact of monetary policy on the welfare outcome, we consider some simple policy rules starting with a rule that responds only to inflation.

The simple policy rule with a very moderate response to inflation (a coefficient of 1.1) does a poor job in terms of welfare outcome. Under this rule, the standard deviation in the inflation process is twice as high as under the more aggressive inflation policies, and this increases the welfare costs of the price and wage dispersion by a factor of four or more. However, all components of the welfare cost increase under the weak inflation policies. A stricter anti-inflation policy (with a reaction coefficient of 1.7) not only reduces the cost of inflation but also helps to overcome part of the other inefficiencies related to the non-fundamental investment shocks. Augmenting this rule with a reaction to the output gap (to 0.5 as in the traditional Taylor rule) further reduces the efficiency costs. These outcomes confirm the results presented by Bernanke and Gertler (2000). The estimated policy rule, which is close to a first difference rule with a relatively strong coefficient on inflation, performs reasonably well in terms of the welfare implications.

The next step would be to evaluate whether the inclusion of a specific response to the price of installed capital in the policy rule might improve the outcome in the fully stochastic model with multiple sources of disturbances. However, with larger and more persistent shocks in the model, the second-order approximation methods often generate unstable solution paths.[15]

To take into account the possible complications that arise due to the non-linearity of the model, we also consider the estimated policy rule augmented with an asymmetric reaction on the growth rate. The asymmetric policy rule that we consider is of the following type:

$$\begin{aligned} R_t &= \rho R_{t-1} + (1-\rho)\left\{\overline{\pi}_t + r_\pi\left(\pi_{t-1} - \overline{\pi}_t\right) + r_Y\left(Y_t - Y_t^p\right)\right\} + \\ &\quad r_{\Delta\pi}\left(\pi_t - \pi_{t-1}\right) + \left(r_{\Delta y}/\kappa\right)\left(1 - \exp\left(\kappa * \left(Y_t - Y_{t-1}\right)\right)\right) + \eta_t^R \end{aligned} \tag{21}$$

The linear impact of output growth in the policy reaction function (16) is replaced by a non-linear asymmetric relation. The parameter κ determines the degree of asymmetry. In Graph 2, the impact of output growth on the interest rate is compared for the linear relation and a weak (κ =10) and a strong (κ =25) asymmetric relation. The persistence in the policy rule spreads this asymmetric effect through time but the degree of asymmetry that is considered remains very moderate.

Although we did not expect a major impact for the case around the CE output level, this rule does seem to improve the welfare results. An asymmetric policy response is able to generate positive efficiency gains in this stochastic setting compared to the deterministic steady state result.[16] These efficiency gains, which are calculated as the residual in Table 5 between the total welfare effect and the identified components, are of a similar magnitude to the costs from inflation, capital adjustment and volatility.

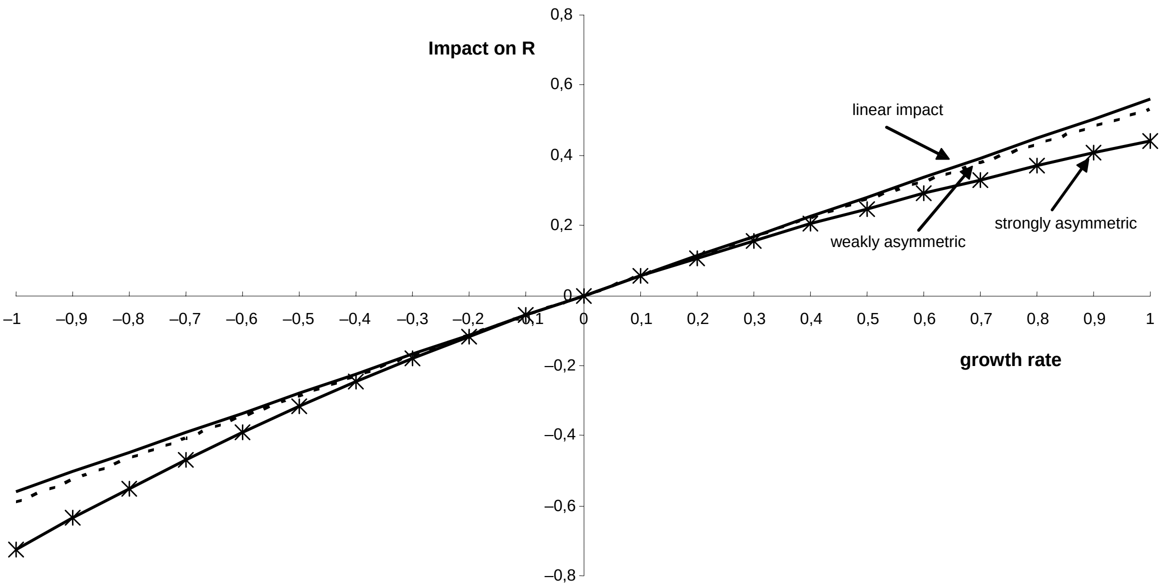
---

[15] Kim et al (2003) discuss the issue of instability of the second-order approximation methods and possible solutions.

[16] At this point, we have no intuition to explain this puzzling result. But given the highly non-linear nature of the model and the utility function, the result is not impossible.

Graph 2

Reaction coefficient on impact of the interest rate on GDP growth



4.2 Monetary policy and non-fundamental investment shocks around the lower MCE output level

The results for the simple rules remain valid for the stochastic simulations around the lower output level in a monopolistic competition context with level distortions. A stricter inflation policy and a reaction to the output gap can limit the costs of the non-fundamental shock, but the impact on the linear term measuring the inefficiency is less sensitive to the monetary policy rule here than it was in the previous table.

In this case, the benefits from an asymmetric monetary policy response to the non-fundamental shock are clear. An asymmetric policy response is able to take full benefit from the positive investment shocks that move output towards the more efficient production level. In contrast, policy is relaxed more rapidly at times of negative investment shocks in order to minimise the negative consequences for output. On average, this asymmetric policy response can be considered as a more accommodating monetary policy because the real interest rate will be lower on average while inflation and the nominal rate will be higher on average. The question then arises whether such a policy can be credible and whether the assumption of commitment to the policy rule is still valid in this context.

The results in Table 6 show that the average inflation rate under the asymmetric policy rule is above the deterministic steady state level. At the same time the average investment and output level in the stochastic simulation are also above the deterministic steady state. The asymmetric policy creates a positive relation between the average long-run inflation outcome and the average output level.

Table 5

Welfare effects of a distortionary iid investment shock under alternative monetary policy rules

Results from the stochastic simulation with the second-order approximation methods
Stochastic simulations around the CE steady state output level

	Benchmark	**Simple rules**			**Asymmetric policy**	
	Estimated rule	**Weak π policy**	**Strong π policy**	**+ Output –gap**	**Weak**	**Strong**
Total welfare effect	–0.1020	–0.1700	–0.1093	–0.0878	–0.0640	–0.0092
In % of steady state consumption level	–6.3052 100.00%	–10.5122 100.00%	–6.7594 100.00%	–5.4316 100.00%	–3.9575 100.00%	–0.5708 100.00%
Price dispersion cost	–0.1013 1.61%	–0.5067 4.82%	–0.1351 2.00%	–0.0676 1.24%	–0.1351 3.41%	–0.3209 56.22%
Wage dispersion cost	–0.1442 2.29%	–0.4532 4.31%	–0.2335 3.45%	–0.1133 2.09%	–0.1614 4.08%	–0.2506 43.91%
Capital adjustment cost	–1.4341 22.74%	–1.5763 14.99%	–1.4829 21.94%	–1.3872 25.54%	–1.4366 36.30%	–1.4401 252.28%
Variance cost	–0.5205 8.25%	–0.9913 9.43%	–0.5843 8.64%	–0.3845 7.08%	–0.5257 13.28%	–0.5509 96.51%
Intra-/intertemporal inefficiency	–4.1051 65.11%	–6.9848 66.44%	–4.3236 63.96%	–3.4790 64.05%	–1.6987 42.92%	1.9917 –348.91%
Average inflation rate q-to-q Standard error	–0.0033 0.0314	–0.0184 0.0699	0.0112 0.0367	0.0214 0.0262	0.0071 0.0308	0.0226 0.0302
Average output level % deviation from steady state Standard error	0.0055 0.5127	–0.0182 0.7298	–0.0378 0.5467	–0.0393 0.4328	0.0671 0.5143	0.1695 0.5206

Table 6

Welfare effects of a distortionary iid investment shock under alternative monetary policy rules

Results from the stochastic simulation with the second-order approximation methods
Stochastic simulations around the lower MCE steady state output level

	Benchmark	Simple rules			Asymmetric policy	
	Estimated rule	**Weak π policy**	**Strong π policy**	**+ Output –gap**	**Weak**	**Strong**
Total welfare effect	–0.0847	–0.1373	–0.1341	–0.1129	0.0126	0.1650
In % of steady state consumption level	–5.2347 100.00%	–8.4884 100.00%	–8.2932 100.00%	–6.9803 100.00%	0.7775 100.00%	10.2053 100.00%
Price dispersion cost	–0.0338 0.65%	–0.1802 2.12%	–0.0450 0.54%	–0.0225 0.32%	–0.0450 –5.79%	–0.1126 –1.10%
Wage dispersion cost	–0.0343 0.66%	–0.1442 1.70%	–0.0710 0.86%	–0.0298 0.43%	–0.0412 –5.30%	–0.0755 –0.74%
Capital adjustment cost	–1.4321 27.36%	–1.5928 18.76%	–1.5114 18.22%	–1.4052 20.13%	–1.4358 –184.67%	–1.4418 –14.13%
Variance cost	–0.1558 2.98%	–0.3115 3.67%	–0.1989 2.40%	–0.1283 1.84%	–0.1568 –20.16%	–0.1637 –1.60%
Intra-/intertemporal inefficiency	–3.5787 68.36%	–6.2597 73.74%	–6.4669 77.98%	–5.3945 77.28%	2.4563 315.92%	11.9988 117.57%
Average inflation rate q-to-q Standard error	–0.0033 0.0219	–0.0034 0.0520	0.0071 0.0276	0.0165 0.0180	0.0047 0.0220	0.0166 0.0228
Average output level % deviation from steady state Standard error	0.0025 0.4691	–0.0122 0.6956	–0.0343 0.5422	–0.0346 0.4154	0.0570 0.4701	0.1449 0.4740

These results illustrate that if output is fluctuating below the first-best output level, the task for an optimal monetary policy from the welfare point of view is much more complicated. Our conclusions are in contrast with most of the results presented in the literature, where the optimal monetary policy is derived as the linear policy rule that is optimising a quadratic approximation of the welfare function subject to the linearised model (Rotemberg and Woodford (1997)). Most of this literature assumes, however, that there exist lump sum taxes and subsidies that compensate for the impact of markups on the steady state equilibrium level. These instruments can be used by fiscal policy to offset the distortions in the economy. The recent paper by Benigno and Woodford (2003) drops this assumption but still retains the assumption that the optimal fiscal policy is stabilising the markup distortion over time, so that the optimal monetary policy can still be described as the solution from a linear-quadratic problem. In the real world it is difficult to imagine that fiscal policy is indeed able to reproduce the first-best outcome or to adjust optimally from period to period. Therefore, the analysis of optimal monetary policy in the presence of markup distortions is more appropriate to mimic real world policy questions.

5. Conclusions

Large asset price and investment cycles that are difficult to motivate by fundamental factors generate complicated decision problems for monetary policymakers. General equilibrium models can be helpful in sorting out the welfare effects of the different inefficiencies that are generated by these cycles. Model solution methods based on higher-order approximations are necessary for this welfare analysis and can increase our understanding of the issues involved. This paper is a first attempt to perform such an analysis using a standard estimated sticky price and wage general equilibrium model.

However, a lot of work remains to be done. First, the estimated non-fundamental equity price shock we analyse in this paper is different from what observers traditionally understand as a typical asset price bubble. More realistic, but exogenously generated bubble processes could be introduced in the model quite easily. These might already change part of our conclusions because these bubble processes are expected to burst at a certain point in the future and generate negative investment and output consequences at that point. If the size of these negative output effects is sufficiently important, this might change the policy reaction drastically as the welfare effects of possible future output declines can easily dominate the welfare gains from more moderate short-run output expansions. This last effect might even be strengthened if the transmission effect of asset price fluctuations to the real economy is also asymmetric with a much larger impact during the bursts. In such a scenario monetary policy actions today may serve as an insurance policy against larger losses in the future.[17] In reality the decision problem might therefore be a much more complicated and dynamic problem.

Furthermore, there is also the identification issue to distinguish between fundamental and non-fundamental asset price movements. However, if the efficiency gains from higher output levels are the dominant factor in the welfare analysis, this difference might not be as important as it is in the case of fluctuations around the first-best output level.

Ideally, asset price booms should be modelled as endogenous processes, probably related to the uncertainty and heterogeneous expectations about fundamental shocks. Alternative monetary policy rules may affect the probability of asset price booms and bursts in such a setup. Asymmetric policy rules may also create a moral hazard issue by providing one-sided protection against the negative risks. Understanding these mechanisms together with more knowledge about the transmission mechanism from these financial variables to the real economy would make the policy conclusions of this type of research much more robust. Introducing financial frictions, firm-specific capital and heterogeneous agents will certainly be ingredients for future research in this context.

[17] See Bordo and Jeanne (2002) for an analysis of this argument.

References

Bank for International Settlements/Centre for Economic Policy Research (1998): *Asset prices and monetary policy: four views*, August.

Benigno, P and M Woodford (2003): "Optimal monetary and fiscal policy: a linear-quadratic approach", *NBER Working Papers*, no 9905, August.

Bernanke, B and M Gertler (2000): "Monetary policy and asset price volatility", *NBER Working Papers*, no 7559.

——— (2001): "Should central banks respond to movements in asset prices?", *American Economic Review Papers and Proceedings*, 91 (2), May, pp 253-57.

Bordo, M D and O Jeanne (2002): "Boom-busts in asset prices, economic instability, and monetary policy", *NBER Working Papers*, no 8966.

Borio, C and P Lowe (2002): "Asset prices, financial and monetary stability: exploring the nexus", *BIS Working Papers*, no 114.

Cecchetti, S G, H Genberg, J Lipsky and S Wadhwani (2000): *Asset prices and central bank policy*, CEPR.

Detken, C and F Smets (2003): *Asset price booms and monetary policy*, manuscript, June.

Dupor B (2001): *Nominal price versus asset price stabilisation*, manuscript, Wharton School, University of Pennsylvania.

——— (2002), *Stabilising non-fundamental asset price movements under discretion and limited information*, manuscript, Wharton School, University of Pennsylvania.

Erceg, C J and A T Levin (2002): "Optimal monetary policy with durable and non-durable goods", *European Central Bank Working Paper Series*, no 179.

Gali, J, M Gertler and D Lopez-Salido (2001): *Markups, gaps, and the welfare costs of business fluctuations*, manuscript, May.

Gilchrist, S, C P Himmelberg and C Huberman (2002): *Do stock price bubbles influence corporate investment?*, manuscript, 11 December.

Greenspan, A (2002): *Economic volatility*, Jackson Hole Conference, Federal Reserve Bank of Kansas City.

Issing, O (2003): "Monetary and financial stability: is there a trade-off?", in "Monetary Stability, Financial Stability and the Business Cycle", *BIS Papers*, no 18, September.

Julliard, M (2003): *Dynare*, manuscript, Cepremap.

Kent, C and P Lowe (1997): "Asset price bubbles and monetary policy", *Reserve Bank of Australia Working Paper* RDP9709.

Kim, J, S Kim, E Schaumburg and C A Sims (2003): *Calculating and using second order accurate solutions of discrete time dynamic equilibrium models*, manuscript, Princeton University.

Levin, A and E Swanson (2003): *Optimal monetary policy in an imperfect world*, manuscript, Federal Reserve Board.

Rigobon, R and B Sack (2003), "Measuring the reaction of monetary policy to the stock market", *Quarterly Journal of Economics*, vol 118 (2).

Rotemberg, J J and M Woodford (1997): "An optimisation-based econometric model for the evaluation of monetary policy", *NBER Macroeconomic Annual*, pp 297-346.

Schmitt-Grohe, S and M Uribe (2002): "Solving dynamic general equilibrium models using a second-order approximation to the policy function", *Working paper*, Rutgers University.

Smets, F and R Wouters (2003a): "An estimated dynamic stochastic general equilibrium model of the euro area", *Journal of the European Economic Association*, vol 1 (5).

——— (2003b): *Output gaps: theory versus practice*, manuscript, ECB-NBB.

——— (2003c): *Shocks and frictions in US and euro area business cycles: a Bayesian DSGE approach*, manuscript ECB-NBB.

Taylor, J (1993): "Discretion versus policy rules in practice", *Carnegie-Rochester Conference Series on Public Policy*, 39, pp 195-214.

Woodford, M (2003): "Inflation stabilisation and welfare", in *Interest and Prices*, Princeton University.

Productivity, monetary policy and financial indicators

Arturo Estrella[1]

Introduction

Labour productivity is widely thought to be informative with regard to inflation and it therefore comes up frequently in discussions about the conduct of monetary policy. However, productivity growth is very difficult to interpret in real time. From a time series perspective, it is an unwieldy mixture of low-frequency trends and cyclical movements, with a generous dose of short-term noise thrown in. The net result is a very volatile series whose implications are difficult to grasp even in hindsight. However, if productivity does offer the prospect of information about inflation, it is worth making an effort to go beyond the surface noise. In that spirit, this paper considers why it may be helpful to pay attention to productivity in monetary policy and examines the possible use of financial indicators to obtain information about cyclical fluctuations in productivity growth in real time.

The main reason that productivity is thought to be helpful in monetary policy is that it may contain information about future inflation. This information may be directly about inflationary trends, or it may be about real trends (say in potential output) which could indirectly shed light on future inflation. A brief review of the relevant research suggests that there are definitely some theoretical relationships that should be explored, but that the empirical obstacles are far from easy to clear. Nevertheless, there is some empirical evidence that monetary policy in the United States has reacted to changes in productivity growth since the 1950s.

If knowing about productivity growth is helpful over the business cycle, but it is hard to measure, can we find any simple indicators that are related to productivity growth with any degree of robustness? We consider here a handful of financial indicators, all easy to track, and all exhibiting some degree of correlation with productivity growth over the business cycle. In order to bring to the fore the cyclical relationships, it is necessary to filter the data to exclude long-term trends and short-term noise. For these purposes, we apply standard techniques that allow us to split the movements of each variable into components that move in a single frequency or in a range of similar frequencies.[2] The results vary substantially across financial indicators, but they suggest that the federal funds rate, the spread between rates on short- and long-dated US Treasury securities, and the returns on the S&P 500 all contain meaningful information about cyclical movements in US labour productivity in real time.

1. Productivity and monetary policy

Why is it important for monetary policymakers to consider the growth in labour productivity in their deliberations? If the main goal of monetary policy is to keep inflation in an acceptable range, we must conclude that knowing about labour productivity is helpful if it ultimately sheds some light on the issue of inflation. We examine three possibilities. First, that productivity may contain direct information about future inflation. Second, that productivity may contain information about other variables, say potential output or the output gap, which in turn may contain information about future inflation. Third, that productivity and inflation may be simply statistically related over the business cycle, and knowing about one or acting on it may have consequences for the other.

---

[1] The views expressed in this paper are those of the author and do not necessarily represent those of the Federal Reserve Bank of New York or the Federal Reserve System.

Federal Reserve Bank of New York, 33 Liberty Street, New York, NY 10045. Tel: +1 212 720 5874. Fax: +1 212 720 1582. E-mail: arturo.estrella@ny.frb.org.

[2] A technical discussion of the techniques applied here can be found in Estrella (2003).

Consider first the possibility of a direct connection between productivity and inflation. In the United States, the 1970s brought the confluence of two unwelcome events, a noticeable drop in labour productivity growth and an even more noticeable rise in the rate of inflation. From the early 1980s, there was a surge in journal articles examining possible direct connections between productivity and inflation.[3] This literature identified a strong negative empirical correlation between the two variables and offered a series of theoretical arguments to help explain the facts.

Some of the arguments suggested that causality went from productivity to inflation. For instance, a slowdown in productivity growth could reduce aggregate supply and, other things equal, lead to a rise in the aggregate price level. Other arguments had causality going in the opposite direction. For example, a rise in inflation could distort incentives and lead to adverse changes in employment, savings, investment and trade. Alternatively, higher inflation could increase aggregate uncertainty, which could then disrupt business plans. Some of the empirical evidence, particularly the evidence based on vector autoregressions, suggested that it was most likely that causality went from inflation to productivity.[4]

Either way, there would be implications for monetary policy. If productivity growth tended to reduce inflation, monetary policymakers would have to factor current productivity growth into their decision-making so as to avoid over- or underreacting. If inflation lowered productivity growth, policymakers would have an added impetus to control inflation, although the information content of productivity would be less of a factor.

The second possibility is that productivity influences variables, such as potential output, which may have either a causal effect on, or a predictive connection with, future inflation. A standard view in current macroeconomics is that the output gap, the difference between actual and potential output, helps predict inflation. This view is embodied in the Phillips curve, based on an original proposal by Phillips (1958).[5] If higher productivity growth is consistent with faster sustainable output growth, a given level of actual output produces a smaller output gap and lower future inflation.

The third possibility is that productivity growth and inflation are not causally related in any clear way, but are merely statistically correlated. If the correlation were such that productivity were a leading indicator of (lower) inflation, and if it were persistent and robust, information about productivity could be used almost as in the causal circumstances, perhaps with a bit more caution and scepticism. However, if productivity is inversely related to contemporaneous inflation, as some of the evidence suggests, policymakers may find that they get extra benefit from keeping inflation under control (and thus have an extra incentive to do so).

To put the discussion in perspective, we end this section with a small macroeconomic model of the United States that clarifies some of the empirical questions and asks whether policymakers have taken productivity into account in their decisions since the 1950s. The model is a vector autoregression (VAR) using quarterly data from the first quarter of 1955 to the first quarter of 2003. There are four variables in the model: non-farm productivity growth, CPI inflation, non-farm output growth and the federal funds rate. The first three variables, obtained from the US Department of Labor, are measured as first differences of logs. The interest rate is in per cent per annum and is obtained from the Federal Reserve. Three lags of each variable are included in each equation.[6]

We present two types of results to examine the implications of the model. First, the results of Granger causality tests for the VAR are shown in Table 1. They indicate the level of importance of lags of each variable in each of the four individual equations, and provide some direct information about the estimates. Second, Figure 1 shows impulse responses to shocks in each of the four variables of the model, which help isolate the effects of the individual variables.

The first line of Table 1 indicates that no variable is very significant in explaining productivity growth. One interpretation could be that productivity growth is to some extent exogenous, that is, it does not

---

[3] For example Clark (1982), Jarrett and Selody (1982), Ram (1984) and Cameron et al (1996).

[4] For example Ram (1984). However, Sbordone and Kuttner (1994) and Saunders (1998) find that the negative relationship disappears if the model controls for monetary policy.

[5] See eg Gali and Gertler (1999) for a recent application.

[6] A preliminary test using the Akaike information criterion suggests this is an appropriate lag length.

react to changes in other variables. Another possibility is that the variable, as mentioned earlier, is dominated by short-term noise, which is impossible to predict. Perhaps if it were possible to eliminate that noise, the relationships would be clearer. We come back to this point in the next section.

The second line in the table shows that output reacts to changes in both productivity and the federal funds rate. The relationship with the funds rate is consistent with earlier research that shows that this rate is a good indicator of the stance of monetary policy, which is expected to affect output within a few quarters, as in the model.[7] Inflation is very persistent, and its own lags are very significant, as indicated in the table.[8] Otherwise, it seems to be affected only by lags in the funds rate. We need to exercise caution with regard to this last relationship, however, since the sign of the relationship is not necessarily what we would expect.[9]

The final line in the table may be interpreted as a crude model of the "Fed reaction function", the extent to which policy reacts to observable macroeconomic variables.[10] The results suggest that policy reacts strongly to the lagged funds rate, output and inflation, and less strongly to productivity growth.

The impulse responses in Figure 1 are based on the ordering shown in the table, and the shock to each variable is of a magnitude equal to the standard error of the corresponding equation. By construction, the results are consistent with the Granger causality results, but afford a somewhat different perspective. The last row, for instance, corresponds to the last row of Table 1, the "Fed reaction function". We see in the figure that the funds rate, output and inflation are all significant, as expected, and that they have the expected signs. In addition, the figure indicates that the policy reaction to productivity is also statistically significant for some horizons, albeit of a smaller magnitude than the reaction to other variables.

In the row corresponding to inflation, we see manifestations of several of the patterns that have been discussed before. First, the persistence of inflation is apparent in the slow decline of this variable in response to a shock in itself. Second, the Phillips curve relationship that predicts that higher growth will lead to higher inflation is clear in the third panel. Third, we see the price puzzle in the final panel of the row: inflation seems to rise in response to a positive shock in the funds rate. Though somewhat disturbing, this response is short-lived and relatively small.

For output, we see indications of an "IS equation" in the last panel of the row. An upward shock to the funds rate leads to a noticeable drop in output, particularly two quarters ahead. Finally, we see in the upper left-hand panels of the figure some evidence that productivity and inflation react negatively to one another, as some of the earlier literature has suggested. Some of these results are statistically significant, though they are all fairly small.

2. Productivity and financial indicators at business cycle frequencies

We turn now to the cyclical correlations between productivity growth and several financial indicators. As noted earlier, the purpose here is to determine whether easily accessible financial indicators can shed some light on the current situation in the productivity cycle. In addition to the federal funds rate, which we used in the previous section, we include the three-month Treasury bill rate, the 10-year Treasury bond rate, and the term spread between these two rates.[11] We also include the return on the S&P 500 Index (first difference of the log) and, to look at the direct correlation between productivity and inflation over the business cycle, the CPI inflation used in the previous section.

---

7 Bernanke and Mihov (1998), for example, suggest that the funds rate is the best single indicator of the stance of monetary policy in the United States.

8 See Fuhrer and Moore (1995).

9 This "price puzzle" is found in virtually every VAR of this type. For a discussion, see Sims (1986).

10 See, for example, Clarida et al (2000).

11 Data for all the interest rates were obtained from the Federal Reserve.

We focus on business cycle frequencies by operating within the frequency domain.[12] This allows us to measure correlations, leads and lags that pertain only to the frequencies of interest. Specifically, we look at averages over frequencies corresponding to cycles of 11 to 28 quarters (roughly three to seven years) in length. Empirical evidence shows that these frequencies are representative of the US business cycle.

To illustrate the effects of focusing on business cycle frequencies only, Figure 2 compares the business cycle component of productivity with the untransformed productivity growth series. The filtered productivity series eliminates both the short-term noise that makes the growth series very hard to interpret and long-term trends that have slow-moving effects on the series. The business cycle pattern that emerges is clear, and we can observe its relationship with NBER-dated recessions, which are shaded in the figure.

Table 2 contains several measures of the relationship between productivity growth and each of the financial indicators (and inflation). Coherence is a correlation measure that indicates how strongly the two variables are related at business cycle frequencies. It ranges from 0 (no correlation) to 1 (perfect correlation). The caveat is that this correlation may not be contemporaneous, but may involve a lead or a lag. A measure of the magnitude of this lead or lag is the phase lead, presented next in the table. The (weighted) average of the business cycle frequencies is about 16 quarters. Thus, a phase lead of 0 quarters means that the relationship is contemporaneous, and a phase lead of eight quarters, or half a cycle, means that the contemporaneous relationship is essentially negative.

The final measure in Table 2 is the in-phase correlation, which is similar to the coherence, but focuses only on contemporaneous correlation. It also has a sign that indicates the direction of the relationship. If the coherence and in-phase correlation of a pair of variables are about the same size (in absolute value), the phase lead is small. Conversely, a high coherence with a low or negative in-phase correlation is indicative of a substantial phase lead or lag.

In Table 2, coherence with productivity is fairly high and statistically significant for all indicators. The usefulness of this result stems from the fact that it confirms that all these variables have substantial variation at business cycle frequencies. Unfortunately, it helps very little in differentiating among indicators of cyclical productivity growth in terms of quality or timing.

Turning to the next measure in the table, the phase leads are small for both the term spread and the stock index, neither of which is significantly different from zero. This is a sign that the relationships with productivity growth are roughly contemporaneous and that these variables show some promise as coincident indicators. The lags for interest rates are all relatively large, but this is not necessarily a problem. None of the lags are significantly different from eight quarters, which indicates that there may be a high, but negative, contemporaneous correlation with productivity growth.

These results are confirmed by looking at the in-phase correlations, which are high and close in absolute value to the coherence for most of the interest rate and stock variables. Only the correlation for the 10-year bond rate is less than one half in absolute value. Graphical evidence that provides visual confirmation of the results of Table 2 is presented in Figure 3, which shows the business cycle components of the variables in the time domain. It is clear from the various panels of Figure 3 that the federal funds rate and the term spread have particularly tight relationships with productivity at these frequencies over most of the sample period, which accords with the ranking of the in-phase correlations in Table 2.

Inflation is also included in Table 2, and the results indicate that inflation, much like the short-term interest rates, has a strong negative relationship with productivity over the business cycle. As argued in the previous section, one interpretation of this result, even if it is purely statistical, is that monetary policymakers have an additional incentive to keep inflation cyclically low, since low cyclical inflation is regularly accompanied by high cyclical productivity growth.

One drawback of using the financial indicators in the foregoing manner is that it requires computation of the cyclical components of the financial variables, as well as for productivity. How much information

[12] Time series are transformed into the frequency domain by taking Fourier transforms. For details, see Appendix 2 in Estrella (2003).

about cyclical productivity may be gleaned from the financial series directly, without resorting to frequency domain methods?

Table 3 suggests that some useful information can in fact be obtained simply by looking at the financial series. The first column of the table shows the in-phase correlation from Table 2. The second column, however, correlates the cyclical component of productivity with the untransformed financial indicators. We see that the correlations are lower (in absolute value), though most are not insubstantial. The largest correlation is for the term spread, which has a value of 41%. The federal funds rate and the stock index are both at 29%, which is still somewhat informative. The worst case is the bond rate, which is clearly not very reliable.[13]

To gauge the gains from the frequency domain analysis of productivity, the final column of Table 3 shows the correlation of the untransformed financial indicator with untransformed productivity growth. The difference between this measure and the others is most notable in the case of the term spread, whose correlation with directly observable productivity growth is only 18%. Once the short-term noise and the trends are removed from productivity growth, the correlation of its cyclical component with the term spread rises to 41%, and the in-phase correlation is 72%. We also see gains for the funds rate and the stock index, although of more modest magnitude.

3. Conclusions

The analysis of this paper suggests that information about the movements of labour productivity growth over the business cycle may be useful to monetary policymakers for various reasons, both direct and indirect. The empirical analysis shows that there are statistically significant relationships consistent with the theoretical usefulness of productivity and, moreover, that the data are consistent with US monetary policy taking productivity growth into account since 1955.

The paper also shows that financial indicators may be somewhat helpful in interpreting the noisy productivity growth series, in particular by serving as coincident indicators of the cyclical component of productivity growth. The strongest signals are derived from the term spread, the federal funds rate and growth in the S&P 500 Index.

Results for some of the financial indicators are statistically significant, though they may not seem particularly impressive. To put these in perspective, however, it is helpful to bear in mind that looking at the productivity growth series itself is not highly informative, as Figure 2 shows. In other words, one needs all the help one can get.

[13] A useful benchmark for these correlations is the correlation between the untransformed productivity growth series and its own business cycle component, which is 30%. Note that the business cycle component of productivity growth is more highly correlated with the observable term spread and about as correlated with the actual federal funds rate and stock index growth. I am grateful to Eduardo Loyo for suggesting this comparison.

Table 1: The table reports p-values for the exclusion tests of the lags of the variables named in each column from the forecasting equation of the variable named in each row.

Table 1

Granger causality tests for four-variable vector autoregression

1955 Q1 to 2003 Q1

Dependent variable	Productivity	Output	Inflation	Fed funds rate
Productivity	.20	.11	.06	.10
Output	.00	.00	.21	.00
Inflation	.70	.32	.00	.00
Fed funds rate	.06	.00	.02	.00

Note: Productivity, output and inflation are first differences of logs and the federal funds rate is in per cent per annum. Each equation includes a constant term and three lags of each variable.

Table 2

Coherence, phase lead of productivity, and in-phase correlation with productivity at business cycle frequencies

1955 Q1 to 2000 Q4

Variable	Coherence (t-statistic)	Phase lead (standard error)	In-phase correlation (t-statistic)
Fed funds rate	.812 (5.32)	–7.03 (1.24)	–.746 (–4.52)
3-month T-bill rate	.772 (4.81)	–7.03 (1.45)	–.691 (–3.99)
10-year T-bond rate	.571 (3.04)	–5.68 (2.53)	–.319 (–1.55)
Term spread	.721 (4.26)	.25 (1.69)	.717 (4.23)
S&P 500 Index	.645 (3.60)	1.03 (2.08)	.597 (3.23)
CPI inflation	.701 (4.08)	–7.26 (1.79)	–.652 (–3.65)

Note: The business cycle is defined by cycle lengths of 11 to 28 quarters, centred at a weighted average of 16 quarters. For coherence and in-phase correlation, significance is calculated with respect to an arctanh transformation and a t-statistic is given. For phase lead, a standard error is provided to gauge the significance of differences from values other than zero, eg from half the mean cycle length of 16 quarters.

Table 3

Correlations with productivity at business cycle frequencies and all frequencies

1955 Q1 to 2000 Q4

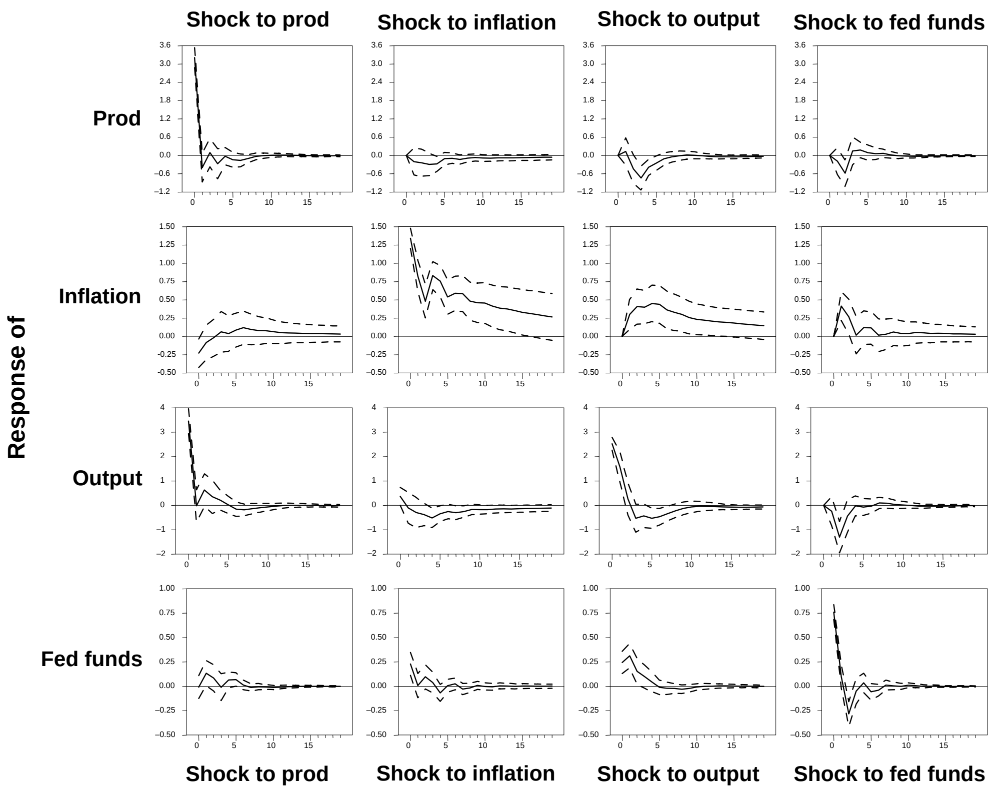
Variable	BCF productivity BCF variable	BCF productivity AF variable	AF productivity AF variable
Fed funds rate	–.746	–.285	–.259
3-month T-bill rate	–.691	–.234	–.230
10-year T-bond rate	–.319	–.073	–.166
Term spread	.717	.405	.184
S&P 500 Index	.597	.288	.215

Note: BCF means business cycle frequencies only; AF means all frequencies (untransformed variable). BCF for both productivity and variable produces the in-phase correlation of Table 2. The business cycle is defined by cycle lengths of 11 to 28 quarters, centred at a weighted average of 16 quarters.

Figure 1

Impulse responses for four-variable vector autoregression

1955 Q1 to 2000 Q4

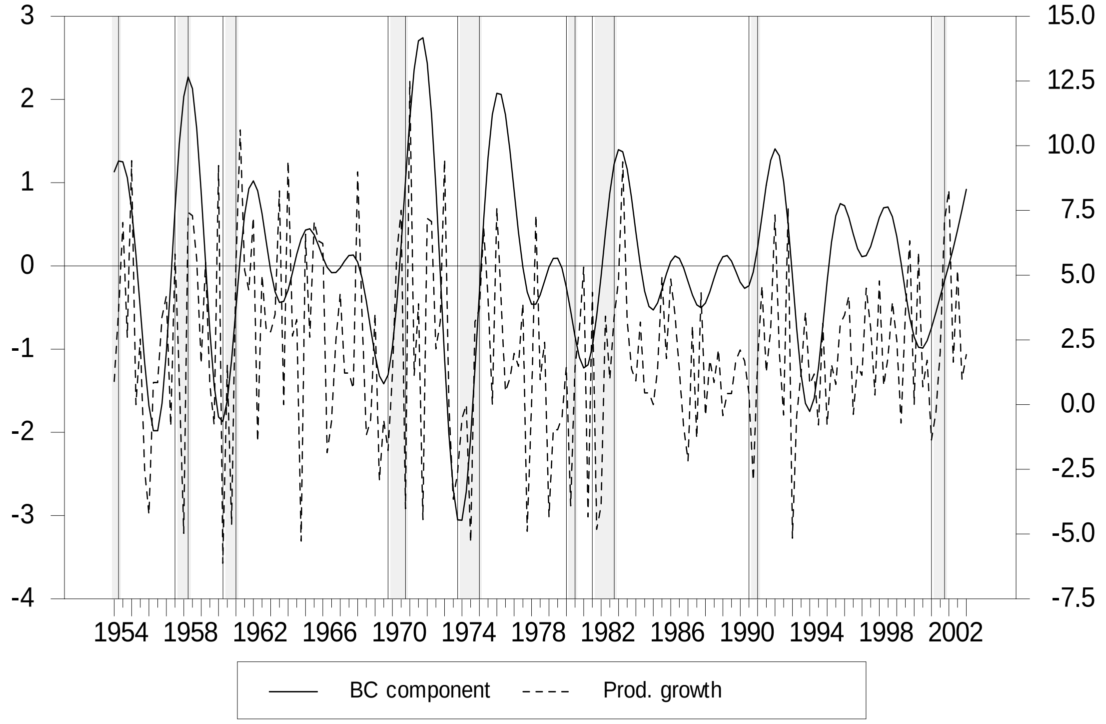


Note: The magnitude of each shock is the residual standard error for the corresponding equation. Contemporaneous ordering is as listed. Dashed lines represent a 95% confidence band using standard errors computed by Monte Carlo integration (see Sims and Zha (1999)).

Figure 2

Productivity growth and its business cycle component

1954 Q1 to 2002 Q4

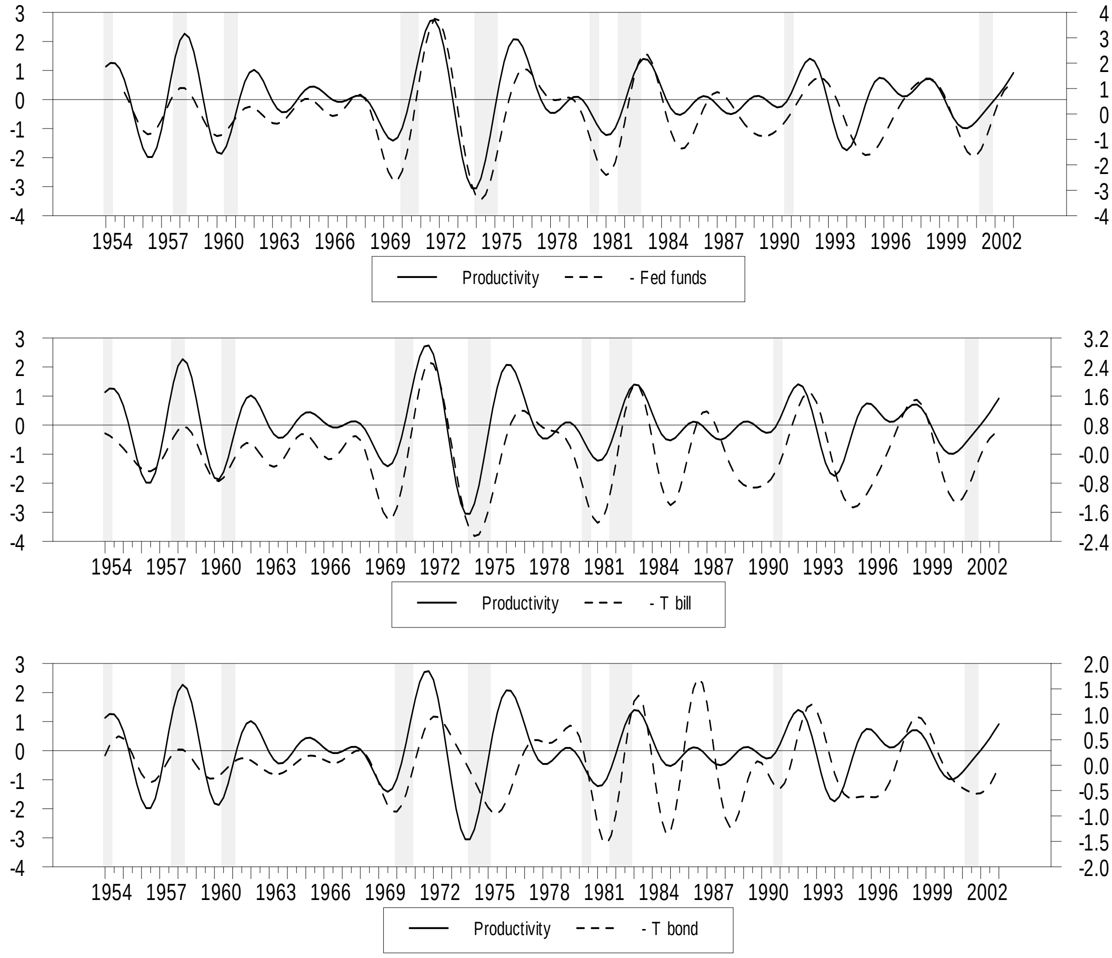


Note: The business cycle component is derived by focusing on business cycle frequencies in the frequency domain, retaining cycles of 11 to 28 quarters. Left scale is for solid line, right scale for dashed line. Federal funds series starts in 1955. Shading denotes NBER recession dates.

Figure 3a

Business cycle components of productivity growth and financial indicators

Interest rate indicators (with negative signs to make the correlations positive)
1954 Q1 to 2002 Q4

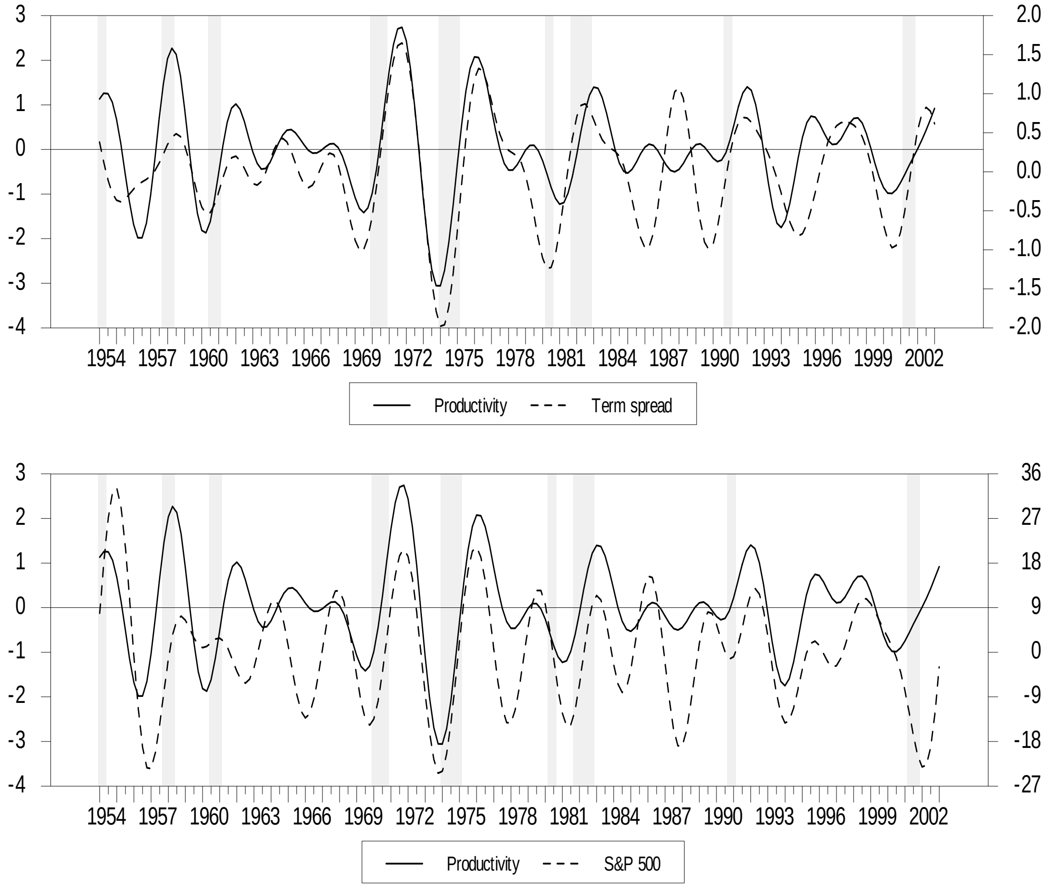


Note: Business cycle components are derived by focusing on business cycle frequencies in the frequency domain, retaining cycles of 11 to 28 quarters. Left scale is for solid line, right scale for dashed line. The federal funds series starts in 1955.

Figure 3b

Business cycle components of productivity growth and financial indicators

Term spread and S&P 500 Index
1954 Q1 to 2002 Q4



Note: Business cycle components are derived by focusing on business cycle frequencies in the frequency domain, retaining cycles of 11 to 28 quarters. Left scale is for solid line, right scale for dashed line.

References

Bernanke, Ben S and I Mihov (1998): "Measuring monetary policy", *Quarterly Journal of Economics*, vol 113, pp 869-902.

Cameron, Norman, Derek Hum and Wayne Simpson (1996): "Stylized facts and stylized illusions: inflation and productivity revisited", *Canadian Journal of Economics*, vol 29, pp 152-62.

Clarida, R, J Gali and M Gertler (2000): "Monetary rules and macroeconomic stability: evidence and some theory", *Quarterly Journal of Economics*, vol 115, pp 147-80.

Clark, Peter K (1982): "Inflation and the productivity decline", *American Economic Review*, vol 72, pp 149-54.

Estrella, Arturo (2003): "Decoding productivity: business cycle properties of labor productivity growth", *Working paper*, Federal Reserve Bank of New York.

Fuhrer, Jeffrey C and George Moore (1995): "Inflation persistence", *Quarterly Journal of Economics*, vol 110, pp 127-59.

Gali, Jordi and Mark Gertler (1999): "Inflation dynamics: a structural econometric analysis", *Journal of Monetary Economics*, vol 44, pp 195-222.

Jarrett, J Peter and Jack G Selody (1982): "The productivity-inflation nexus in Canada, 1963-1979", *Review of Economics and Statistics*, vol 64, pp 361-67.

Phillips, A W (1958): "The relation between unemployment and the rate of change of money wage rates in the United Kingdom, 1861-1957", *Economica*, vol 25, pp 283-99.

Ram, Rati (1984): "Causal ordering across inflation and productivity growth in the post-war United States", *Review of Economics and Statistics*, vol 66, pp 472-77.

Saunders, Peter J (1998): "Inflation, productivity, and monetary policy: are they related?", *Studies in Economics and Finance*, vol 18, pp 3-25.

Sbordone, Argia and Kenneth Kuttner (1994): "Does inflation reduce productivity?", *Federal Reserve Bank of Chicago Economic Perspectives*, pp 2-14.

Sims, Christopher A (1986): "Are forecasting models usable for policy analysis?", *Federal Reserve Bank of Minneapolis Quarterly Review*, vol 10, pp 2-16.

Sims, Christopher A and Tao Zha (1999): "Error bands for impulse responses", *Econometrica*, vol 67, pp 1113-55.

The term structure as a predictor of real activity and inflation in the euro area: a reassessment

Jesús Crespo Cuaresma, University of Vienna
Ernest Gnan and Doris Ritzberger-Grünwald,[1]
Austrian National Bank

1. Introduction

The slope of the yield curve is often used by financial and policy analysts as an indicator of future real activity and inflation. Empirical research tends to confirm the predictive power of the yield spread both for real activity and for inflation. Empirical research has focused mostly on the US economy and, to some extent, on the larger individual pre-EMU EU countries, whereas there are hardly any estimates for the euro area.[2]

Berk and van Bergeijk (2000, 2001) attempted an empirical assessment for the euro area. They concluded that both for individual euro area countries and for the euro area as a whole, the yield spread contains only very limited information on future inflation rate and output growth changes beyond the information contained in the history of the latter variables.

The present paper makes a new attempt to evaluate empirically the predictive power of the yield spread for euro area output and inflation. It makes use of the longer time series that have become available since Berk and van Bergeijk (2000, 2001). More importantly, the paper proposes a simple method to estimate time-varying term premia that may have caused the poor forecasting performance of the yield spread quoted in the above-cited contribution. We believe that this issue is of particular relevance when working with longer euro area financial market series, since the pre-1999 part (ie normally the larger part!) of the series is usually composed of raw country aggregates potentially plagued by changing risk premia related, among other factors, to the exchange rate mechanism of the European Monetary System. Equally, convergence phenomena in the run-up to the start of EMU may have heavily influenced national European bond rates. Working with synthetic pre-EMU bond rates for the euro area which are not adjusted for these changing risk premia can be expected to strongly influence empirical estimates of economic relationships.[3] In a recent contribution, Carstensen and Hawellek (2003) show that, for German data, assessing the time-varying nature of the term premium improves the quality of inflation forecasts obtained using term structure models.

In this contribution, we show that using a simple adjustment method for risk premia contained in bond rates significantly improves the information content of the term spread for future euro area output and, to a lesser extent, for future inflation rates. The basic idea behind the adjustment procedure is to approximate the (time-varying) term premium by making use of the relationship implied by the rational expectations hypothesis of the term structure (henceforth REHTS). By means of an out-of-sample forecasting exercise, we provide evidence that, for forecasting horizons ranging up to two years, the yield curve adjusted for risk premium improves significantly upon the observed term spread as a predictor of industrial production in the euro area. The results for the inflation rate are less clear-cut, but indicate that the use of the term premium adjustment can lead to improvements in the accuracy of the forecasts of inflation and core inflation rates.

---

1 The authors would like to thank Ernst Glatzer for research assistance and Martin Scheicher as well as participants at the Autumn 2003 Central Bank Economists' Meeting organised by the Bank for International Settlements and an internal Austrian National Bank seminar for helpful comments. The views expressed in this paper are those of the authors and do not necessarily represent the position of the Austrian National Bank or of the Eurosystem.
(E-mail: jesus.crespo-cuaresma@univie.ac.at; ernest.gnan@oenb.at; doris.ritzberger-gruenwald@oenb.at.)

2 For an extensive survey of the literature on using asset prices to forecast growth and inflation, see Stock and Watson (2003).

3 An example of this influence in the context of the estimation of monetary policy reaction functions is given by Crespo Cuaresma et al (2004).

The remainder of the paper is structured as follows. Section 2 summarises the theory underlying the predictive capabilities of the term spread for output and inflation, including the conditions by which they are influenced and limited. Section 3 proposes a simple risk premium adjustment method for euro area bond rates. Section 4 presents evidence for the euro area on the predictive content of the term spread for real activity and inflation, juxtaposing the results based on the premia-adjusted term spreads against results from unadjusted series. Section 5 concludes.

2. Theoretical underpinnings for a leading indicator property of the term structure

The theoretical background underlying the use of the term structure of interest rates as an indicator for market expectations of future inflation and/or real growth is based on the combination of the Fisher equation and the REHTS. The REHTS states that the yield to maturity of a bond with n periods to maturity can be decomposed into expected one-period yields and a risk premium, so that

$$R(n,t) = \frac{1}{n}\sum_{i=0}^{n-1} E_t R(1,t+i) + \Phi(n,t) \tag{1}$$

where $E_t(\cdot)$ is the conditional expectation operator using the information available up to period t, $R(n,t)$ is the yield to maturity of a bond with n periods to maturity, and $\Phi(n,t)$ is the average risk premium on an n-period bond until it matures.

Using the Fisher decomposition, equation (1) can be rewritten as

$$R(n,t) = E_t r(n,t) + E_t \pi(n,t) + \Phi(n,t) \tag{2}$$

where $E_t r(n,t)$ is the average real ex ante interest rate over the periods t to $t+n-1$, and $E_t \pi(n,t)$ is the average expected inflation rate over the periods $t+1$ to $t+n$. Under the REHTS, the risk premium is assumed to be constant over time. We will address this restrictive assumption in the next section.

The slope of the yield curve between maturities m and n can be decomposed into changes in the real rate and in expected inflation making use of (2). Consider equation (2) for a long-term interest rate of maturity n and a short-term interest rate of maturity m. Subtracting the latter from the former, we obtain

$$R(n,t) - R(m,t) = E_t[r(n,t) - r(m,t)] + E_t[\pi(n,t) - \pi(m,t)] + \Phi(n,t) - \Phi(m,t) \tag{3}$$

If real activity is related to changing real interest rates and if the term premium is constant, then equations (2) and (3) imply that the term spread should contain information about future economic activity and inflation.

While the literature on the theoretical background of the relationship between the term spread and future inflation rates is, to the knowledge of the authors, exclusively based on the Fisher decomposition and the REHTS[4] as described above (see Tzavalis and Wickens (1996)), different theoretical underpinnings have been proposed to the link between the term spread and output growth. From a theoretical point of view, the term spread may be related positively or negatively to future real output, depending on the channel at work. Various explanations have been put forward in the literature (see eg Estrella and Mishkin (1997), Berg and van Bergeijk (2000, 2001) and Estrella (2003)).

A first channel derives from the "common factor" effect of current monetary policy on both the term spread and real activity. As a credible central bank, for instance, tightens monetary policy, short-term interest rates rise, while long-term rates rise by less or are not affected at all, leading to a flattening of a previously positively sloped yield curve. After a lag of a few quarters, real activity is also dampened by the restrictive policy. Given the faster reaction of the term spread, the latter leads the slowdown in economic activity.

A second channel works through expectations about future monetary policy changes, in the presence of nominal rigidities. For instance, the expectation of a future monetary tightening (which can be

---

[4] Notable exceptions, discussed below, are Smets and Tsatsaronis (1997) and Estrella (2003).

thought of as a future shift of the LM curve) would imply higher future short-term rates, thus higher current long-term rates, and, consequently, an increase in the term spread. The expected upward shift in the future LM curve implies a shift to the left in the current IS curve and a fall in current and future output.

A third channel operates through real demand shocks. In terms of a standard IS/LM framework, an expected economic upswing as represented by a future outward shift in the IS curve raises expected future short-term rates (the expected outward shift in the IS curve raises future money demand). Due to the REHTS arbitrage condition, this expectation translates into higher current long-term rates.

In a fourth category of explanations, Harvey (1988) and Hu (1993) explain the correlation between the term spread and future economic growth from intertemporal consumption smoothing by using the consumption capital asset pricing model. The first-order condition of the consumption-based asset pricing model proposed by Campbell (1988) implies that expected returns and consumption growth are linearly related. Consequently, one should observe a comovement between the (real) term structure and the business cycle.

Two attempts to embed the link between the term spread and real activity into a broader analytical framework warrant specific mentioning. Smets and Tsatsaronis (1997) model the joint movements of output, inflation and the nominal term structure as the combined effect of four distinct fundamental shocks: aggregate demand, aggregate supply, monetary policy, and a long-term interest rate shock (driven by unwarranted “inflation scare”). They find that in both Germany and the United States about half of the medium-term variability in the term spread is accounted for by demand and monetary policy shocks, the other half being driven by supply shocks in Germany but by fears about long-term inflation prospects in the United States. They attribute this difference to the higher anti-inflationary credibility enjoyed by the Deutsche Bundesbank. They also find that the big role of supply shocks in explaining term-spread variability is the main reason for the much stronger leading indicator properties of the term spread for output growth. Finally, they show that the predictive content of the term spread is time-varying.

Estrella (2003) systematically investigates factors influencing the predictive power of the term spread for inflation and real variables in the framework of a single formal model comprising a (backward- or forward-looking) Phillips curve, a (backward- or forward-looking) IS equation, the Fisher equation, the term structure, and various monetary policy reaction functions. He finds that the yield curve should be a useful predictor of output and inflation under most circumstances. A positive relationship between the term spread and future output is predicted by the backward-looking form of the model. The prediction capabilities of the term spread importantly depend on the specific form of the policy reaction function. Thus, the predictive relationship, though robust, is not “structural”. In most specifications, further information beyond the term spread is useful in forecasting output. Finally, he finds that, since 1987, reflecting a regime of “strict inflation targeting”, the predictive power of the yield spread, though not entirely absent, has been diminished.

What are the implications from the theoretical literature for the paper at hand? First, there are sufficient sound theoretical underpinnings to justify a further investigation of the empirical leading indicator properties of the term spread for the euro area. Second, most channels and models suggest a positive relationship between the lagged term spread and real activity. However, there are also channels and shocks suggesting a negative relationship. Third, the monetary policy regime may affect the predictive power of the term spread. Thus, any reading of empirical relationships between the term spread and real activity or inflation requires a structural interpretation against the background of prevailing economic circumstances and the monetary policy regime in place. The regime change implied by the transition from the ERM to EMU appears to deserve particular attention in this context. The remainder of this paper pursues this latter aspect by proposing an empirical method to gauge the time-varying term premium in the euro area in the run-up to EMU. We take the theoretical literature as mere background motivation for our research and do not attempt to assess the validity of any of the above theoretical channels. Instead, we will exclusively concentrate on empirically assessing the out-of-sample forecasting abilities of the term spread for output growth and inflation taking into account the time-varying nature of the risk premium.

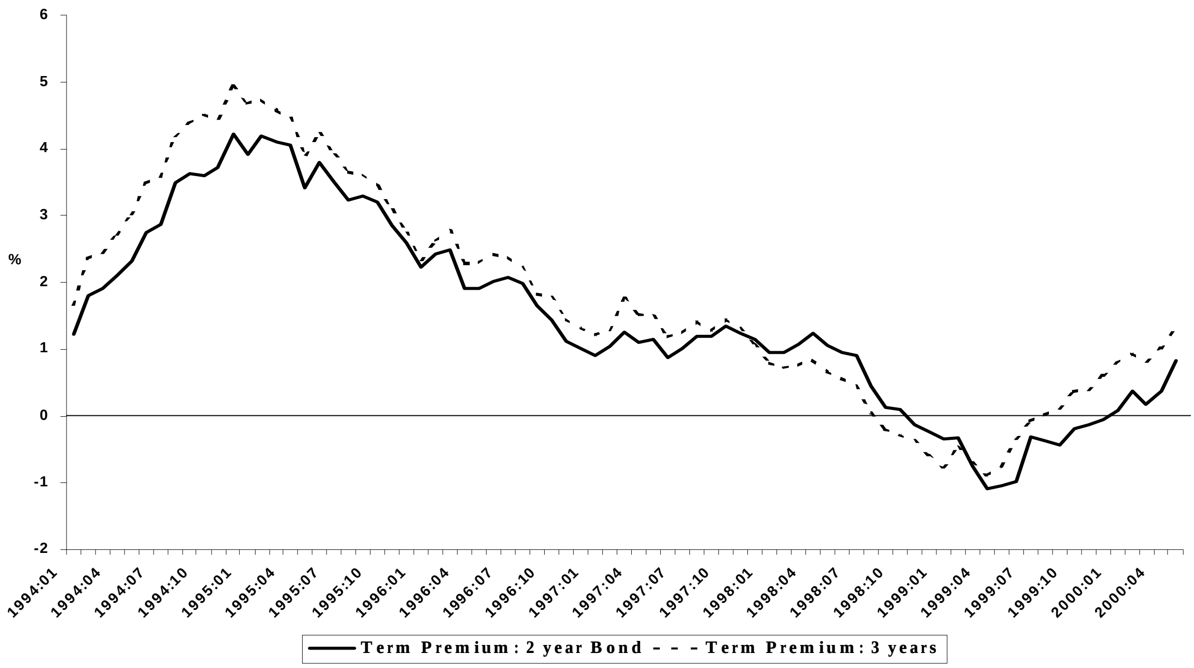
3. A simple risk premium adjustment

The assumption of constant risk premia is unlikely to have held in individual euro area countries and therefore in the euro area as a whole during the time of the ERM and in the run-up to EMU. This section will provide evidence of the existence of time-varying risk premia for long rates in the aggregate euro area. We also propose a simple method to obtain a (potentially time-varying) estimate of $\Phi(n,t)$ in (1).

Some evidence for our claim can be found by extracting the risk premium from equation (1) for the observed two- and three-year bond yields in the euro area.[5] Graph 1 presents the risk premium estimates implied by (1) for these maturities under the assumption of perfect foresight, that is, substituting the expected values with those which were actually realised. The one-month interest rate was used as the short rate. The implied risk premia are plotted for the period ranging from January 1994 (first available observation) to April 2000 (last period for which it is possible to obtain an implied premium for the three-year bond). The risk premium is far from being constant for both cases, reaching a global maximum in late 1994 of around 4 percentage points for the two-year bond and 5 percentage points for the three-year bond. A clear convergence pattern towards zero is observed during the run-up to EMU, culminating in premia around zero in the second half of 1998. The negative risk premium for both long rates during practically the whole of 1999 is due to the increase of short-run nominal interest rates which ran parallel to the rise in inflation after the inception of EMU. The subsequent stabilisation of inflation rates, which was followed by a reduction in the one-month nominal interest rate in the last period of the sample, results in positive risk premia from January 2000 onwards.

Graph 1

Implied risk premia, perfect foresight



However, the estimates presented in Graph 1 can only be obtained a posteriori. If the aim is to correct the term spread for time-varying term premia in order to use the information contained in the adjusted yield curve for predicting future growth rates of output or inflation rates, a real-time estimate of the risk

---

5 Much of the empirical literature tends to use the 10-year bond on the long side of the term spread. The relatively short sample existing for the aggregate euro area does not allow for sensible empirical work based on such long maturities if the REHTS is to be taken literally.

premium needs to be obtained with information ranging up to the time period in which the forecasts are carried out. We will use a simple expectation formation method to overcome this difficulty. For each time period, we will assume that expectations are formed as forecasts of the variables of interest (the nominal short rate) given the history of this variable up to period t. We will assume that individuals obtain point forecasts of the short-term nominal interest rate using simple autoregressive models. Using the information up to period t on one-month nominal rates, an autoregressive process of order p (AR(p)) model[6] is fitted to the data, and forecasts of the short rate are obtained for $n-1$ periods ahead, where n is the maturity of the bond whose risk premium we are estimating.

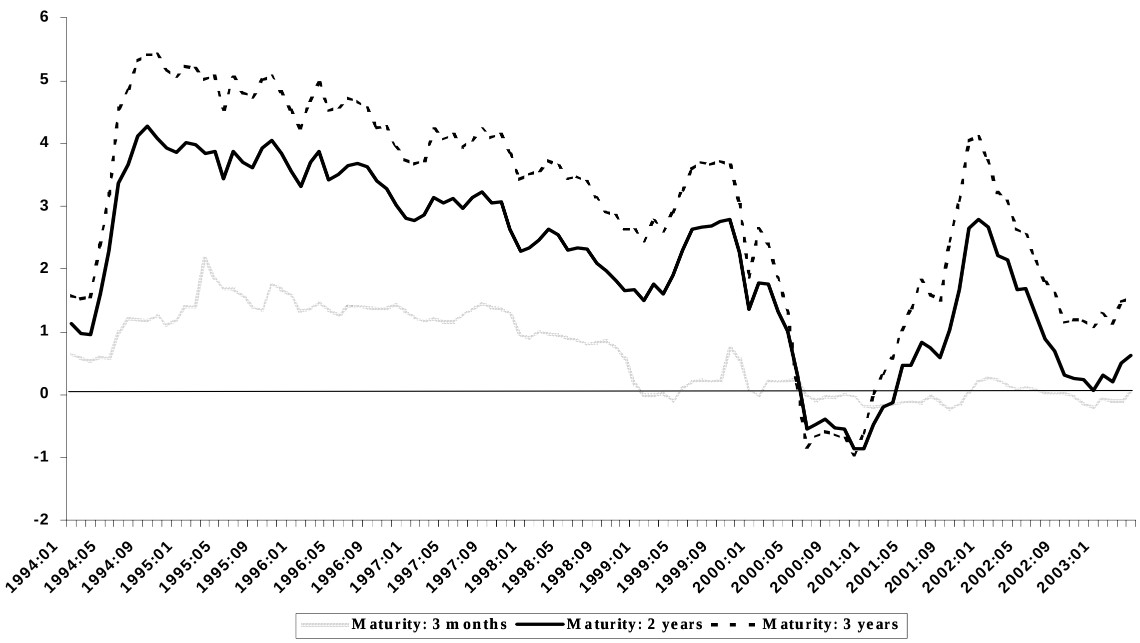
The estimate of the risk premium of the bond with maturity n at period $t(\widehat{\Phi}(n,t))$ is then given by the difference between the actual bond yield and the yield implied by the first terms on the right-hand side of (1)

$$\widehat{\Phi}(n,t) = R(n,t) - \frac{1}{n}\sum_{i=0}^{n-1}\hat{R}(1,t+i) \tag{4}$$

where $\hat{R}(1,t+i)$ is the one-month real interest rate in period $t+i$ predicted by the autoregressive model. Analogously to the definition of $\Phi(n,t)$ in (1) if perfect foresight is not assumed, the estimate given by (4) is not only composed of a risk premium, but also includes the forecast error of individuals when forming expectations.

Graph 2

Risk premia estimates: autoregressive expectations



Graph 2 presents the estimates of the risk premia obtained by applying this method to the euro area data for three-month interest rates and the two- and three-year bond yields.[7] Significant deviations from zero, ranging up to 200 basis points, appear already for the three-month interest rate in the pre-EMU sample, with a downward-sloping trend since 1996. The term premium associated with the three-month interest rate practically disappears for the EMU period. The overall dynamics and range

6 A trend was included in the AR(p) specification to account for the departure from stationarity which is observable in the short-term nominal interest rate series for the euro area. At each time period, the length of the AR(p) model was chosen to be the one that minimises the Schwarz criterion among lags one to 12.

7 See the Appendix for a description of the data and their source.

of the term premium for the three-month interest rate resemble closely the estimates obtained by Crespo Cuaresma et al (2004), who model pre-EMU interest rate spreads with the German short-term interest rate as depending upon expected inflation and output gap differences. For the long-term interest rates, the pre-EMU convergence to a zero term premium occurs with some delay compared to the three-month interest rate and is followed by a resurgence in the risk premia in the EMU period. The risk premia for the long-term rates estimated by this method present more persistence and higher values in the first part of the sample compared to the perfect foresight case due to the fact that the AR(*p*) model produced downward-sloping projections of the short rate also for the period where the one-month interest rate showed a stable dynamic pattern. The same line of reasoning applies to the increase in risk premia after 2001, where the decrease in nominal interest rates observed in the data was expected, according to the projections of the AR(*p*) model, to continue for longer than it actually did.

4. The predictive content of the term spread for real activity and inflation: evidence for the euro area

The results in the previous section suggest that the assumption of a constant risk premium may not hold for euro area data spanning long enough periods of time. This section investigates whether the predictive abilities of the term spread for industrial production growth and for inflation are improved by adjusting for a time-varying term premium. The adjusted term spread is given by

$$\hat{R}(n,t) - \hat{R}(m,t) = [R(n,t) - \hat{\Phi}(n,t)] - [R(m,t) - \hat{\Phi}(m,t)]$$

which can be rewritten using (4) as

$$\hat{R}(n,t) - \hat{R}(m,t) = \frac{1}{n}\sum_{i=0}^{n-1}\hat{R}(1,t+i) - \frac{1}{m}\sum_{i=0}^{m-1}\hat{R}(1,t+i) \tag{5}$$

ie we are proposing the use of the term spread implied by the REHTS with expectations formed using a simple AR(*p*) model. The differences between the adjusted and observed term spread are shown in Graphs 3 and 4, where both of them are plotted for two- and three-year bonds as the long rate and the one-month interest rate as the short rate. Graph 3 presents the observed term spread together with the term spread implied by the adjustment with perfect foresight, ie replacing expected short rates with the actually realised one-month nominal rate.[8] The discrepancies between both measures are more extreme in the pre-EMU period, where the level and dynamics of the observed term spread are interpreted mainly as premium dynamics when using the adjustment method. The same qualitative conclusion applies if the three-month interest rate is used as the short rate. Graph 4, on the other hand, presents the observed term spread and the term spread implied by the adjustment using expectations formed by means of an AR(*p*) model. Due to the fact that the simple expectation-formation mechanism tended to overestimate the decrease of the nominal short-term interest rate in the pre-EMU period, the resulting synthetic long rates are very low compared to the one-month interest rate. This implies that a negative term spread prevails for the whole pre-EMU sample, which only turns positive at the end of 1999.

The potential improvement in the predictive content of the term spread for future developments in real activity and inflation will be tested and measured in the framework of an out-of-sample forecasting exercise for the growth rate of industrial production as well as headline and core inflation in the euro area.[9]

---

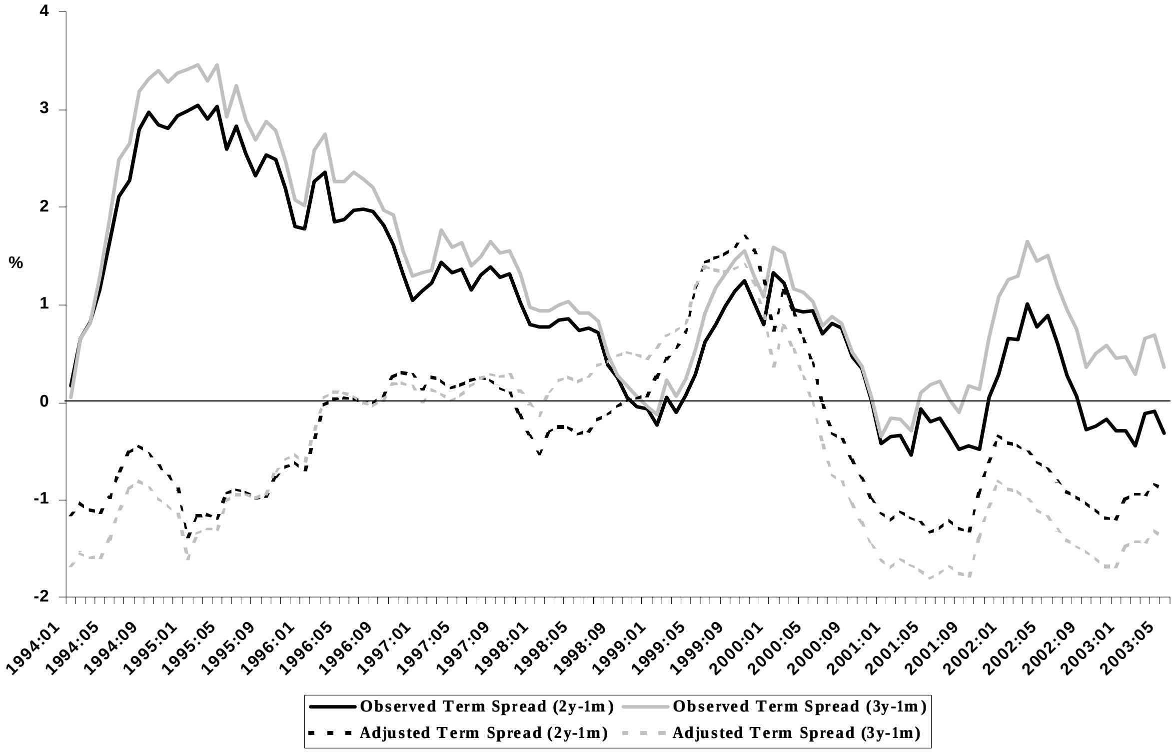
8 The adjusted long rate for the last part of the sample was computed using simple projections of the short-term interest rate using all the available data.

9 We will thus only consider what Estrella et al (2003) label a "continuous model", as opposed to a "binary model", with the latter aiming exclusively at forecasting the occurrence of recessions or the direction of change in inflation rates. Estrella et al (2003) provide evidence that binary models are more stable than those offering point forecasts of real activity. The choice of a continuous type of model for our exercise is conditioned by the fact that only one single recessionary episode has been observed in the aggregate euro area since 1990.

Graph 3

Observed and adjusted term spreads, perfect foresight

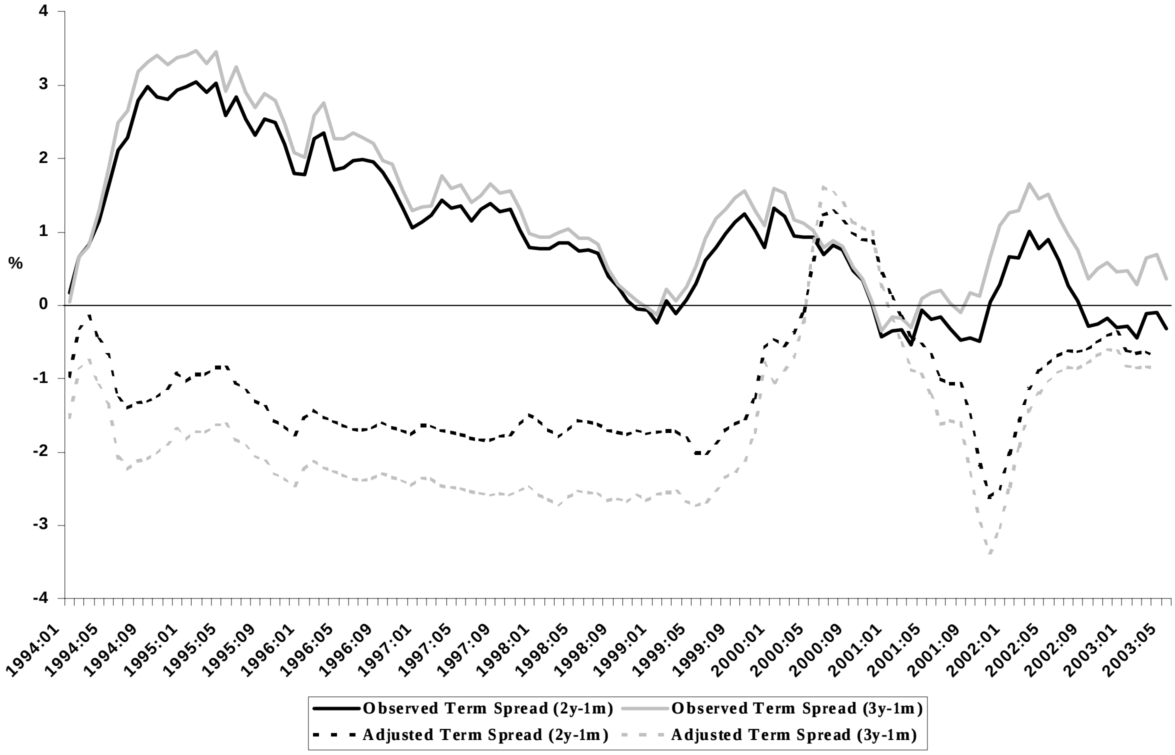
Long rate: two and three years; short rate: one month



Graph 4

Observed and adjusted term spreads, autoregressive expectations

Long rate: two and three years; short rate: one month



We will consider simple autoregressive distributed lags (ARDL(p,q)) models for forecasting industrial production growth and the inflation rate. For a given forecasting horizon h, the models estimated and used in the forecasting exercise are of the type

$$y_{t+h} = \delta + \sum_{i=0}^{p} \alpha_i y_{t-i} + \sum_{j=0}^{q} \beta_j x_{t-j} + \varepsilon_t \tag{6}$$

where y_t will alternatively be the yearly growth rate of industrial production or the inflation rate for the euro area. For a given dependent variable, x_t will alternatively be the observed and adjusted spread and ε_t is an iid random error with constant variance.

The forecasting exercise is carried out as follows. For a given value of the forecasting horizon, h, equation (6) is estimated using data up to period T using the observed spread as the x variable. With the estimated model, an h-steps-ahead out-of-sample forecast is generated. The observations for period $T+1$ are added to the estimation sample, (6) is re-estimated, and another h-steps-ahead forecast is computed. This is repeated until forecasts are obtained for all available observations of industrial production growth or the inflation rate since period $T+h$. The same procedure is then repeated for the adjusted spread as an x variable in (6). Notice that the adjustment procedure with AR(p) forecasts as expectations for the short rate which was described in the preceding section only requires data up to time t in order to obtain an estimate of $\Phi(n,t)$. The adjusted term spread assuming perfect foresight, however, uses future information for the adjustment method, so the results concerning this variable do not fulfil the usual requirements of a proper out-of-sample forecasting exercise, but are presented here for obvious comparison reasons.

The predictive ability of the different models used in the analysis will be compared in terms of root mean square forecasting error (RMSE). The h-steps-ahead RMSE of the model including variable x is given by

$$RMSE(x,h) = \sqrt{\frac{1}{N} \sum_{n=T+h}^{T+h+N} \left(y_n^{x,h} - y_n \right)^2}$$

where $y_n^{x,h}$ is the forecast of y_n obtained by the model with variable x and data ranging up to $T+n-h$, and N is the number of out-of-sample forecasts carried out. The Diebold-Mariano (Diebold and Mariano (1995), henceforth DM) test, which is described in the Appendix, will be used to compare the predictive accuracy of the models with the observed and adjusted term spread.

The results of the forecasting exercise for the rate of growth of industrial production are presented in Table 1. The procedure described above was carried out for adjusted and unadjusted term spreads with the two- and three-year bond as the long rate and the one- and three-month interest rate as the short rate. The lag lengths of the estimated ARDL(p,q) models are allowed to change with each new observation added to the in-sample period. In each replication, the lag lengths (p,q) chosen are the ones that jointly minimise the Schwarz criterion among those in the set $\{0,1,\ldots,6\} \times \{0,1,\ldots,6\}$. Table 1 reports the results of the forecasting exercise for forecasting horizons from six months to two years ahead, at six-month steps. In all cases, the first in-sample period was January 1994-January 1998, and forecasts were computed up to December 2002, the last observation of annual industrial production growth available. The last row of Table 1 presents the forecasting results for a simple autoregressive (AR) process, which is the natural benchmark of comparison if we want to evaluate the predictive content of the term spread in models such as (5).[10] The AR process is defined like in (6) without the second summation term on the right-hand side. The DM test statistic is provided in the table for those models that show better predictive abilities than the benchmark, and refers to the test for equal predictive accuracy against the AR model.

The results for the observed term spread confirm and expand the conclusions in Berk and van Bergeijk (2000, 2001). The simple AR model, which excludes the information contained in the term spread, performs better than the models including the unadjusted yield curve information in terms

[10] The procedure based on the Schwarz criterion was also used for choosing the optimal lag length for the AR process in each period. Qualitatively, the results remain unchanged if an unconstrained vector autoregression (VAR) using inflation and output growth data is used as the benchmark model. At most forecasting horizons, the simple AR model actually outperforms the VAR model in terms of forecasting error for output growth and inflation.

of RMSE for all forecasting horizons with the exception of two-years-ahead forecasts. For this forecasting horizon, only the model containing the term spread between the two-year bond and the three-month interest rate obtains a marginal improvement in the RMSE compared to the AR model, which is, however, insignificant according to the DM test.

Table 1

Forecasting comparison: industrial production growth

		RMSE			
		6 months	**12 months**	**18 months**	**24 months**
Adjusted spread (perfect foresight)					
Long rate	**Short rate**				
2 years	1 month	1.69 (–2.43***)	2.38 (–1.38*)	3.23	4.45
	3 months	1.71 (–2.09**)	2.29 (–1.61*)	3.36	4.03
3 years	1 month	1.73 (–2.03**)	2.08 (–1.87**)	2.90 (–0.01)	4.77
	3 months	1.72 (–2.07**)	1.99 (–2.06**)	2.95	4.56
Adjusted spread (AR expectations)					
2 years	1 month	2.67	2.63 (–0.46)	1.78 (–2.45***)	2.62 (–1.31*)
	3 months	2.65	2.67 (–0.32)	1.76 (–2.47***)	2.57 (–1.29*)
3 years	1 month	2.75	2.95	1.91 (–2.94***)	2.57 (–1.40*)
	3 months	2.73	2.97	1.89 (–2.82***)	2.63 (–1.52*)
Observed spread					
Long rate	**Short rate**				
2 years	1 month	2.82	3.53	4.05	6.07
	3 months	2.27	3.40	3.60	3.06 (–0.08)
3 years	1 month	2.78	3.50	4.35	5.83
	3 months	2.27	3.44	3.56	3.14
Benchmark AR model		2.09	2.82	2.91	3.07

Note: Numbers in parenthesis refer to the DM test statistic of the corresponding model against the AR model, asymptotically standard normal distributed. * (**) [***] refers to significance at 10% (5%) [1%] significance level.

While the results for the observed spread caution against the use of the information contained in the yield curve when forming predictions for real activity developments in the euro area, the forecasting exercise reaches a very different conclusion for the adjusted term spread. For forecasting horizons up to and including one year, the models including the premium-adjusted term spread with perfect foresight uniformly outperform all other models, independently of the interest rates used as long and short rates in the computation of the spread. The results of the DM test against the AR model conclude that the observed difference in predictive ability is significant in all cases. The predictive content of the adjusted term spread with perfect foresight ceases to exist, however, for longer forecasting horizons. For 18-months-ahead predictions, only one of the models with adjusted term spreads and perfect foresight presents an insignificantly lower forecasting error than the AR model, and for the two-year forecasting horizon, all models including the adjusted term spread are outperformed by the minimal benchmark AR model.

The improvement in the predictive ability of the premium-adjusted term spread with perfect foresight is not surprising, as it includes actual information on the development of short-term interest rates in the out-of-sample period. The forecasts obtained from the premium-adjusted term spread using AR(*p*) expectations, by contrast, are based exclusively on in-sample data. The results for long-term forecasts with the model containing the adjusted term spread using AR(p) expectations indicate an overwhelming improvement of the prediction error for forecasting horizons higher than a year ahead. Independently of the rates used to form the term spread, all models including this variable outperform significantly the benchmark model at 18- and 24-months-ahead horizons, with reductions of the RMSE up to 40% compared with the simple AR model and 55% if compared to the model including the observed spread. The fact that the forecasting horizon where improvements are significant has shifted forward as compared to the perfect foresight case is explained by the relatively high inertia of the autoregressive forecasts (changes in direction of the trend which is estimated when forming expectations tend to be picked up with around 12 months' delay).

The results are very different if the variable to be predicted is inflation. Table 2 presents the results for the headline inflation rate in the euro area (defined as yearly change in the harmonised index of consumer prices), and Table 3 presents the results for the core inflation rate (defined as yearly change in the harmonised index of consumer prices excluding energy and unprocessed food).

Table 2

Forecasting comparison: headline inflation

		RMSE			
		6 months	**12 months**	**18 months**	**24 months**
Adjusted spread (perfect foresight)					
Long rate	**Short rate**				
2 years	1 month	0.58	1.07	1.76	2.29
	3 months	0.59	1.06	1.68	2.26
3 years	1 month	0.55	1.00	1.53	2.22
	3 months	0.55	0.99	1.55	2.21
Adjusted spread (AR expectations)					
2 years	1 month	0.48	0.89 (–0.71)	1.64	2.44
	3 months	0.46 (0.52)	0.76 (–1.33*)	1.40	2.25
3 years	1 month	0.49	2.33	1.67	2.39
	3 months	0.48	0.91 (–0.24)	1.52	2.27
Observed spread					
Long rate	**Short rate**				
2 years	1 month	0.55	0.98	1.43	2.16
	3 months	0.51	1.15	1.59	2.12
3 years	1 month	0.56	1.02	1.46	2.18
	3 months	0.51	1.10	1.57	2.15
Benchmark AR model		0.48	0.96	1.24	1.56

Note: Numbers in parenthesis refer to the DM test statistic of the corresponding model against the AR model, asymptotically standard normal distributed. * refers to significance at 10% significance level.

Although the adjusted term spread using AR(p) expectations achieves lower forecast errors than all other models in some cases for forecasting horizons up to one year, only the model with the adjusted two-year–three-month spread is able to outperform the benchmark significantly for one-year-ahead predictions. Neither the information contained in the observed term spread nor that contained in the adjusted term spread with perfect foresight improves the predictions on inflation based on its own past history at any forecasting horizon.[11]

Table 3

Forecasting comparison: core inflation

		RMSE			
		6 months	**12 months**	**18 months**	**24 months**
Adjusted spread (perfect foresight)					
Long rate	**Short rate**				
2 years	1 month	0.35	0.77	1.17	1.46
	3 months	0.35	0.75	1.17	1.46
3 years	1 month	0.28 (–0.67)	0.45 (–0.97)	0.89 (–0.88)	1.23 (–1.39*)
	3 months	0.28 (–0.58)	0.48 (–0.80)	0.93 (–0.77)	1.26 (–1.21)
Adjusted spread (AR expectations)					
2 years	1 month	0.52	1.29	0.93 (–1.61*)	1.00 (–1.24)
	3 months	0.49	1.36	0.92 (–1.16)	0.89 (–2.04**)
3 years	1 month	0.58	2.18	1.78	0.99 (–1.16)
	3 months	0.54	1.66	1.89	0.92 (–1.72**)
Observed spread					
Long rate	**Short rate**				
2 years	1 month	0.46	0.95	1.03 (–0.96)	1.21 (–2.19**)
	3 months	0.41	0.82	1.20	1.66
3 years	1 month	0.45	0.89	1.00 (–1.34*)	1.21 (–2.35**)
	3 months	0.41	0.78	1.27	1.64
Benchmark AR model		0.34	0.68	1.13	1.39

Note: Numbers in parenthesis refer to the DM test statistic of the corresponding model against the AR model, asymptotically standard normal distributed. * (**) refers to significance at 10% (5%) significance level.

However, the term spread, in both its adjusted and unadjusted form, seems to be partly useful for obtaining forecasts of core inflation. The results in Table 3 show that the models including the observed term spread with the one-month interest rate significantly outperform the benchmark model in predicting core inflation rates at long horizons. The improvement is still greater if the adjusted spread with AR(*p*) expectations is used, with reductions of the RMSE over the benchmark of more than 35%. The model with the adjusted term spread using the difference between the adjusted two-year bond rate and the adjusted three-month interest rate presents the best forecasting abilities at the two-years-ahead horizon, and outperforms (with a DM test statistic of ˜1.71) the best model among those using the observed spread. Surprisingly, marginal improvements over the benchmark are observed for the adjusted term spread with perfect foresight only for two-years-ahead forecasts, and these are of a small magnitude compared to the improvements obtained using the adjustment with AR(*p*) expectations.

[11] Estrella et al (2003) note that the relationship between real activity and the term spread is of a more stable nature than that between inflation and the term spread. Our results for the inflation rate may as well reflect the existence of one or more structural breaks in the underlying data-generating process.

Given the way in which the adjustment takes place with AR(p) expectations, the adjusted term spread is computed using exclusively information on the short-term interest rate. The results presented above could thus be interpreted as evidence that the predictive power of the term spread is determined by the dynamics in the short-term rate. The aggregation implied by the REHTS is, according to the results presented, a useful way of disentangling the part of the term spread whose dynamics actually contain information on future macroeconomic developments. If the adjustment method is to be relied upon, one would expect that no significant information on future developments in real activity and inflation should be present in the risk premia estimates plotted in Graph 2. Table 4 presents the results of the forecasting exercise explained above using the risk premia implied by the decomposition with AR(p) expectations as the x variable.

Table 4

Forecasting comparison results for risk premia estimates

Risk premia		**RMSE**			
		6 months	**12 months**	**18 months**	**24 months**
Industrial production growth					
Long rate	**Short rate**				
2 years	1 month	2.78	4.37	4.35	4.21
	3 months	2.80	4.52	3.83	3.73
3 years	1 month	2.81	4.74	4.54	4.14
	3 months	2.81	5.22	3.99	3.65
Headline inflation					
2 years	1 month	0.50	0.94 (0.14)	2.45	2.34
	3 months	0.70	1.08	1.59	2.16
3 years	1 month	0.50	0.94 (0.17)	1.73	2.25
	3 months	0.67	1.10	1.59	2.38
Core inflation					
2 years	1 month	0.45	0.88	1.26	1.81
	3 months	0.42	0.78	1.47	1.32 (0.95)
3 years	1 month	0.45	1.07	1.34	1.50
	3 months	0.44	0.81	1.55	1.41

Note: Numbers in parenthesis refer to the DM test statistic of the corresponding model against the AR model, asymptotically standard normal distributed.

The results in Table 4 present the RMSE obtained in the forecasts when using the risk premium with respect to the one- and three-month interest rate as explanatory variables in the out-of-sample exercise presented above. There is no improvement on the models where industrial production growth, headline inflation or core inflation are explained by their own past for any forecasting horizon and any risk premium estimate. These results indicate that the decomposition used tends to be successful in isolating the part of the term spread with predictive properties for industrial production growth and, notwithstanding the limitations of this link, also with inflation.

The method used to adjust the term spread for time-varying risk premia renders an adjusted term spread composed exclusively of autoregressive expectations on the short rate, which are aggregated according to the REHTS using (5). Whether imposing the structure implied by (5) actually improves the forecasting abilities of the term spread as compared to using exclusively the information embodied in the short rate data without the restrictions implied by the aggregation method can also be checked empirically. Table 5 presents the results of the forecasting exercise using the monthly change in the short rate as an explanatory variable in (6).[12] There is no evidence of significant improvement over the forecasts of the benchmark model for any variable at any forecasting horizon. The results for the short rate can be interpreted as a robustness check of the simple methodology proposed, and they draw attention to the empirical relevance of the method of aggregation of expectations implied by the REHTS when assessing the predictive abilities of the term spread for output growth and inflation.

Table 5

Forecasting comparison results for the short rate

RMSE			
6 months	**12 months**	**18 months**	**24 months**
Industrial production growth			
2.11	2.84	2.86 (–0.54)	3.23
Headline inflation			
0.50	0.87 (–0.89)	1.57	2.27
Core inflation			
0.36	0.73	1.15	1.36 (–1.04)

Note: Numbers in parenthesis refer to the DM test statistic of the corresponding model against the AR model, asymptotically standard normal distributed.

5. Conclusions and paths of further research

This paper reinvestigates the informational content of the yield spread for real activity and inflation for the euro area aggregate. The motivation is threefold. First, at the theoretical level, a number of possible channels have been put forward in the literature that would suggest a systematic empirical relationship between the yield spread and current and/or future real activity. Second, at the level of data availability, four and a half years of genuine euro area data make it worthwhile to investigate the issue empirically. Third, previous research has not paid attention to the substantial difference of the monetary policy regime in place prior to the start of EMU, which may have strongly influenced risk premia over time. Contrary to previous research on the euro area, this paper explicitly pays attention to disturbances of the term spread from time-varying risk premia. We put forward a simple, purely empirical adjustment procedure for a time-varying term premium based on the rational expectations hypothesis of the term structure, and find that significant improvements can be achieved in the predictive content of the term spread if the dynamics of the risk premium are taken into account in its computation.

The results of a forecasting exercise using adjusted and unadjusted term spreads show that, for the euro area aggregate, modelling the risk premium adequately is a necessary requirement in order to exploit the information embodied in the term spread for predictions in the development of real activity and inflation. Regarding real activity, of all possible models including the term spread, only those where the adjustment was performed were able to deliver significantly better medium-run forecasts than simple models where the growth rate of industrial production is explained by its own past history.

[12] Augmented Dickey-Fuller tests could not reject the existence of a unit root in the series of one-month rates at any reasonable significance level.

For forecasting horizons exceeding one year, the models including the premium-adjusted term spread, where the expectations on the short rate are modelled through a simple autoregressive model, uniformly outperform all other models. This result arises independently of the interest rates used as long and short rates in the computation of the spread. For the case of inflation, however, the results are more mixed, but evidence of improvement in the forecasting abilities of the term spread after the premium adjustment was provided for two-years-ahead forecasts of core inflation.

We conclude that, if distortions arising from time-varying risk premia are filtered out, the term spread can - despite the substantial limitations imposed on econometric estimates by the necessity to use *synthetic* pre-EMU data - nevertheless serve as one useful indicator (among others) to gauge future developments in real activity and, to a lesser extent, (core) inflation. In this sense, it seems worth monitoring as part of the "economic analysis" within the framework of the Eurosystem's monetary policy strategy. In particular, after adjusting for the existence of a time-varying risk premium, the term spread could be useful in order to check the robustness of forecasts produced by more extensive macroeconomic models.

An alternative reading of our results is that - for the euro area - using information embodied in short-term interest rates yields better forecasting results for both real activity and (core) inflation than the term spread. In other words, the medium-term end of the yield curve used in our study seems to contain no additional information. However, our results show that the aggregation of expectations on short rates implied by the REHTS seems to play an important role in the predictive properties of the adjusted term spread. This interpretation would raise serious questions about the widespread reference by financial analysts and policy commentators to the (term-spread-unadjusted) yield curve as a market expectations indicator.

Finally, it may also be that the policy regime break induced by the inception of EMU pollutes empirical analysis at this stage too much. In this case, the issue might be resolved over time, as longer time series become available and the regime break becomes an event which is only relevant for the beginning of the sample. Linked to that, it may also be that the use of more sophisticated econometric methods will in the future be able to shed some light on the reasons for the predictive failure of the observed spread in the euro area.

In this vein, Venetis et al (2003) provide evidence concerning the existence of threshold effects in the relationship between the term spread and real activity for Canada, the United Kingdom and the United States. The use of non-linear time series models to assess the informational content of the term spread on future developments in real activity can thus be seen as a possible avenue of future research in order to provide further evidence on the leading indicator properties of the slope of the yield curve.

Appendix

Data sources

- One-month interest rate, euro area aggregate. Source: Datastream.
  Range: November 1990-May 2003.
- Three-month interest rate, euro area aggregate. Source: Bank for International Settlements.
  Range: January 1990-May 2003.
- Two-year bond yield, secondary market, benchmark, euro aggregate. Source: Bank for International Settlements.
  Range: January 1994-May 2003.
- Three-year bond yield, secondary market, benchmark, euro aggregate. Source: Bank for International Settlements.
  Range: January 1994-May 2003.
- Industrial production index, euro aggregate, seasonally adjusted. Source: Eurostat.
  Range: January 1990-December 2002.
- Harmonised index of consumer prices (all items and all items excluding energy and unprocessed food), euro aggregate, seasonally adjusted using the Census X12 method. Source: European Central Bank.
  Range: January 1990-May 2003.

The Diebold-Mariano test for comparing predictive accuracy

The DM test is an asymptotic test for the null of equal predictive accuracy of two models. In the framework proposed above, consider two models using variables x_1 and x_2 respectively. For a given forecasting horizon h, the null hypothesis in the DM test is that

$$d_n = E\left[g(e_{1n}) - g(e_{2n})\right] = 0$$

where e_{1n} is the forecasting error produced by the model with variable x_1 when forecasting Δy_n (that is, $e_{1n} = \Delta y_n^{x_1,h} - \Delta y_n$), e_{2n} is defined analogously for x_2, and $g(z)$ is a prespecified loss function associated with the forecast error. In our case, the loss function is a quadratic one, so that $g(z) = z^2$. The DM test is based on the observed average forecast error difference, $\bar{d}$, The DM test statistic is given by

$$S_1 = [\hat{V}(\bar{d})]^{-1/2}\bar{d}$$

where $\hat{V}(\bar{d})$ is an estimate of the asymptotic variance of $\bar{d}$, given by

$$\hat{V}(\bar{d}) = \frac{1}{N}\left(\hat{\gamma}_0 + 2\sum_{k=1}^{h-1}\hat{\gamma}_k\right)$$

where $\hat{\gamma}_k$ is the k-th order sample autocovariance of the forecasting error difference series. The asymptotic distribution of S_1 is standard normal, so tests for equality of predictive accuracy between different models can be easily carried out.[13]

[13] The DM test methodology is not free of criticism. For a recent critical assessment of testing predictive accuracy using the DM test statistic, see Kunst (2003).

References

Berk, J M and P van Bergeijk (2000): "Is the yield curve a useful information variable for the Eurosystem?", *European Central Bank Working Paper Series*, no 11.

——— (2001): "On the information content of the yield curve: lessons for the Eurosystem", *Kredit und Kapital*, 34, pp 28-47.

Campbell, R H (1988): "The real term structure and consumption growth", *Journal of Financial Economics*, 22, pp 305-33.

Carstensen, K and J Hawellek (2003): "Forecasting inflation from the term structure", *Review of World Economics - Weltwirtschaftliches Archiv*, 139, pp 306-23.

Crespo Cuaresma, J, E Gnan and D Ritzberger-Grünwald (2004): "The natural rate of interest: a euro-area perspective", *Empirica*, vol 31, pp 185-204.

Diebold, F X and R S Mariano (1995): "Comparing predictive accuracy", *Journal of Business and Economics Statistics*, 13, pp 253-63.

Estrella, A (2003): *Why does the yield curve predict output and inflation?*, Federal Reserve Bank of New York, mimeo.

Estrella, A and F S Mishkin (1997): "The predictive power of the term structure of interest rates in Europe and the United States: implications for the European Central Bank", *European Economic Review*, 41, pp 1375-401.

Estrella, A, A P Rodrigues and S Schich (2003): "How stable is the predictive power of the yield curve? Evidence from Germany and the United States", *The Review of Economics and Statistics*, 85, pp 629-44.

Hu, Z (1993): "The yield curve and real activity", *International Monetary Fund Staff Papers*, 40, pp 781-806.

Kunst, R M (2003): "Testing for relative predictive accuracy: a critical viewpoint", Institute for Advanced Studies, *Economics Series*, no 130, Vienna.

Smets, F and K Tsatsaronis (1997): "Why does the yield curve predict economic activity? Dissecting the evidence for Germany and the United States", *BIS Working Papers*, no 49.

Stock, J H and M W Watson (2003): "Forecasting output and inflation: the role of asset prices", *Journal of Economic Literature*, vol XLI, pp 788-829.

Tzavalis, E and M R Wickens (1996): "Forecasting inflation from the term structure", *Journal of Empirical Finance*, 3, pp 103-22.

Venetis, I A, I Paya and D A Peel (2003): "Re-examination of the predictability of economic activity using the yield spread: a nonlinear approach", *International Review of Economics & Finance*, 12, pp 187-206.

Extracting growth and inflation expectations from financial market data

Lauri Kajanoja,[1] Bank of Finland

1. Introduction

Financial market prices are affected by market participants' expectations concerning future macroeconomic developments. However, expectations regarding real GDP growth, for example, cannot be directly observed in the price quotations for a financial market instrument. In order to gain information on such expectations, one needs to employ economic models in addition to financial market data.

Financial market participants' expectations concerning macroeconomic developments are obviously of great interest, not least to economic policymakers. These expectations, as they are reflected in financial market prices, are based on a huge amount of information. Naturally, they can be wrong, and each individual may disagree with them. Nonetheless, knowledge of market expectations does make it easier to understand current economic developments and to form one's own expectations concerning the future.

Various measures of market expectations concerning macroeconomic developments have been put forward. A widely used measure of inflation expectations is the so-called "break-even" inflation rate derived from the yield of an inflation-indexed bond. Break-even inflation rates are discussed for example by eg Sack (2000) and Scholtes (2002). Another market-based measure of inflation expectations can be derived from inflation-linked swaps, as reported by the ECB (2003). Measuring market expectations of real output growth seems to be a more formidable task. For measuring the market's perception of the output gap, Martin and Sawicki (2003) propose a method based on an inverted Taylor rule. Taking a broader view on measuring market expectations, one can also consider indicator models of growth and inflation that use financial market variables as inputs. Such models are widely used in short-term macroeconomic forecasting.

Stock prices and interest rates can be interpreted to yield information concerning market expectations of future output growth and inflation. High stock prices indicate fast expected growth of companies' earnings and dividends in the future. Long-term interest rates reflect expectations concerning both inflation and output in the long run; according to standard macroeconomic theory, the long-term interest rate is related to expected long-run output growth. However, stock prices and interest rates do not as such provide direct measures of real output growth or inflation expectations.

This paper presents a new framework for measuring market expectations concerning long-run inflation and real output growth. The method combines items of information contained in stock prices and interest rates. The framework can be directly applied to measuring expectations in real time. As inputs, it uses interest rates and dividend/price ratios for equity indices. In addition, equity index futures are utilised in gauging short-run expectations. The framework is based on economic theory. It builds on three elements: first, a dividend discount model of stock prices is used; second, it is assumed that expected long-run dividend growth is proportional to expected long-run GDP growth; and third, it is assumed that there is a stable linear relationship between the long-term real interest rate and the expected long-run real GDP growth.

The paper is organised as follows. Section 2 describes the methodology used. Section 3 presents the results; that is, the series of extracted growth and inflation expectations for the euro area and the United States. Section 4 concludes.

[1] I am grateful to Jarmo Kontulainen, Hanna-Leena Männistö, Nicolas Rautureau, Tuomas Saarenheimo, Juha Tarkka, Nico Valckx, Jouko Vilmunen, and the participants in the Bank of Finland Research Department and Economics Department seminars and the BIS Central Bank Economists' Meeting for useful comments and suggestions.

2. Framework

In this study, market expectations for long-run GDP growth and inflation are measured using the following data as inputs: interest rates, dividend/price ratios of equity indices, and equity index futures. This section presents the framework used to carry out the measurement. Section 2.1 describes how the dividend discount model is used in the framework. The method for deriving long-run expectations is further developed in Section 2.2. Section 2.3 then describes how near-term dividend growth expectations are measured using data on equity index futures. The near-term expectations are measured in order to improve the measurement of long-run expectations. Finally, Section 2.4 gives parameter values.

2.1 Expected dividend growth

Following the dividend discount model, we start from the assumption that stock prices equal expected discounted future dividends. The discount rate is the expected return on equity capital, which can be approximated by the risk-free interest rate plus an equity premium. Here, the latter is assumed to be constant. Therefore, the price of a stock at the end of period t, P_t, can be expressed as:

$$P_t = \sum_{j=1}^{\infty} \frac{D_{t+j|t}}{\left(1+i_{j,t}+\omega\right)^j} = D_t \sum_{j=1}^{\infty} \frac{\left(1+n_{j,t+j|t}\right)^j}{\left(1+i_{j,t}+\omega\right)^j}, \tag{1}$$

where $D_{t+j|t}$ denotes dividends paid during period $t+j$ as expected at the end of period t, $i_{j,t}$ denotes the risk-free interest rate in maturity j at the end of period t, ω denotes the equity premium, and $n_{j,t+j|t}$ denotes the end of period t expectation of the growth rate of nominal dividends from period t till period $t+j$, in annual terms. In other words, the first subscript denotes the length of the time horizon for the variable, and the second subscript indicates when the value of the variable is realised. We take the length of a time period to be one year.

We do not assume that dividend growth is expected to be constant in the future. Instead, we decompose the expected dividend growth into short-run and long-run expectations. We use the following "term structure" assumption for the expected nominal dividend growth:

$$1+n_{j,t+j|t} = \left(1+n_{1,t+1|t}\right)^{\frac{1}{j}} \left(1+n_{LR|t}\right)^{\frac{j-1}{j}}, \tag{2}$$

where $n_{LR|t}$ denotes expected long-run nominal dividend growth. In addition, we use a similar approximation for the term structure of the discount rate:

$$1+i_{j,t}+\omega = \left(1+i_{1,t}+\omega\right)^{\frac{1}{j}} \left(1+i_{LR,t}+\omega\right)^{\frac{j-1}{j}}, \tag{3}$$

where $i_{LR,t}$ denotes the long-term risk-free interest rate at the end of period t. The empirical definition of this variable will be given in Section 2.4 below.

Equations (1) to (3) imply, as an approximation, that:

$$n_{LR|t} = i_{LR,t} + \omega - \frac{D_t}{P_t} \frac{1+i_{LR,t}}{1+i_{1,t}} \left(1+n_{1,t+1|t}\right). \tag{4}$$

According to equation (4), market expectations concerning long-run nominal dividend growth can be inferred from current financial market prices, past dividends, and estimates of near-term dividend growth expectations and the equity premium.

2.2 Measuring GDP growth and inflation expectations

We assume that expected long-run dividend growth varies in proportion to expected long-run GDP growth. For an imaginary stock price index covering all firms in an economy, one could argue that these two should move one-to-one. For the stock price indices used here it is natural to assume that the expected long-run dividend growth rate varies more than the expected long-run GDP growth rate for the whole economy. Therefore, we assume that:

$$n_{LR|t} - \pi_{LR|t} = \alpha + \beta g_{LR|t}, \tag{5}$$

where $\pi_{LR|t}$ denotes expected long-run inflation, $g_{LR|t}$ expected long-run real GDP growth, and α and β positive constants. Equation (5) states the relationship between the expected real long-run dividend growth and the expected real long-run GDP growth. As discussed in Section 2.4 below, we will set β close to 2 when the S&P 500 Index for the United States is considered.

Let us next introduce an assumption concerning the relationship between expected long-run real GDP growth and the long-term real rate of interest. A standard consumption Euler equation from a representative consumer model combined with a market clearing condition, saying that consumption equals output, yields:

$$r_{LR,t} + \omega_Y \approx \frac{1}{\delta}\frac{u'[Y_t]}{u'\left[(1+g_{LR|t})Y^t\right]} - 1, \tag{6}$$

where δ denotes the discount factor, u' is the first derivative of the period utility function, and Y_t denotes period t real consumption, which equals real output. ω_Y denotes a risk premium. It is not assumed to be equal to the ω of equation (1), since β is allowed to differ from 1 in equation (5). The long-term real interest rate is denoted by $r_{LR,t}$, and defined as:

$$r_{LR,t} = i_{LR,t} - \pi_{LR|t}. \tag{7}$$

Equation (6) can be linearised to yield, as an approximation:

$$r_{LR,t} = \rho + \lambda g_{LR|t}. \tag{8}$$

This linearisation holds for positive constants ρ and λ, the latter of which represents the inverse of the elasticity of intertemporal substitution.

Equations (4), (5), (7) and (8) can be combined to yield the following system of equations:

$$g_{LR|t} = \frac{\rho+\gamma}{\beta-\lambda} - \frac{1}{\beta-\lambda}\frac{D_t}{P_t}\frac{1+i_{LR,t}}{1+i_{1,t}}\left(1+n_{1,t+1|t}\right), \tag{9}$$

$$\pi_{LR|t} = i_{LR,t} - \rho - \frac{\lambda(\rho+\gamma)}{\beta-\lambda} + \frac{\lambda}{\beta-\lambda}\frac{D_t}{P_t}\frac{1+i_{LR,t}}{1+i_{1,t}}\left(1+n_{1,t+1|t}\right), \tag{10}$$

where γ denotes a constant defined as $\gamma \equiv \omega - \alpha$. Equations (9) and (10) express the expected long-run real GDP growth and inflation in terms of (1) current observable variables: D_t, P_t, $i_{LR,t}$, $i_{j,t}$; (2) parameters: λ, ρ, γ; and (3) expected one-year-ahead growth in nominal dividends $n_{1,t+1|t}$.

In the next section we deal with the near-term expectations $n_{1,t+1|t}$. After that, we set values for the parameters λ, ρ, γ and β. Then we are ready to use equations (9) and (10) empirically to extract market expectations for the euro area and for the United States.

2.3 Measuring short-run dividend growth expectations

The framework presented in this paper is constructed in order to extract market expectations concerning long-run developments. Sometimes when the dividend discount model is utilised in extracting market expectations, the expected dividend growth rate is assumed to be constant in the future. Expectations derived in such a way reflect, to a large extent, short-run expectations. This is because short-run expectations seem to vary more than long-run expectations, and because they have a larger weight in the dividend discount model due to the discounting. Therefore, we deal with long- and short-run expectations separately, as shown in Section 2.1.

Regarding expectations concerning near-term stock returns, one way to proceed would be to use stock analysts' bottom-up predictions. However, in the current context this approach would have an obvious drawback: the predictions are not available on a real-time basis. In addition, such predictions are known to have a significant upward bias. The approach chosen here is therefore to use the information contained in equity index futures quotations.

2.3.1 Expectations and equity index futures: the idea

Short-run dividend growth expectations can be extracted from financial market data using the principle of equation (1) and the prices of equity index futures, financial derivatives whose underlying assets are equity indices. The value of an equity index future reflects the market expectations concerning the value of the index in the future as well as expectations regarding dividends paid out before the future matures.

Let us start by stating that:

$$\frac{D_{t+1|t} + F_{t+1,t}}{P_t} = 1 + i_{1,t} + \omega_{D1}, \tag{11}$$

where $F_{t+1,t}$ denotes the end of period t market delivery price for an equity index future concerning a contract maturing at the end of period $t+1$,[2] and ω_{D1} denotes a risk premium.

The left-hand side of equation (11) is the expected gross return from an investment strategy where equities underlying the index are bought in period t and sold in period $t+1$ for a price set in a futures contract made in period t. In practice this means holding the stocks for one period and hedging against stock price movements by selling short equity index futures in period t. The expected return from this strategy must equal the right-hand side of the equation, that is, 1 plus the risk-free interest rate plus the risk premium ω_{D1} related to the uncertainty concerning D_{t+1} as of time t. This risk premium is related to but not equal to the ω of equation (1). It is likely to be very small and can safely be ignored in the following calculations.

Equation (11) shows that the difference between the current equity index value P_t and the futures contract price $F_{t+1,t}$ reflects two things: expected next period dividends and the discount rate. The larger the expected next period dividends, the smaller the futures contract price, other things being equal. This reflects the fact that next period dividends will be paid out before the futures contract is settled, and paying out dividends decreases the value of a firm, ceteris paribus. Equation (11) naturally holds only for equity indices not adjusted for cash dividends, that is, those that are not so-called total return indices. Most widely used equity indices, including the ones used in this study, are not total return indices.

Using the notation $1 + n_{1,t+1|t} = D_{t+1}/D_t$ and ignoring the risk premium, equation (11) can be rewritten as:

$$n_{1,t+1|t} = \frac{P_t}{D_t}\left(1 + i_{1,t} - \frac{F_{t+1,t}}{P_t}\right) - 1. \tag{12}$$

Equation (12) shows that one can infer the expected one-period nominal dividend growth $n_{1,t+1|t}$ from the values of $F_{t+1,t}$, P_t, D_t and $i_{1,t}$, all of which are observable at the end of period t.

2.3.2 Expectations and equity index futures: a detailed account

In practice, equity index futures exist only for certain maturity dates. Rather than using equation (12) empirically, it is simpler to use data on equity index futures with different maturities. In this section, we use daily frequency in time notations. In the empirical analysis we will use data from money markets which have adopted the so-called actual/360 method for interest rate calculations, also known as the 365/360 method. Therefore, using daily frequency and annualised interest rates, equation (11) should be written as:

$$P_t = \frac{F_{T,t} + D_{(t,T)|t}}{1 + \frac{T-t}{360} \cdot i_{T-t,t}}, \tag{13}$$

where T denotes the maturity date of the future, and $D_{(t,T)|t}$ denotes the day t expectation of the day T value of dividends that will be paid during the days $t+1, t+2, ..., T$. The maturity of the relevant interest rate is now $T-t$ days. When equation (13) is written for two different future dates, $T = T_1$ and $T = T_2$, combining these two equations by eliminating P_t yields:

[2] Here, we use an index future for which the delivery price is quoted in terms of the value of the index itself.

$$\frac{D_{(t,T_2)|t}}{1+\frac{T_2-t}{360}\cdot i_{T_2-t,t}}-\frac{D_{(t,T_1)|t}}{1+\frac{T_1-t}{360}\cdot i_{T_1-t,t}}=\frac{F_{T_1,t}}{1+\frac{T_1-t}{360}\cdot i_{T_1-t,t}}-\frac{F_{T_2,t}}{1+\frac{T_2-t}{360}\cdot i_{T_2-t,t}}. \tag{14}$$

Equation (14) can be rewritten to express the expected divident stream as:

$$D_{(T_1,T_2)\,|\,t}=\frac{1+\frac{T_2-t}{360}\cdot i_{T_2-t,\,t}}{1+\frac{T_1-t}{360}\cdot i_{T_1-t,\,t}}F_{T_1,\,t}-F_{T_2,\,t}\,, \tag{15}$$

where $D_{(T_1,T_2)|t}$ denotes the day t expectation of the day T_2 value of the dividends that will be paid out during the days $T_1+1, T_1+2, ..., T_2$.

We approximate the near-term annual expected dividend growth rate $n_{1,t+1|t}$ by:

$$n_{1,t+1|t}=\frac{D_{(T_1,T_2)|t}}{D_{(T_1-365,T_2-365)}}-1\,, \tag{16}$$

where the dividends in the denominator are already observed on day t. Equation (16) shows how the near-term dividend growth expectations can be extracted from the prices of equity index futures and money market interest rates.[3] In practice, the equity index futures used mature at the end of each quarter. Therefore, the numerator of equation (16) refers to dividends paid out during some quarters in the future, and the denominator refers to the dividends that were paid out during the same quarters one year earlier.

Based on equation (16), Figures 1 and 2 show the expected near-term nominal dividend growth rates for two equity indices. The Standard & Poor's 500 Index represents US stocks, and the Dow Jones EURO STOXX 50 represents euro area stocks. The series depicted in Figures 1 and 2 result from using equity index futures such that the growth rate given by the right-hand side of equation (16) refers to the one that is expected to prevail about half a year into the future.

To be exact, for the S&P 500 Index, we use prices for the futures contract that is the next one to mature and for the fourth one to mature. This means that we are measuring market expectations concerning the dividends that will be paid out during the next three full calendar quarters. For the DJ EURO STOXX 50 index, we deal with dividends to be paid out during the next two full calendar quarters. In the case of the DJ EURO STOXX 50 the contracts that mature further in the future do not exist, and in the case of the S&P 500 they have a shorter history. As the empirical counterparts for the interest rates that appear in equation (15) we use money market interest rates, linearly interpolated for different maturities when necessary.

The $n_{1,t+1|t}$ series shown in Figures 1 and 2 are not given directly by equation (16). Two modifications are made to the series. First, to smooth out what seems to be noise, we use moving averages: the past 90 days moving average for the S&P 500 and the past 30 days for the DJ EURO STOXX 50. In addition, there seem to be some premia affecting the futures prices or some other institutional factors that remain unaccounted for, so that the variances of the series given by equation (16) are implausibly large. Therefore, we regress realised ex post dividend growth series on the series given by equation (16) and use the fitted values from those linear regressions as the $n_{1,t+1|t}$ series shown in Figures 1 and 2.[4]

---

[3] For leap years the figure 365 is replaced by 366 in the subscript of the denominator.

[4] All available data are used in the estimations. For the S&P 500 the estimation period extends from the third quarter of 1991 to the third quarter of 2002, and for the DJ EURO STOXX 50 from the first quarter of 1991 to the third quarter of 2002. The data for the explanatory variable are mid-quarter values. We use ordinary least squares. The slope estimates are greater than zero, as expected, and in the case of the S&P 500 the estimate is statistically significant.

2.4 Parameter values

Now we are almost ready to use equations (9) and (10) in extracting expectations. What remains to be done is to set the values for the parameters λ, ρ, γ and β, using published macroeconomic forecasts. In the case of λ, we also make use of restrictions stemming from economic theory. The daily financial market data used here are provided by Bloomberg.

For the United States, we start by setting β so that it equates the standard deviations of the two sides of equation (5). We use the right-hand side of equation (4) as the empirical counterpart of $n_{LR|t}$. The series for the long-term interest rate $i_{LR,t}$ is calculated by solving equation (3) for it with $j = 10$, $i_{10,t}$ being the 10-year government bond yield and $i_{1,t}$ the 12-month money market interest rate.[5] Here, for $\pi_{LR|t}$ and $g_{LR|t}$ we use the inflation and real GDP growth forecasts made by the Congressional Budget Office (CBO). We use the forecasts for three to five years ahead, so as not to include the forecasts for the first two years. The standard deviations of the two sides of equation (5) are then calculated for the period 1991-2002, as the data on equity index futures start in 1991. For $n_{LR|t}$ we use end-of-year values, since the CBO publishes its forecasts close to the end of the year. The resulting β equals 1.98.

Once the value of β is set, we obtain the rest of the parameter values for the United States by estimating equations (9) and (10) as a system with parameter restrictions. As left-hand side variables, we use CBO long-run forecasts. As such, the statistical model of the forecasts is of no interest to us. The purpose of the estimation exercise, loosely speaking, is to set the values for the parameters so that equations (9) and (10) produce expectation series with averages and variances in line with the published forecasts that are used as benchmarks here. As the empirical counterparts of $\pi_{LR|t}$ and $g_{LR|t}$ we again use CBO long-run forecasts. The data set extends from 1991 to 2002 in annual frequency, and we again use end-of-year values for the financial market variables.

We estimate the system using the method of maximum likelihood, with the assumption that the error terms are normally distributed, assuming the parameter restrictions given below. The estimation is performed numerically. The parameter restrictions include having the same values for λ, ρ and γ in both equations. In addition, we restrict the value of λ to greater than or equal to 0.5. This restriction for the elasticity of intertemporal substitution is based on macroeconomic literature. Unrestricted, the estimate of λ would be lower and thus inconsistent with the theoretical starting point of the framework. The restriction for λ turns out to be binding, and the estimate of λ is thus 0.5. The estimate of ρ equals 0.022, and the estimate of γ is 0.0375. The fitted values of the equations are depicted in Figures 3 and 4.

For the euro area, data on futures prices have been available only since the beginning of 1999. The value of β is set as for the United States, now based on the forecasts made by Consensus Economics Inc. We use the forecasts for three to seven years ahead, so as not to include the forecasts for the first two years. The forecasts for euro area averages are approximated by weighted averages of the five largest euro area economies. The inflation forecasts are calculated from the real and nominal GDP growth forecasts. The long-run consensus forecasts are published in August each year. Therefore, we use end-of-July values for $n_{LR|t}$. The resulting β parameter for the euro area is 4.98.

With the data series for the euro area being very short, we do not estimate equations (9) and (10) for the euro area. Rather, we set λ at 0.5, following the United States value. The values for ρ and γ are then calculated using equations (9) and (10), setting the variables in these two equations at their average values for 1999-2002. For $g_{LR|t}$ and $\pi_{LR|t}$ we again use consensus forecasts. The resulting values for ρ and γ are, respectively, 0.0214 and 0.103.

3. Results: long-run expectations

The long-run expectations given by the framework are presented in Figures 5 to 13. That is, these figures show the $g_{LR|t}$ and $\pi_{LR|t}$ series given by equations (9) and (10), using the parameter values

[5] Here, we make the approximation that $\omega = 0$.

described in the previous section. The series are derived from daily data on equity indices, dividends, interest rate and equity index futures. The data are provided by Bloomberg. The last data point in these figures is 7 November 2003.

Figures 5 and 7 present the expected long-run inflation and GDP growth rates for the United States. According to the results, the markets' long-run inflation expectations have been on a declining trend since 1991. This is not surprising, since in the 1990s the inflation rate slowed considerably in the United States, as seen in Figure 7. The credibility of US monetary policy regarding price stability seems to have increased since the early 1990s. Figure 8 shows that inflation expectations evolve to some extent similarly to long-term inflation forecasts and to a break-even inflation rate derived from an inflation-indexed bond. Recently, however, the level of inflation expectations has been very low. In terms of the framework, this is mainly due to the very low long-term interest rates.

While inflation expectations were lowered in the 1990s, the expected long-run GDP growth rate increased over the same period, according to the results. This is in line with the accelerated productivity growth seen in the United States, and with upward revisions in growth forecasts, such as those presented in Figure 6. The strong upward movement in expected real GDP growth implied by the framework largely reflects the increases in stock prices seen in the late 1990s. Similarly, the recent fall in stock prices implies a fall in the growth expectation series. The turning point towards lower growth expectations is earlier than the corresponding turning point in the forecasts shown in Figure 6.

The results for the euro area are presented in Figures 9 to 12. These results cover only the period since 1999, because the data for the equity index futures are unavailable before that. Even during this period, there are some gaps in the data, which can be seen in the figures. Similarly to US expectations, euro area growth and inflation expectations have diminished somewhat since the year 2000, and the turn for the worse in the growth expectation series takes place earlier than in the published growth forecasts (Figure 10).

In order to consider longer time series for the euro area, we investigate a version of the framework that differs from the one presented in the preceding sections. Figure 13 depicts growth and inflation expectations derived under the assumption that growth and inflation rates are expected to be constant in the future. That is, the series shown in Figure 13 are derived making the assumptions that $n_{1,t+1|t} = n_{LR|t}$ in equation (2), and $i_{1,t} = i_{LR,t}$ in equation (3). Since the near-term expectations are not treated explicitly in this version of the framework, data on equity index futures are not required. Therefore, the results can be shown for a longer time span than in the case presented in Figures 9 to 12. According to Figure 13, changes in growth and inflation expectations in the euro area since the early 1990s have been to some extent similar to those in the United States. However, changes in growth expectations have been somewhat smaller in the euro area, and the drop in inflation expectations has been greater.

4. Conclusions

This study presents a framework for measuring financial markets' expectations concerning long-run real GDP growth and inflation. The framework is based on economic theory, and uses as inputs data on equity indices, dividends, interest rates and equity index futures prices. Using the framework, market expectations can be measured in real time.

Obviously, it is impossible to determine the "true" market expectations. First, there is no unique set of market expectations, in the sense that the expectations of individual market participants differ. What is being measured in all attempts to gauge market expectations is a sort of noisy weighted average of the individual market participants' expectations. Second, that weighted average can only be observed with limited accuracy: we do not know which of the different measures presented in Figures 6, 8, 10 and 12 are closest to the "truth".

To some extent, the measures of market expectations produced by the framework presented here are similar to some other measures and published forecasts. However, in several instances this is not the case. For example, our measures of US growth expectations differ from other measures during the early part of the 1990s and again the period 2002-03, as shown in Figure 6. During the latter period the same is true for US inflation expectations, as shown in Figure 8.

Regarding recent developments, one can speculate whether the results of this framework imply that the growth forecasts shown in Figure 6 will be revised down in the near future, following the downturn in the expectation series. After all, the level of stock prices compared with past dividends has settled at a level considerably lower than that which prevailed in the late 1990s. Generally, the measures of growth expectations presented here have been strongly influenced by changes in stock prices.

The recent low inflation expectations shown in Figure 8, in turn, reflect low long-term nominal interest rates. This suggests that the fall in interest rates during recent years has been large even relative to the fall in stock prices. This interpretation is based on the fact that our measure of inflation expectations is affected by both nominal interest rates and stock prices as shown in equation (10): stock prices have not fallen enough to counteract the effect of the fall in interest rates. If this result is taken seriously, then one of the following must be true: (1) the market currently expects long-term inflation to be lower than the published forecasts indicate; (2) stock prices are still too high compared with expectations concerning the macroeconomy; (3) inflation expectations are not, for some reason, fully priced into long-term interest rates.

The framework presented here is new and experimental. In addition to the discussion above, one interpretation of the differences between the measures of expectations is that the framework is flawed in one way or another. Naturally, one can identify some potential problems with the approach. One is that equation (5), presenting the relationship between the expected GDP and dividend growth rates, may not hold empirically. It is difficult to assess how severely this equation may be misspecified.

In addition, it is possible that some other parameters of the model framework are not stable. For example, one might think that the assumption of a constant equity premium does not hold, even though this assumption is often made in applied work. In this respect, one way to try to improve the framework presented here would be to consider modelling the variation in the equity premium. Further, it is possible that the relationship between the long-term real interest rate and expected long-run real GDP growth is not stable. Finally, international linkages in the bond and equity markets have not been taken into account in the framework. For example, the real interest rates in the United States and in the euro area undoubtedly also reflect developments in other parts of the world. Dealing explicitly with such international linkages would be another way to possibly improve this framework in future work.

While the framework presented here provides measures of market expectations, it does not attempt to determine whether the expectations later turn out to be correct or not. Therefore, we do not need to take a stand when it comes to the question of whether there are bubbles in financial markets. We simply interpret market prices to reflect market expectations. However, if bubbles exist, they can be problematic for the method we use. This is because we assume that similar growth and inflation expectations are reflected in both stock prices and bond prices. When talking about bubbles, macroeconomists often have stock prices in mind more than interest rates. After all, there is strong evidence that the value of the S&P 500 Index, for example, has tended to vary too much with respect to the subsequent changes in dividends. Large swings in the dividend/price ratio have been followed by large movements in stock prices and not in dividends, as documented by Campbell and Shiller (2001), among others. In addition to stock prices, however, it is naturally possible to argue that there are bubbles in bond prices as well. For example, some economists explained the very low level of long-term interest rates in mid-2003 in terms of a bond market bubble.

Figure 1

S&P 500: short-run growth in nominal dividends

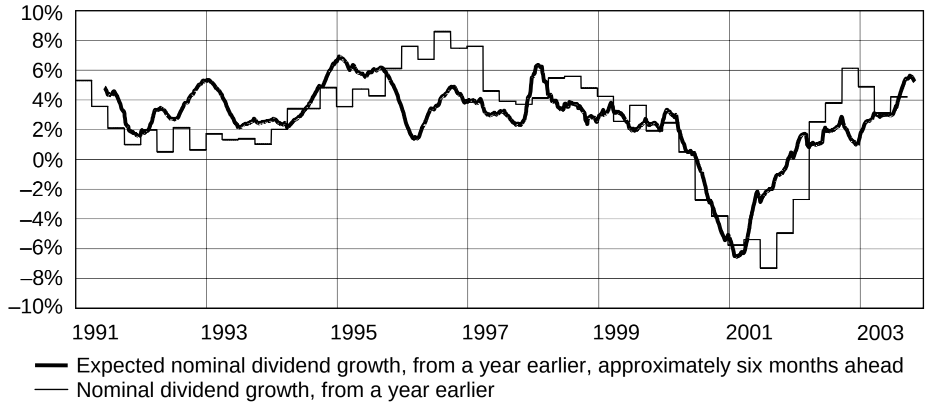


Figure 2

DJ EURO STOXX 50: short-run growth in nominal dividends

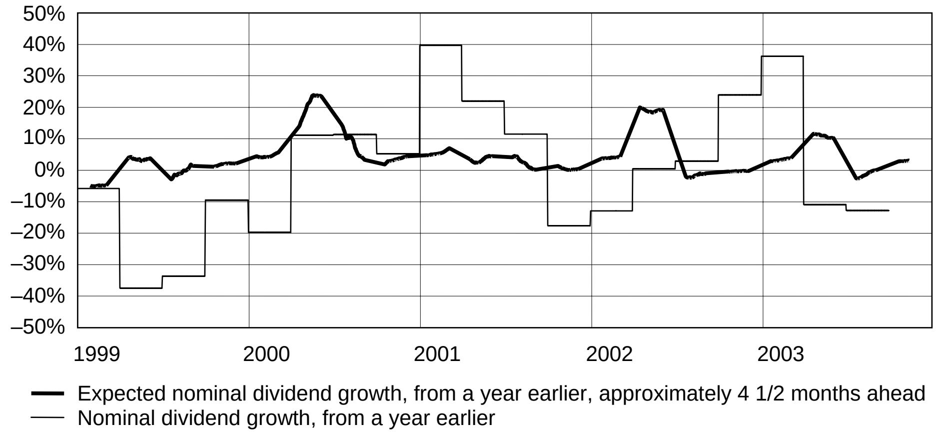


Figure 3

United States: long-run growth forecast (CBO) and fit

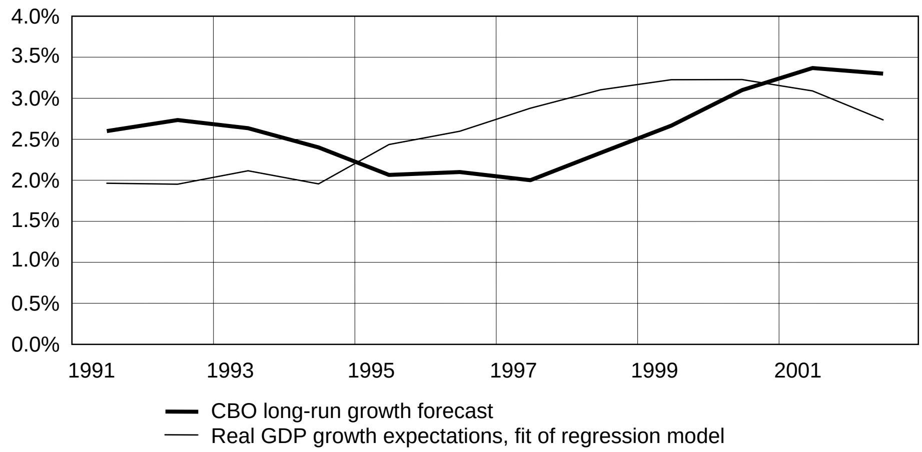


Figure 4

United States: long-run inflation forecast (CBO) and fit

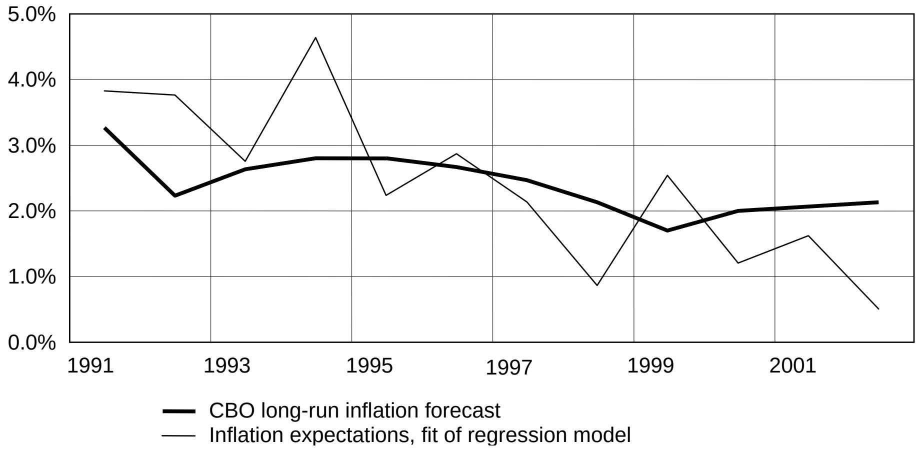


Figure 5

United States: expected long-run growth

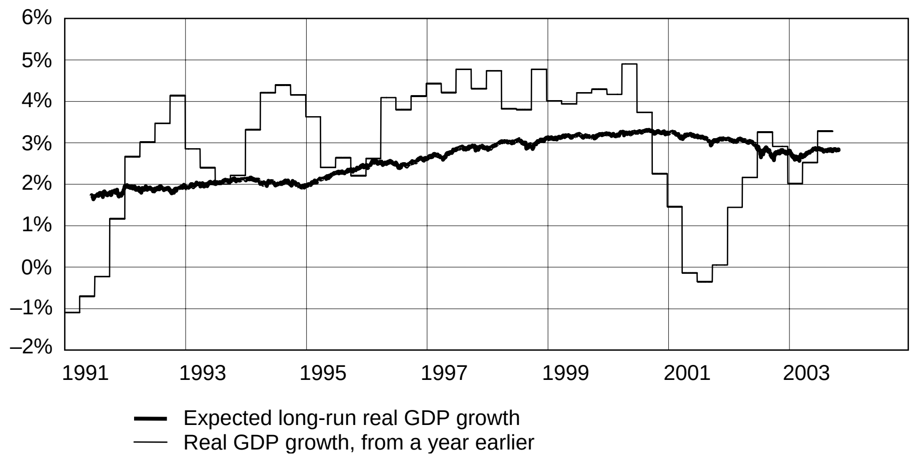


Figure 6

United States: expected long-run growth

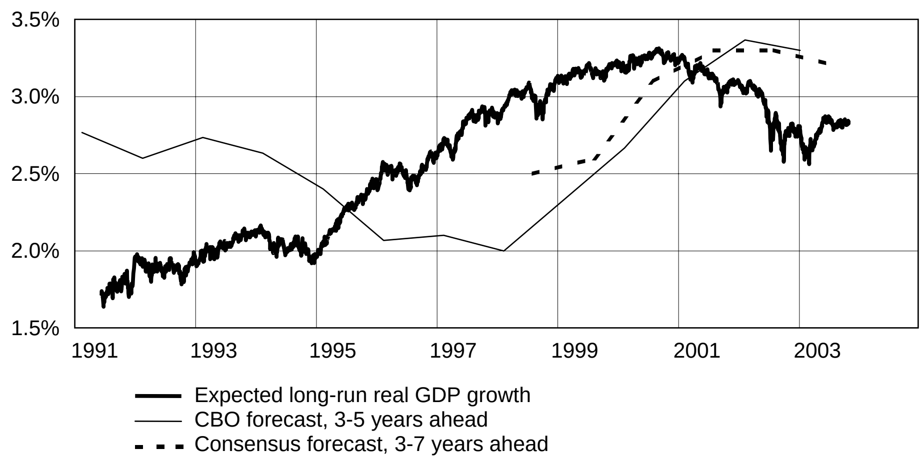


Figure 7

United States: expected long-run inflation

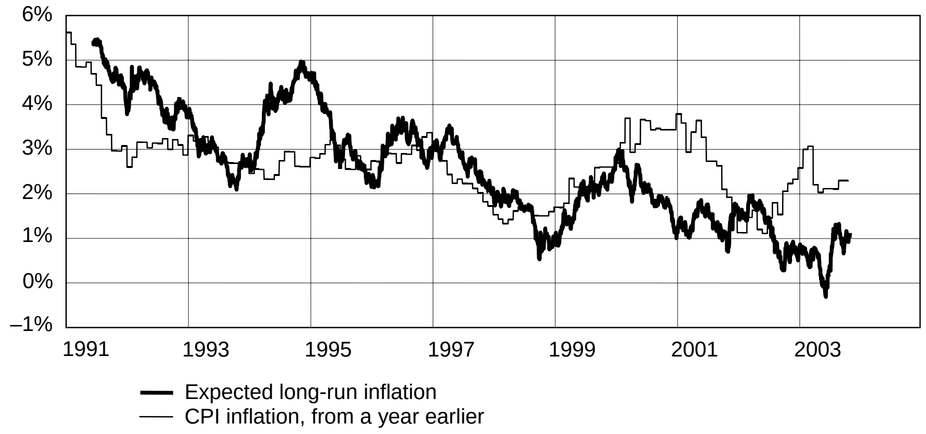


Figure 8

United States: expected long-run inflation

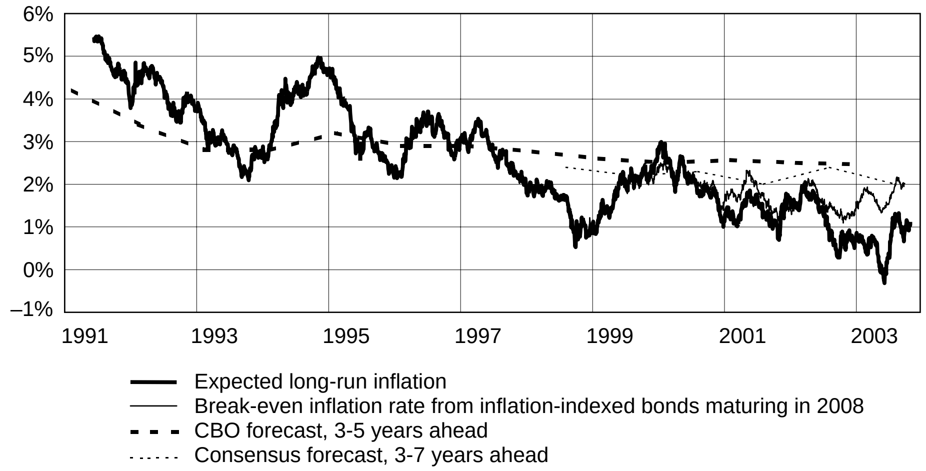


Figure 9

Euro area: expected long-run growth

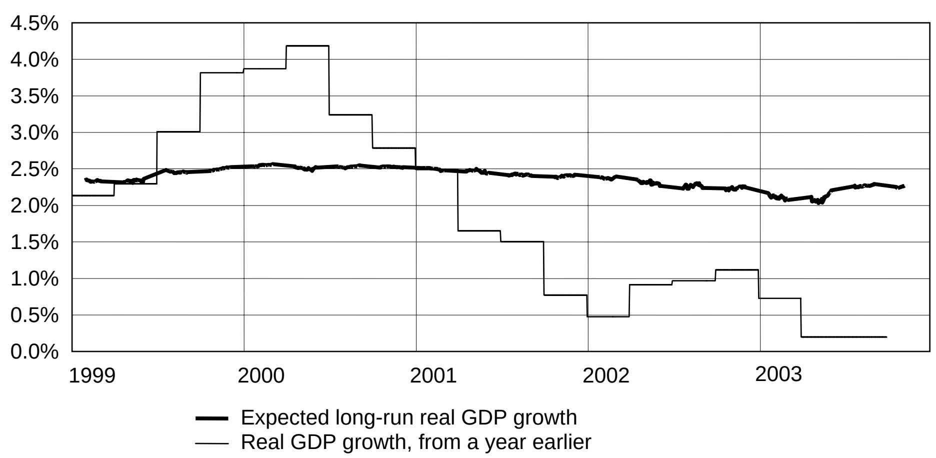


Figure 10

Euro area: expected long-run growth

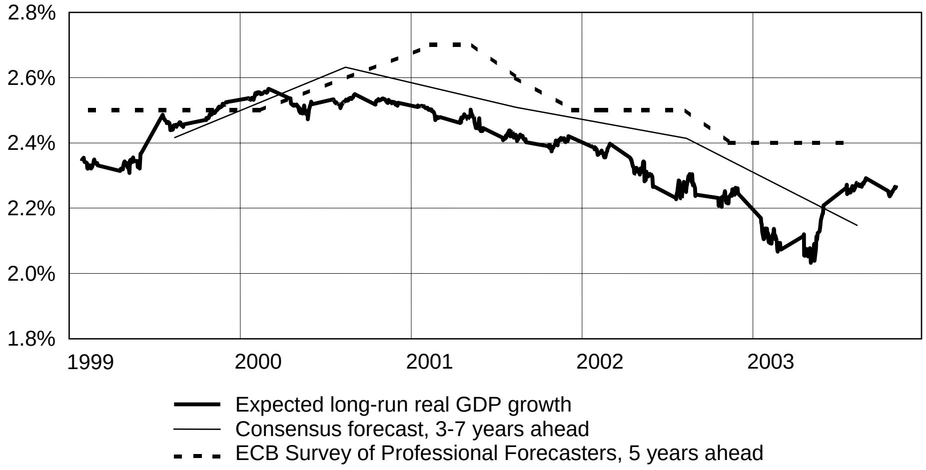


Figure 11

Euro area: expected long-run inflation

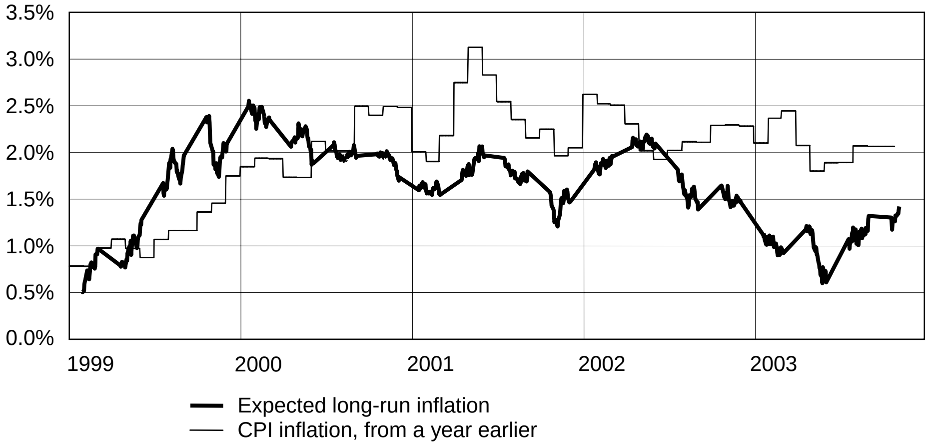


Figure 12

Euro area: expected long-run inflation

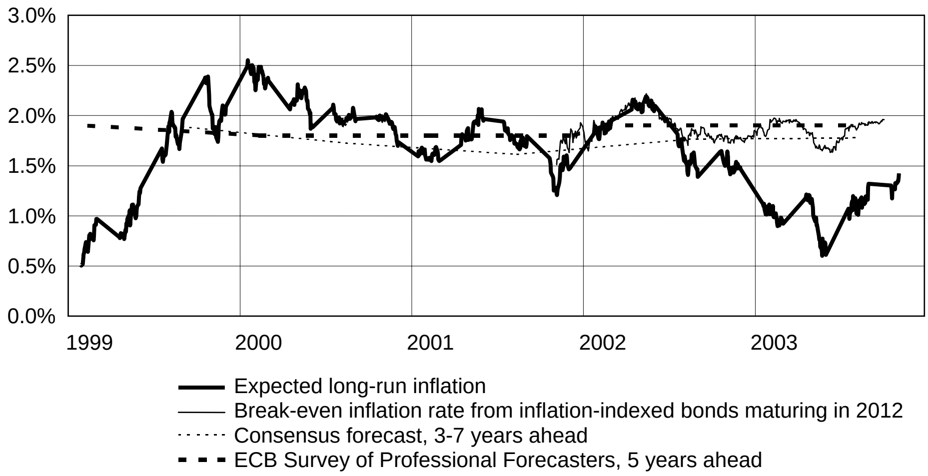
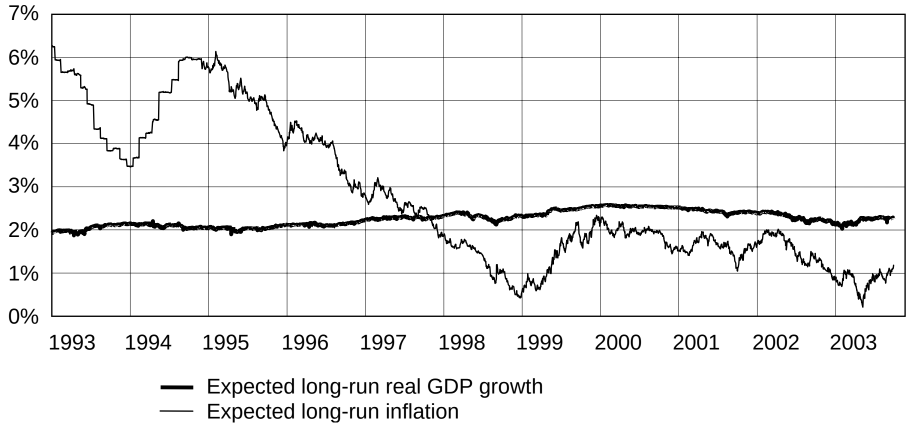


Figure 13

Euro area: long-run expectations, assuming constant expected rates of growth and inflation



References

Campbell, John Y, Robert Shiller (2001): "Valuation ratios and the long-run stock market outlook: an update", *NBER Working Paper*, no 8221.

European Central Bank (2003): "Deriving euro area inflation expectations from inflation-linked swaps", *ECB Monthly Bulletin*, September, pp 31-2.

Martin, Ben and Michael Sawicki (2003): "Market-based estimates of expected future UK output growth", *Bank of England Quarterly Bulletin*, Spring, pp 20-6.

Sack, Brian (2000): "Deriving inflation expectations from nominal and inflation-linked treasury yields", *Journal of Fixed Income*, vol 10, pp 6-17.

Scholtes, Cedric (2002): "On market-based measures of inflation expectations", *Bank of England Quarterly Bulletin*, Spring, pp 67-77.

Forecasting aggregate investment in the euro area: do indicators of financial conditions help?

Marie Diron, Maria Cruz Manzano and
Thomas Westermann, European Central Bank[1]

1. Introduction

The past few years have seen a resurgence of interest in the role that financial conditions play in corporate investment decisions, stemming essentially from the presumption that the current economic cycle is partly shaped by developments in asset prices and gearing. More specifically, in the second half of the 1990s both equity valuations and corporate indebtedness rose sharply to unprecedented levels. The subsequent bursting of the stock market bubble and the protracted slowdown in demand might have led to higher cyclical sensitivity of companies' investment expenditure if companies had had to adjust more rapidly in order to meet debt obligations and adjust their balance sheets. As pointed out by Jaeger (2003), this has important implications for forecasters and policymakers. Indeed, the investment outlook in recent forecasts and projections from international (and private) organisations mostly incorporated some dampening effect from corporate balance sheet adjustments.[2]

There are strong theoretical considerations for taking into account balance sheet effects when assessing corporate investment. Modern finance theory suggests that informational asymmetries can introduce a wedge between (lower) internal and (higher) external costs of finance. If large enough, such a wedge implies that investment projects may have positive net present values but may nevertheless not go ahead or be delayed if there is a lack of internal funds. Adverse financial conditions can also take the form of outright quantity constraints, implying that firms cannot raise external funds at any given cost. In general, constrained firms are likely to be those with relatively small amounts of liquid assets and net worth, where the latter implies lower values of debt collateral. In examining financial constraints in investment, most of the empirical literature has focused on microeconomic data, given that cost and quantity constraints are likely to be related to firm-specific characteristics and that aggregation can blur the identification of important parts of firms' behaviour.

By contrast, forecasts of capital investment are typically undertaken in the context of macroeconomic models with no explicit role for financial constraints. Indeed, the aggregate investment equations in macroeconomic models are typically of a "demand accelerator" or "Q" type and do not normally allow for an impact from financial conditions on investment, other than through cost of capital terms or Q-ratios. At the same time, Bond and Meghir (1994) argue that empirical findings in such equations of investment-profit sensitivities might not reflect financial constraints but simply pick up investment opportunities that are not properly captured by (expected) demand variables and the available proxies for the Q-ratio. Similar problems may exist with regard to other indicators of financial conditions such as share prices. Thus, even if the inclusion of financial variables improves the explanatory power of aggregate investment equations, the economic interpretation of this effect could still be ambiguous.

In this paper we assess the predictive power of various financial indicators in parsimonious aggregate investment equations. Abstracting from theoretical underpinnings, we conduct a horserace exercise where the criterion for incremental predictive power of these indicators is a reduction in the root mean square error of out-of-sample forecasts. We use financial indicators that are more or less readily available to forecasters in order to assess whether ad hoc judgment is the best way to take account of financial variables in projections, or whether there could be a role for a more systematic treatment in

---

1 The views expressed in this paper are those of the authors and do not necessarily represent the views of the European Central Bank.

2 See, for instance, the ECB's *Monthly Bulletin* of June 2003, the IMF's *World Economic Outlook* of April 2003, the European Commission's *Spring 2003 Forecast* of April 2003 and the *OECD Economic Outlook* of June 2003.

investment equations. The exercise confirms a number of well known problems in estimating aggregate investment equations, in particular the difficulty of finding a significant and stable relationship between financial developments and investment. This may reflect the fact that the typical linear aggregate investment equations used in macroeconomic models are ill-suited to capture the impact of financial variables, given that financing conditions may be more relevant in downturns than in upturns or may start being binding beyond certain thresholds only.

The structure of this paper is as follows. Section 2 discusses some stylised facts of adjustment processes in the corporate sector's capital and financial accounts. This helps to understand the various options - in addition to adjusting investment - which firms may have in reacting to cost of capital and balance sheet problems. It also helps to identify financial quantity variables that are potentially useful in signalling financial constraints on investment. Section 3 examines the statistical significance of financial variables in investment equations and their ability to improve the out-of-sample forecasts. The finding is that improvements in forecast errors - if any - are quantitatively limited. One possible explanation for this is that investment and financial indicators do not have the linear relationship assumed in conventional equations. We test this possibility in terms of regime dependency, but only in very few cases find the estimated sensitivity of investment to financial indicators to be significantly different between regimes. The apparent lack of statistical significance could reflect the failure of those financial variables that are readily available to forecasters to accurately capture the nature and extent of financial constraints. Section 4 concludes.

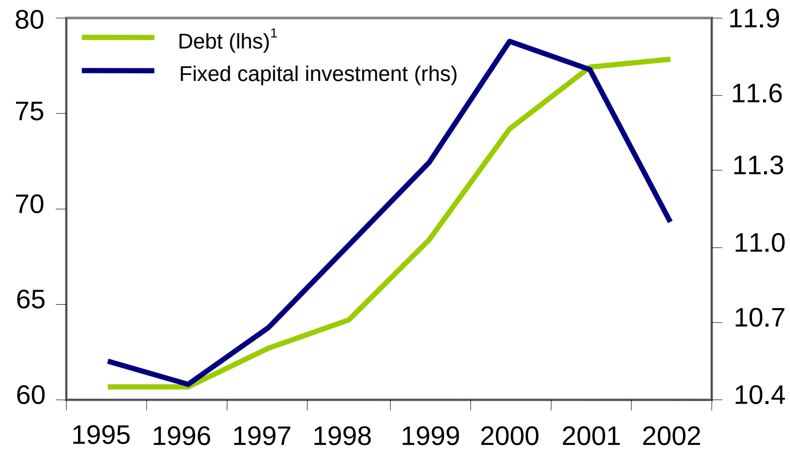
2. Stylised facts of balance sheet adjustments in the corporate sector

This section introduces a general flow of funds framework for analysing balance sheet adjustments in the non-financial corporate sector. The framework is used to review the buoyant investment developments in the second half of the 1990s and their relation to the run-up in corporate debt. As a ratio to GDP, corporate investment increased relatively strongly - by more than 1 percentage point - between 1995 and 2000, and the debt ratio at the same time increased quickly to very high levels of around 75% (Graph 1). Looking at these developments in terms of associated flows helps to assess the adjustments made in the past few years and also gives some indications with regard to the options for further balance sheet corrections in the period ahead.

Graph 1

Investment and debt of euro area non-financial corporations

As a percentage of nominal GDP



[1] Includes loans and debt securities (excluding financial derivatives) issued by and pension fund reserves of non-financial corporations.

Sources: ECB; OECD; Eurostat; authors' own calculations.

The real and financial sides of corporate investment decisions are tied together by a budget constraint. In general terms, outlays for capital investment (*I*) and financial investment (*FI*) are financed by changes in internal funds (*IF*) and external funds, where the latter can take the form of debt (*D*) and/or equity (*E*):

$$I + FI = \Delta IF + \Delta D + \Delta E \qquad (1)$$

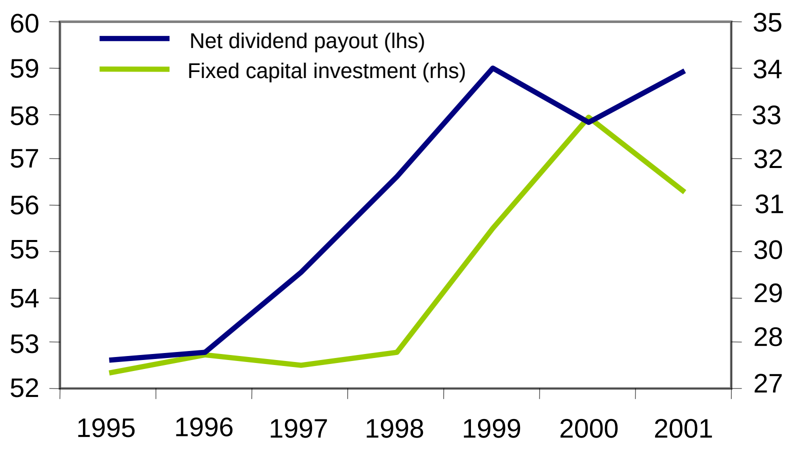
Conversely, the identity implies that in order to reduce debt, businesses have to cut back on investment, generate more internal funds or issue new shares. For tax-paying corporations, the flow of internal funds available for investment essentially reflects profits after subtracting taxes, interest payments and dividend payouts. In addition, the national accounts identify a number of other positions that can affect changes in internal funds, such as net transfers, net acquisitions of non-financial, non-produced assets, or net property incomes from rents and reinvested earnings of foreign direct investment. However, these other positions are relatively small and amount on balance to only 2-3% of the gross operating surplus in the euro area corporate sector. Moreover, due to their nature they are unlikely to play an important active role in businesses' balance sheet adjustment considerations. As official euro area-wide national accounts data for institutional sectors are not yet available, we constructed own estimates for the non-financial corporate sector in the period 1995 to 2001. The estimates are based on OECD data for the individual countries and complement the information from the ECB's monetary and financial accounts available for the period 1995 to 2002.

The pecking order theory of finance establishes a general preference for internal over external funds, and, with regard to the latter, for debt over equity as firms issue the safest security first (Myers (2001)). Looking first at the developments in internal funds, towards the end of the 1990s an increasing part was absorbed by the upturn in corporate spending on capital investment. In 2000, the ratio of fixed capital investment to gross operating surplus peaked at around 58%. Funds were also increasingly absorbed by net dividend payouts, which amounted to around one third of the gross operating surplus at the end of the 1990s (Graph 2). In addition, taxes paid on profits and wealth saw a relatively strong increase to around 10% of gross operating surplus. By contrast, relatively low interest payments took some of the strain off the internally available funds, falling to around 12% of gross operating surplus at the end of the 1990s (Graph 3). Taken together, however, these expenditures exceeded the available internal funds by an increasing margin, reflected in higher net borrowing requirements. This became particularly apparent when in 2000 corporate accounts, mainly in the telecommunications sector, were burdened down by the purchase of UMTS licences.

Graph 2

Investment and dividends of euro area non-financial corporations

As a percentage of gross operating surplus

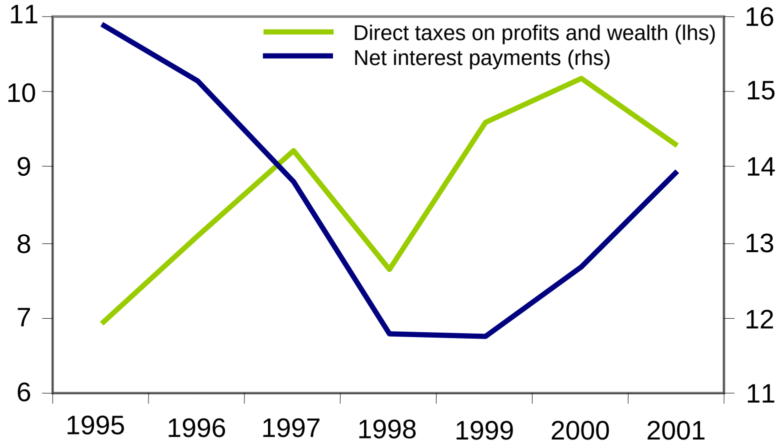


Sources: OECD; authors' own calculations.

Graph 3

Taxes and interest payments of euro area non-financial corporations

As a percentage of gross operating surplus



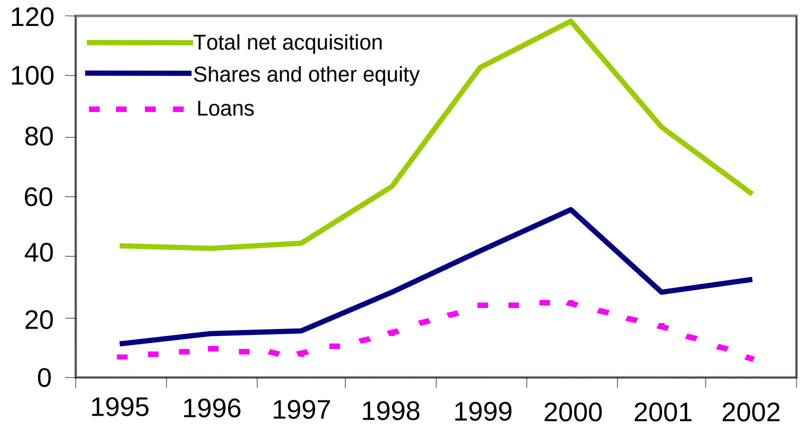
Sources: OECD; authors' own calculations.

The late 1990s were also a period of relatively buoyant financial investment activity. This activity to some extent reflected portfolio investments in a period where stock market prices kept climbing to unprecedented levels. In addition, there was a strong pickup in mergers and acquisitions (M&A) activity, explained by a combination of structural and cyclical factors which fostered, mainly in some sectors like high-tech and telecommunications, the expansion and the scale of the activity of euro area firms domestically and abroad. Overall, net financial investment increased much more strongly than fixed capital investment and in 2000 clearly exceeded the latter while in 1995 it had been less than half of it. Equity investment alone amounted to almost 60% of fixed capital formation in 2000 and intercompany loans accounted for another 20% (Graph 4).

Graph 4

Net financial investment of euro area non-financial corporations

As a percentage of fixed capital formation

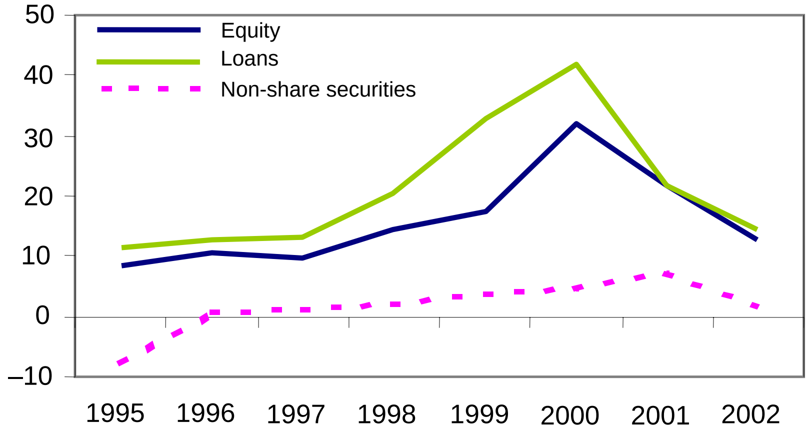


Sources: ECB; authors' own calculations.

Graph 5

Net incurrence of liabilities of euro area non-financial corporations

As a percentage of gross operating surplus[1]



[1] Operating surplus for 2002 estimated from economy-wide data.

Sources: ECB; OECD; authors' own calculations.

The sum of capital and financial investment implied a widening financing gap vis-à-vis the available internal funds and showed in a strongly rising incurrence of liabilities. In 2000 this almost reached the volume of corporate profits, with loans being the largest component of gross operating surplus at around 40% (Graph 5). While over the second half of the 1990s overall debt financing (loans plus debt securities issued) gained relative importance vis-à-vis the issuance of shares and other equity, the latter was particularly strong in 2000 at the height of the stock market boom. Given the buoyant stock price developments until early 2000, some conventional leverage indicators (eg debt in relation to financial or total assets) did not immediately reflect the rising indebtedness of euro area corporations, while others, such as ratios of debt to operating surplus or to GDP, started to reflect it earlier. The strong and protracted fall in stock prices from 2000 onwards not only had repercussions on firms' leverage ratios but in an environment of relatively low interest rates also significantly increased the cost of equity in relation to that of debt. As a consequence, financing via quoted shares was cut back and the relative importance of debt issuance rose again in 2001. In particular, the issuance of debt securities continued to rise relatively strongly right into the early phases of the downturn, reflecting in part the fact that some of the earlier M&A activities were financed through short-term bridge loans which were later substituted by the issuance of debt securities.

The more moderate recourse to external funds that took place after 2000 reflects the lower demand for finance associated with the economic slowdown and the stock market decline but also the return to more normal levels after the one-off boost related to the purchase of UMTS licences. In addition, supply factors could also have played a role if the high level of indebtedness had signalled risks to financial market participants and given rise to more cautious lending policies by banks. Such supply side considerations could have affected the availability of new funds for firms (mainly in the case of the most heavily indebted firms) and/or the risk premia incorporated in their cost. Since 2000, euro area non-financial corporations seem to have been under pressure to improve their financial structure and rationalise investments they have carried out in the past. In some cases (such as telecoms), this involved not only debt restructuring but also business reorganisation, including asset sales in order to generate internal financing resources, despite lower market values.

The adjustment process towards lower financing gaps also involved lower capital investment, while dividend payouts seem to have remained more resilient as a ratio to the gross operating surplus. The role of dividends in the impact of balance sheet adjustment on investment depends on the ranking of business and shareholder objectives. For some corporations, continuity of dividend payments may be on a par with investment and consolidation, given that dividend payout policies can have important signalling effects for financial markets and shareholders. However, with stock prices being low, share repurchases could be an alternative use of available funds in providing positive signals to financial markets. The debt service burden remained subdued in 2001 and 2002 despite the high level of indebtedness, but, given that profit developments have also remained weak, interest payments took up a slightly rising share in gross operating surplus. By early 2003, the efforts made to generate more internal funds and deleverage balance sheets had not yet translated into visible improvements in debt ratios. Looking forward, more adjustment might thus be needed, but this may be easier once the recovery is fully under way and allows for some "growing-out" effect in terms of higher profits.

The analysis above points to a number of financial variables that interact with fixed capital investment in balance sheet adjustment processes. Forecasting investment in the presence of potential financial constraints would thus ideally consider all the accounting identities implied by the flow of funds. However, in practice, the data set of timely financial variables that is normally available to forecasters tends to be limited and to consist of prices rather than quantities. Moreover, feedback loops between the financial sector and the real economy are typically not taken into account. Forecasters are therefore typically obliged to inform their judgment on the basis of cruder tools. This issue is addressed below.

3. Including financial indicators in investment equations - some empirical results

3.1 Preliminary steps

In this section we establish a benchmark investment equation, which we then use in out-of-sample forecast exercises to examine the statistical relevance of financial variables. The ECB's forecast models are based on quarterly data. A breakdown of quarterly euro area-wide investment by main types of products has recently become available, but a breakdown according to institutional sectors is not available as yet. For the purpose of this paper, it was therefore necessary to choose an investment series on the basis of the available breakdown that is as close as possible to corporate investment. Two measures were considered: non-housing investment and non-construction investment, which, respectively, account for around three quarters and half of total euro area investment. Excluding all construction investment has the drawback of not taking into account the increasing share of buildings and office space in corporate investment as the services sector gains in importance. On the other hand, using non-housing investment implies the drawback of including public infrastructure investment, which does not follow the same determinants as business investment. As this was perceived to be a lesser problem, the focus below is on non-housing investment. This implies looking at investment activities that reflect - to around three quarters - decisions made in the corporate sector (Graph 6).

The analysis presented is carried out with the aim of drawing possible practical conclusions for forecasters. In this respect, we "let the data speak" as much as possible. In particular, we remain agnostic in terms of which measure (growth rates, ratios, etc) to use for the various financial indicators and about the leads and lags involved in their relationship with investment.

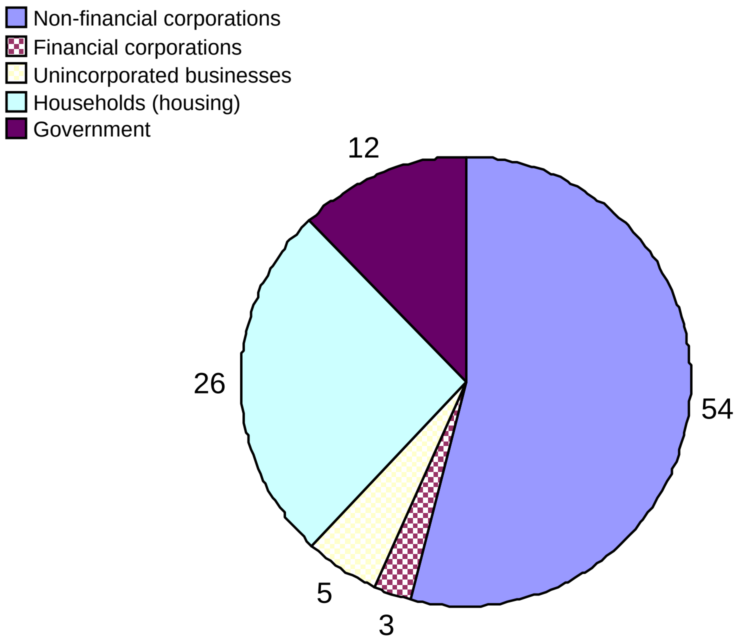
Correlation analysis

As a first step, we compute cross-correlation coefficients in order to obtain some initial indication of which indicators are likely to be useful in explaining developments in the investment ratio. Correlations can also point to a specific measure for a given indicator and specific leads or lags at which it may be relevant. Table 1 shows average correlation coefficients between a series of variables and quarter-on-quarter differences in the ratio of real non-housing investment to GDP. The range of indicators attempts to capture demand conditions and expectations of economic activity as well as financing conditions. Section 2 provided some guidance as to which financial variables would be useful to include, but most of these indicators are financial quantity variables that are not part of the data set used in the ECB's macroeconomic projections. This reflects the fact that for the euro area as a whole these data mostly cover only a very short time period, which makes it difficult to derive reliable empirical evidence on their relevance in structural equations underlying macroeconomic models.

Graph 6

Investment by institutional sector in 2001

Values, as a percentage of total economy investment



Sources: OECD; author's own calculations.

For the purpose of this paper, the choice of financial indicators was therefore guided, first and foremost, by data availability for longer time horizons and, second, by the availability of proxy forecasts or exogenous assumptions for the future developments of these variables in forecasting exercises. As far as possible, both price and quantity aspects of financing conditions are included in the set of financial indicators, although data are more readily available for prices than quantities. Details on data sources and definitions are provided in the annexes. Various measures are tested for each variable, such as quarter-on-quarter growth rates or ratios to gross operating surplus. For some volatile variables, such as share prices, a smoothed growth rate (taking a two-quarter moving average) is also tested. The shaded cells denote the highest correlation coefficients (including those close to, ie an arbitrary ±0.03 from, the maximum) for each indicator and measure.

The main features emerging from this analysis are the following.

As regards *demand variables*, developments in GDP and final demand are strongly correlated with those in the investment ratio, while the correlation between euro area foreign demand and investment is not significant. Similarly, the correlation between the growth rate of GDP excluding investment and the investment/GDP ratio is rather low. The latter observation probably reflects the fact that investment is determined by specific factors which may not affect other expenditure components, and that there exist spillover effects within different investment categories that are missed when investment is excluded from the demand indicator. *Capacity utilisation* seems to be lagging investment, when considered in level terms, while its changes are coincident or leading. The drawback of this indicator is that it refers to the manufacturing sector only, while the share of corporate investment accounted for by services sector companies is likely to have increased in recent years, to significant levels.

The various *financial indicators* show similar results, with most of them apparently being coincident at correlation coefficients of 0.3-0.5. The three measures of financing costs considered here (long-term interest rates, cost of equity, and the composite cost of financing measure) show the expected negative correlation with investment. Over the common sample of available data for the three cost measures, the cost of equity shows the strongest link with investment. This may reflect the fact that developments in share prices which underpin this variable are linked to corporate investment not only

Table 1

Correlation with change in non-housing investment/GDP ratio

Quarters (q) lead or lag	Measure	Lead 4q	Lead 3q	Lead 2q	Lead 1q	Coincident	Lag 1q	Lag 2q	Lag 3q	Lag 4q
GDP	GR	0.21	0.34	0.36	0.42	0.49	0.40	0.39	0.24	0.19
Final demand	GR	0.26	0.42	0.47	0.49	0.57	0.47	0.38	0.21	0.07
GDP excluding non-housing investment	GR	0.14	0.21	0.21	0.28	0.14	0.26	0.23	0.09	0.08
Foreign demand	GR	0.21	0.03	0.12	0.03	0.05	0.00	–0.01	0.12	0.09
Capacity utilisation rate	L	–0.06	0.06	0.18	0.28	0.40	0.46	0.55	0.57	0.55
	D	0.39	0.40	0.40	0.37	0.38	0.19	0.29	0.09	–0.08
Stock market capitalisation	GR	0.29	0.26	0.28	0.42	0.40	0.21	0.28	0.13	0.05
	GRS	0.32	0.32	0.32	0.42	0.48	0.37	0.29	0.24	0.10
Share price index	GR	0.28	0.23	0.27	0.40	0.37	0.17	0.25	0.09	0.00
	GRS	0.30	0.30	0.29	0.40	0.45	0.33	0.25	0.20	0.05
Price/earnings ratio	L	0.25	0.27	0.31	0.34	0.31	0.29	0.27	0.25	0.21
Dividend yield	L	–0.32	–0.35	–0.38	–0.42	–0.42	–0.40	–0.39	–0.36	–0.33
Dividend/earnings ratio	L	–0.14	–0.20	–0.27	–0.36	–0.47	–0.51	–0.55	–0.53	–0.53
Long-term interest rates (COST)	L	0.05	0.07	0.05	0.06	0.03	0.06	0.04	0.00	–0.02
	D	–0.17	–0.01	–0.04	–0.06	0.01	0.10	–0.01	–0.07	0.02
Cost of equity issuance	L	–0.33	–0.39	–0.54	–0.55	–0.48	–0.48	–0.38	–0.26	–0.17
	D	–0.24	–0.14	–0.31	–0.13	0.05	–0.03	0.10	0.10	0.12
Composite cost of financing	L	–0.33	–0.35	–0.38	–0.41	–0.38	–0.34	–0.33	–0.31	–0.28
	D	–0.32	–0.09	–0.19	–0.17	0.08	0.25	0.14	0.18	0.29
Yield curve	L	0.39	0.38	0.35	0.30	0.28	0.21	0.17	0.06	–0.02
Corporate loans	GR	–0.06	0.12	0.21	0.24	0.29	0.36	0.52	0.41	0.42
	GRS	–0.16	0.04	0.20	0.26	0.32	0.39	0.53	0.57	0.52
	RX	–0.24	–0.25	–0.33	–0.38	–0.42	–0.43	–0.42	–0.38	–0.37
	RXD	–0.21	–0.02	–0.07	–0.13	–0.06	0.09	0.23	0.22	0.27
Gross operating surplus	GR	0.28	0.17	0.31	0.37	0.31	0.16	0.08	0.01	–0.08
	GRS	0.41	0.30	0.32	0.45	0.44	0.32	0.16	0.06	–0.05
Expected earnings	GR	0.24	0.24	0.23	0.32	0.42	0.22	0.29	0.12	0.17
	GRS	0.22	0.31	0.30	0.35	0.48	0.41	0.33	0.26	0.19
Corporate debt	RX	–0.27	–0.28	–0.36	–0.42	–0.46	–0.47	–0.45	–0.41	–0.39
	RXD	–0.26	–0.07	–0.10	–0.17	–0.08	0.09	0.24	0.23	0.27

Note: Sample 1980:1-2003:1, except for cost of equity: 1988:1-2003:1. Financial variables expressed in real terms, except ratios, cost of equity issuance and composite cost of financing deflated (see Annex 1). L refers to levels; D is the quarter-on-quarter difference; GR is the quarter-on-quarter growth rate; GRS refers to the quarter-on-quarter growth rate of the two-quarter moving average level; RX is the ratio to gross operating surplus; and RXD is the quarter-on-quarter difference in this ratio.

via the implied cost of share issuance but also because both variables are influenced by expectations of future economic activity. As regards variables capturing the availability of internal and external funds, correlations of 0.4-0.5 are found between investment, on the one hand, and loans or profits, on the other. The ratios of loans and debt to operating surplus capture developments in the leverage of the corporate sector. These variables show a negative correlation with investment, which is consistent with the idea that a worsening in balance sheet conditions may act as a constraint on investment expenditure.

Benchmark equation

As a second step, we derive a benchmark equation for investment against which we can subsequently analyse the possible impact of financial variables. Quarter-on-quarter differences in the ratio of non-housing investment to GDP (NHIR) are regressed on real GDP growth and COST, the real long-term interest rate adjusted for the relative decline in non-housing investment good prices.[3] Although relatively standard, this equation differs from the investment equations which are included in some macroeconomic models such as the ECB's area-wide model (Fagan et al (2001)). The latter are often derived from production functions where investment growth is explained within an error correction format, with a long-term relationship between the capital stock and real GDP and cost of external finance. However, for the euro area, no data on the capital stock are available and own estimates would have introduced considerable data uncertainty in the estimates.

The lag structure of the equation is determined using PC-GETS,[4] starting with a maximum of four lags for each variable and using instrumental variable estimation in order to account for collinearity. The list of instruments comprises lagged values of the dependent and explanatory variables, as well as euro area exports and the rate of capacity utilisation. The results of IV estimation were very similar to that from OLS estimation. Using PC-GETS has the advantage of "letting the data speak", which seems particularly convenient for the purpose of this paper, considering that there is little a priori knowledge as to the combination and lag structure in which the real economy and financial variables should enter the equation. For instance, GDP growth may account for both current demand conditions and expectations of future activity. Remaining agnostic a priori as regards the lag structure of the equations thus seems a sensible approach. The benchmark equation takes the following form:

$$d(NHIR) = C + \sum_{i=1}^{4} \alpha_i * d(NHIR(-i)) + \sum_{i=0}^{4} \beta_i * d\log(GDP(-i)) + \sum_{i=0}^{4} \gamma_i * d(COST(-i))$$

The estimation results are shown in Table 2. The dummies for the second and third quarters of 1984 were selected by PC-GETS and capture the impact of the strikes in the German industrial sector at the time, related to disputes about the introduction of the 35-hour working week. The results shown in Table 2 imply, upon recalculation, that demand is the main explanatory factor of investment, with an elasticity of around 2.5. This importance is in line with the empirical literature and specifications typically used in macroeconomic forecasting models. Moreover, a 100 basis point increase in nominal interest rates cuts investment by around 50 basis points instantaneously and 80 basis points in the long term. The equation passes the usual residual and stability tests. However, there is some evidence of heteroskedasticity, which may be a sign that some information is missing and/or that the relationship between investment, on the one side, and demand and interest rates, on the other, is non-linear. Moreover, the standard error is of the same order as the average absolute value of the dependent variable and similar to the standard error of a simple autoregressive equation.

---

3 Cointegration analysis within the standard Johansen approach showed no cointegration relationship between investment, GDP and long-term interest rates. This may be due to the fact that the sample is relatively short, with the investment/GDP ratio exhibiting large and protracted swings. Given the absence of any stable long-term relationship, the equation only includes short-term dynamics.

4 PC-GETS is a software designed to implement D Hendry's general-to-specific approach, one of the main elements of the LSE approach to econometrics. This method is particularly suitable when, as in the case at hand, the precise formulation of the equation under analysis is not known a priori.

Table 2

Benchmark equation - estimation results

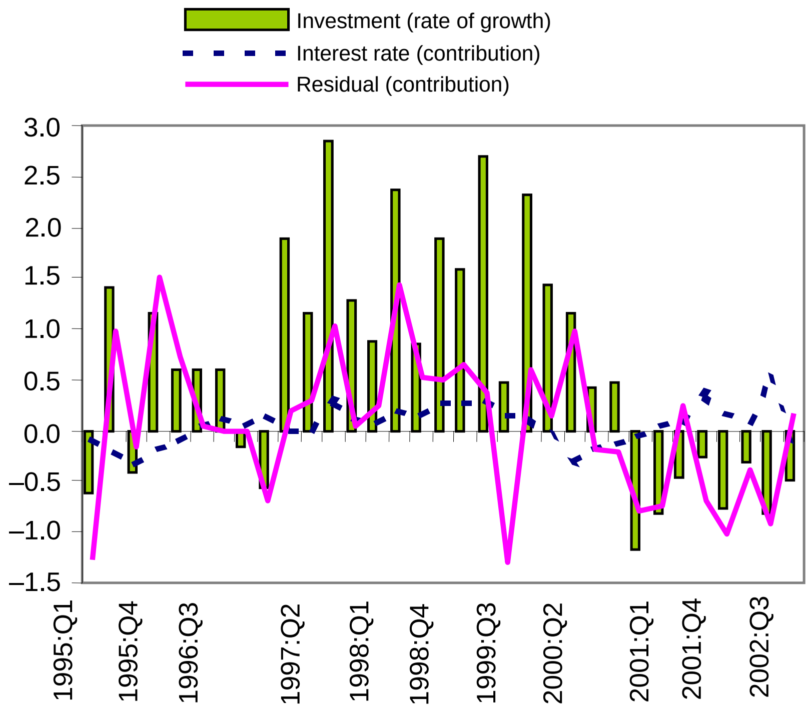
Dependent variable: d(NHIR)
Sample: 1980:1 to 2003:1
White Heteroskedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	Std error	t-statistic	Prob
C	–0.06	0.02	–3.27	0.0016
D(NHIR (–1))	0.25	0.08	3.01	0.0035
D(NHIR (–2))	0.23	0.09	2.55	0.0127
Dlog(GDP)*100	0.12	0.03	3.99	0.0001
D(COST(–4))	–0.25	0.12	–2.21	0.0298
D842	0.36	0.04	9.21	0.0000
D843	–0.43	0.03	–12.81	0.0000
R-squared	0.54	Mean dependent variable		–0.0024
Adjusted R-squared	0.50	S D dependent variable		0.15
S E of regression	0.11	Akaike info criterion		–1.50
Durbin-Watson statistic	2.08	Schwarz criterion		–1.31
F-statistic	15.6	Prob (F-statistic)		0.0000

Graph 7

Contributions of interest rates and unexplained part in benchmark equation

Quarter-on-quarter growth in investment, in per cent and percentage points



Sources: Eurostat; authors' own estimates.

As regards recent developments, compared with the predictions of the benchmark equation, investment was consistently higher in the late 1990s and has been consistently weaker since the end of 2000 (Graph 7). This gives rise to the possibility that other factors have raised and then dampened euro area investment. The remainder of this section looks at whether some of these unexplained developments in investment may be accounted for by financing conditions.

3.2 Linear analysis

Linear estimates of the impact of financial variables

In order to assess the role of financial variables in determining investment, the benchmark equation is augmented by the financial indicators reported in Table 1 (including their various measures such as quarter-on-quarter rates and ratios to gross operating surplus). The variables are included one by one, as taking into account several at the same time was perceived to be too onerous in terms of degrees of freedom. As before, PC-GETS is used to determine the lag structure. The approach admittedly amounts to data mining: the objective is to find significance for a measure or a set of measures for a given financial indicator. At the same time, deciding a priori on a given measure and lag structure is not feasible as most indicators probably capture various channels through which they could affect investment, which could correspond to different measures or lags of the indicators. Table 3 shows the indicators and measures which are significant, together with the estimated lag structure. Most financial indicators are found to be significant, although introducing them in the benchmark equation sometimes implies that the interest rate term is no longer significant.

The forecasting performance of the benchmark and the augmented equations are compared in terms of an out-of-sample forecasting exercise carried out on a rolling basis. More precisely, each equation is estimated up to a particular quarter Q and forecasts are produced for investment for the four following quarters. These forecasts are saved. Then, the equation is estimated up to Q+1, with forecasts again produced for the next four quarters, and so on. The average of root mean square errors (RMSE) for one-, two-, three- and four-quarter-ahead forecasts is shown in Table 4. Three different out-of-sample periods are used: one for forecasts over a six-year period (1997:1 to 2003:1),[5] the two others corresponding to a split of this period between the upturn (1997:1 to 2000:1) and the recent slowdown (2001:1 to 2003:1). In this exercise, financial variables are assumed to be known over the forecast horizons, while, in real forecasting conditions, financial variables also need to be forecast or, more often, derived from technical assumptions. Forecast or assumption errors as regards developments in financial variables would thus tend to worsen the forecasting performance of the augmented equations compared with what is shown in Table 4. GDP and long-term interest rates are also assumed to be known, but as this is the case in both the benchmark and the augmented equations, it should not affect the relative reliability of the forecasts. A further difference compared with real-time forecasting conditions is that currently available series, ie including possible revisions to back data, are used. In the absence of a database of vintages of national accounts data going far enough into the past, the impact of data revisions on the results could not be tested. In this respect, financial variables have the advantage that they are not revised.

Table 4 shows in-sample standard errors and out-of-sample RMSEs for the benchmark equation and the improvement (in bold) or worsening in these measures obtained from the augmented equations. For reference, the results of forecasts of investment based on an autoregressive equation are also reported.

In several cases, taking into account financial variables yields lower RMSEs. However, the improvement is rarely statistically significant,[6] or, when it is, it is relatively small. Graphs 8 and 9 illustrate these results. Graph 9 shows examples of the forecasts produced with the benchmark equation and with two augmented equations: the patterns of these three forecasts are very similar.

---

[5] The choice of 1997:1 as a starting quarter for the out-of-sample exercise is to a large extent arbitrary. It represents a trade-off between leaving enough in-sample data points to have reliable estimates and having a long enough out-of-sample period for the comparison of RMSEs to be meaningful. Moreover, starting in 1997 presents the advantage of having both upturn and downturn phases in the out-of-sample period.

[6] According to a Diebold and Mariano (1995) test, corrected for the small sample bias as advised by Harvey et al (1997).

Taking GDP as known, the forecasts are transformed in terms of quarter-on-quarter investment growth, and Graph 9 shows the part of investment growth which is not accounted for by determinants in the benchmark and some augmented equations. While both graphs show that including financial variables helps capture investment developments somewhat better, a significant part of investment developments remains unexplained. In particular, the estimated impact of financial variables cannot account for the observed large declines in investment of the past two years.

Table 3

Linear estimations with financial indicators

Indicator	Measure	Financial variable		Demand		Interest rates	
		Lag	Coeff	Lag	Coeff	Lag	Coeff
Benchmark				0	0.13	4	–0.25
Stock market capitalisation	GR	–1	0.005	0	0.13	Not significant	
Share price index	GR	–1	0.005	0	0.13	Not significant	
	GRS	0	0.006	0	0.12	Not significant	
Dividend yield	L	0	–0.02	0	0.13	Not significant	
Dividend/earnings ratio	L	0	–0.002	0	0.13	4	–0.24
Cost of equity issuance	L	0	–0.07	0	0.18	Not significant	
Composite cost of financing	L	1	–0.07	0	0.13	Not included[1]	
		3	0.06	3	0.06		
	D	1	–0.12	0	0.12	Not included[1]	
				3	0.05		
Yield curve	L	0	0.03	0	0.14	Not significant	
				3	0.07		
Corporate loans	GR	4	–0.03	0	0.13	4	–0.25
	GRS	0	0.05	0	0.10	4	–0.26
		4	–0.07				
	RX	3	–3.4	0	0.13	4	–0.26
	RXD	4	–59	0	0.14	Not significant	
				3	0.05		
Gross operating surplus	GR	1	0.02	0	0.15	4	–0.26
		2	0.02	3	0.08		
	GRS	4	0.02	0	0.12	Not significant	
Expected earnings	GR	0	0.006	0	0.13	4	
	GRS	0	0.011	0	0.12	Not significant	
Corporate debt	RX	0	–1.43	0	0.10	4	–0.33
		3	1.30	3	0.06		
	RXD	2	–2.3	0	0.14	4	–0.32
				3	0.06		

[1] Interest rates are already included in the composite cost of financing measure. OLS and IV estimations generally give the same results except for the dividend yield, the dividend/earnings ratio, the level of and the difference in the ratio of loans to gross operating surplus, and the quarter-on-quarter difference in the ratio of debt to gross operating surplus.

Table 4

In-sample and out-of-sample performance

		In-sample standard error	Out-of-sample RMSE		
			1997:1-2003:1	**1997:1-2000:4**	**2001:1-2003:1**
PC-GETS benchmark		0.11	0.11	0.12	0.09
AR equation		14.1	17.1	15.4	20.2
Stock market capitalisation	GR	0.1	**–7.5**	**–6.7**	**–9.7**
Share price index	GR	0.2	**–9.3**	**–7.7**	**–13.9**
	GRS	**–2.2**	**–7.0**	**–2.1**	**–21.1**
Dividend yields	L	2.1	**–2.1**	**–8.8**	13.3
Dividend/earnings ratio	L	**–1.0**	4.3	**–9.6**	33.8
Cost of equity issuance	L	10.0	0.8	5.8	**–13.7**
Composite cost of financing	L	**–5.6**	**–8.0**	**–15.2**	8.4
	D	**–8.0**	**–10.9**	**–9.1**	**–15.8**
Yield curve	L	24.4	4.1	3.2	6.2
Corporate loans	GR	**–6.7**	**–0.4**	0.8	–3.6
	GRS	**–10.1**	**–5.9**	**–4.9**	**–8.3**
	RX	0.5	12.4	22.2	**–18.6**
	RXD	**–0.6**	1.2	13.1	**–40.7**
Gross operating surplus	GR	**–3.1**	14.6	12.8	19.3
	GRS	1.1	2.1	5.5	**–7.2**
Expected earnings	GR	0.4	**–7.8**	**–14.2**	7.3
	GRS	**–1.7**	**–12.0**	**–10.7**	**–15.3**
Corporate debt	RX	**–3.3**	7.5	13.7	**–10.4**
	RXD	**–0.1**	7.8	8.4	6.4

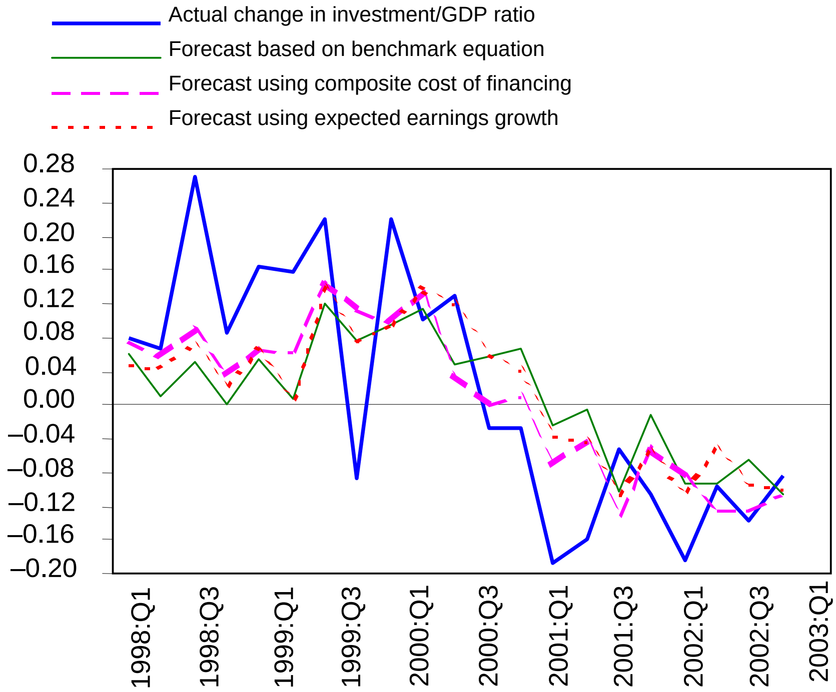
Note: Benchmark: standard error and average of RMSEs for one- to four-quarter-ahead forecasts in percentage points. Other equations: percentage improvement (–) or worsening (+) compared with benchmark.

Several factors may account for the failure to find stronger quantitative evidence of financial indicators in aggregate investment equations. For instance, available indicators may not capture accurately the nature and extent of the financing constraints faced by corporations. Moreover, some sector- or firm-specific factors may not be adequately captured within the macroeconomic framework. Another possibility is that the relationship between investment and financial indicators is non-linear. This latter issue is addressed in the following subsection. From the perspective of projections, resorting to non-linear representations of investment poses significant problems, since including such representations within a macroeconomic model is fraught with difficulties. The idea is therefore to investigate whether non-linear relationships may help understand the relevance of financial variables for investment in the past. This would then guide judgment about the possible effect of financial variables within the projections horizon, while any adjustment would probably have to remain largely ad hoc.

Graph 8

Four-quarter-ahead forecasts of quarter-on-quarter change in investment/GDP ratio

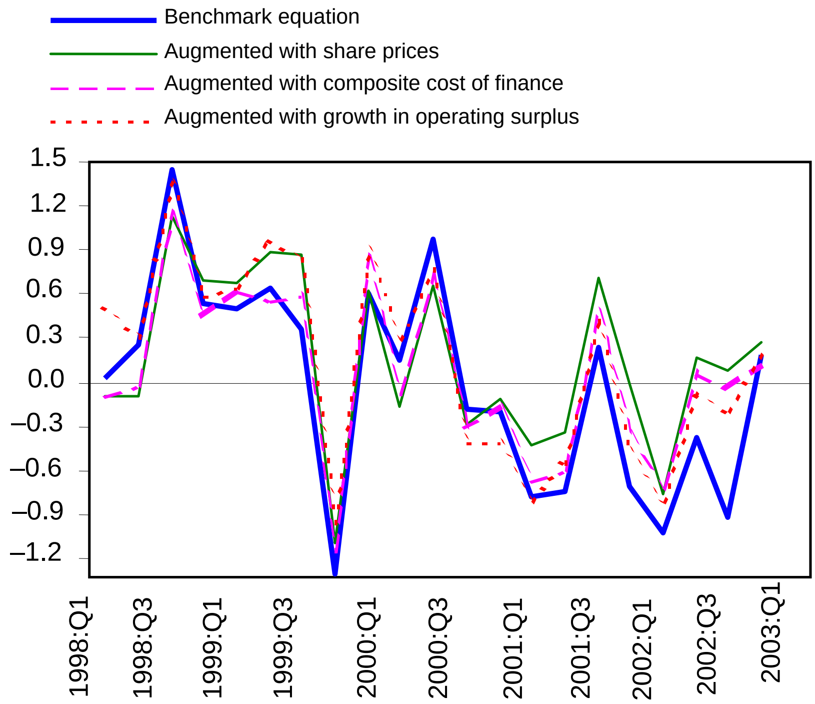
In percentage points



Graph 9

Quarter-on-quarter growth in investment unexplained by determinants from various equations

In percentage points



3.3 Non-linear analysis

Non-linearities in the relationship between investment and financial factors may arise for two reasons. First, financial factors may affect investment decisions differently depending on the stage of the business cycle. A second non-linear aspect relates to different elasticities of investment to the financial variables depending on the state of the financial indicator itself. The underlying idea is that, as long as financing conditions are broadly in line with historical averages, they may not matter for investment. Financing conditions may affect corporate investment to a significant extent only once particularly buoyant or unfavourable conditions prevail. Obviously, periods of favourable (respectively unfavourable) financial conditions are likely to match broadly the phases of higher (respectively lower) growth. Therefore, the two tests of possible non-linearities carried out in this paper, while complementary, are not fully independent.

Non-linearity over the business cycle

A business cycle chronology is determined using a two-stage Markov switching model of quarter-on-quarter real GDP growth:

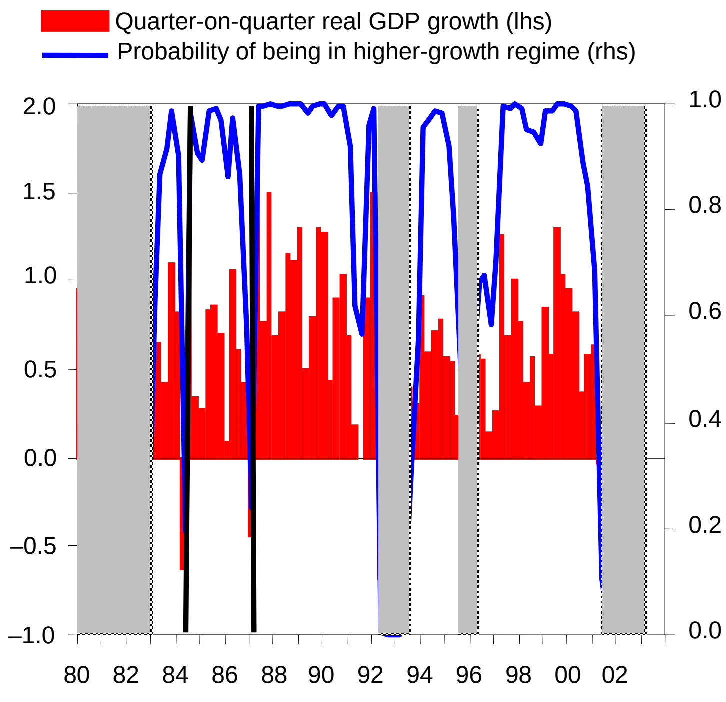
$$d\log(GDP) = \mu_S + \sigma_t v_t \text{ for s = 1,2} \tag{2}$$

where ν_t are independent and identically distributed random variables with zero mean and unit variance and μ_s corresponds to the average real GDP growth in regime *s*. The estimated average quarter-on-quarter GDP growth rates are 0.06% in the lower-growth phase and 0.74% in the higher-growth phase. Graph 10 shows the estimated probability of being in the high-growth phase. In this framework, three periods of lower growth are identified: the early 1980s, the early 1990s and the current slowdown. As usual in non-linear analysis, an important caveat to bear in mind when interpreting these results is the relatively low robustness. Graph 10 shows that the lower-growth regime has been a relatively rare event over the past two decades (32 out of 92 quarters in the sample considered), which tends to undermine reliability of the estimation of different elasticities over each regime.

Graph 10

Real GDP growth and probabilities of high and low growth regimes

In per cent



Note: Shaded areas denote low-growth phases.
Sources: Eurostat; authors' own estimates.

Table 5

Elasticities of investment to financial indicators in higher- and lower-growth regimes[1]

Indicator	Measure	Lag of financial indicator	Low growth	High growth	Significant difference[2]
Stock market capitalisation	GR	1	0	0.002	No
Share price index	GR	1	–0.001	**0.003**	No
	GRS	0	0	**0.004**	No
Dividend yield	L	0	–0.01	–0.02	–
Dividend/earnings ratio	L	0	–0.07	**–0.10**	No
Cost of equity issuance	L	0	0	0	–
Composite cost of financing	L	1	–0.03	**–0.08**	No
		3	0.02	**0.06**	No
	D	1	**–0.07**	**–0.11**	No
Yield curve	L	0	**0.08**	0	**Yes**
Corporate loans	GR	4	**–0.07**	**–0.02**	**Yes**
	GRS	0	**–0.08**	**0.07**	**Yes**
		4	–0.02	**–0.07**	No
	RX	3	**20.6**	–6.3	**Yes**
	RXD	4	**–124**	–31	**Yes**
Gross operating surplus	GR	1	**0.05**	0.00	**Yes**
		2	0.01	0.01	–
	GRS	4	**0.04**	**0.03**	No
Expected earnings	GR	0	**0.008**	**0.006**	No
	GRS	0	**0.017**	**0.005**	No
Corporate debt	RX	0	**–1.19**	**–0.70**	No
		3	**1.39**	0.35	No
	RXD	2	–0.15	–0.25	–

[1] Significant values are highlighted in bold. [2] Based on the standard errors of the estimated coefficients.

Switching regression equations are estimated in order to assess possible asymmetries over the business cycle in the response of investment to financial indicators, generically labelled *FIN_INDIC*. The following equation is estimated:[7]

$$d(NHIR) = C(s) + \alpha_1(s) * d(NHIR(-1)) + \alpha_2(s) * d(NHIR(-2)) + \beta_1(s) * d\log(GDP) + \gamma(s) * dCOST(-4) + \mu(s) * FIN\_INDIC$$

where s = 1 and 2 according to the chronology shown in Graph 10. That is, starting from the structure of the benchmark equation which had been selected by PC-GETS in the linear case, we include one

[7] More parsimonious specifications in which only the elasticity of investment to the financial indicator is regime-dependent have also been estimated. These failed to show any significant differences in the response of investment to financial variables across the stages of the business cycle. The results are available from the authors upon request.

financial indicator at a time and allow elasticities to differ between the two identified phases of the business cycle.

Table 5 shows the estimated elasticities of investment to financial variables in each of the two growth regimes. Only in a few cases are elasticities found to be significantly different between high- and low-growth phases. Moreover, within these cases, some indicators seem to be more relevant during the higher-growth phase, while others are more relevant during the lower-growth regime. An interesting feature stemming from this exercise relates to the elasticity of investment to long-term interest rates, which, in most cases, is found to be more negative during higher-growth periods. This result also holds when no financial indicator is included in the estimating equation. Moreover, it is usually the case that, during lower-growth periods, the elasticity of investment to long-term interest rates is not significant. This result supports the view that, at times in which the outlook is uncertain, companies tend to hold back their investment projects, even when cost of finance is attractively low.

Non-linearity according to state of financial indicators

Asymmetry of the response of investment to financial variables is analysed in a similar manner. For each financial variable, a Markov switching model with two regimes is estimated, thereby defining phases of "favourable" and "unfavourable" financial conditions. For instance, for share prices, the favourable phase corresponds to high-growth periods. Conversely, for corporate debt, the favourable phase corresponds to the regime of lower debt growth.[8] As before, switching regression equations are estimated. For some indicators, the phases defined by the Markov switching model do not lend themselves to such an estimation. Indeed, the dividend yield and the ratios of loans and debt to operating surplus are found to have been in the same regime since the mid-1980s. As a result, these variables are excluded from the analysis.

Table 6 shows the results, presented in the same way as in Table 5. In most cases, financial variables are found to be significant when they are favourable. As regards periods of unfavourable financing conditions, the various indicators give different results. Indicators of stock market developments are not found to be significant. This result could reflect the fact that companies have usually been able to find alternative sources of finance when the stock market declined (namely bank loans). However, corporate loans and gross operating surplus, two indicators reflecting the availability of funds for investment, seem to matter more during their unfavourable periods. For the latter indicator, attention needs to be drawn to the fact that, even for the phase of "unfavourable" conditions, the average growth rate is positive. The significantly positive investment elasticity in periods of high growth in operating surplus reduces to zero in phases where growth in gross operating surplus is relatively low. This finding on loans and operating surplus fits the argument of the existence of financial accelerator effects often found in studies based on firm-level data. When profit growth is low and/or leverage ratios are high, the extra effort needed to restore balance sheets acts as an additional negative factor on investment.

Overall, the econometric analysis presented in this paper suggests that financial variables add little information, if any, to explaining and forecasting developments in investment. There is some tentative evidence of asymmetries in the response of investment to financial variables depending on the state of the cycle and of financing conditions. First, when demand conditions (and hence prospects) are particularly bad, cost of finance does not seem to have any significant impact on investment. Second, when corporate profit growth is relatively low and/or corporate leverage is relatively high, investment seems to react more strongly to financing conditions.

---

8 For the growth rate of corporate loans, the classification between favourable and unfavourable phases is ambiguous. Loans as a reflection of availability of funds suggest that the higher growth phase would be the "favourable" one, while loan growth as an indicator of corporate leverage suggests that the lower growth phase would be the "favourable" one. Based on the positive correlation between investment and loans, "favourable" loan conditions in Table 6 correspond to periods of higher loan growth, but this is only a matter of presentation as elasticities are not found to be significantly different between phases.

Table 6

Elasticity of investment to financial indicators in favourable and unfavourable phases of financing conditions[1]

Indicator	Measure	Lag of financial indicator	Unfavourable	Favourable	Significant difference[2]
Stock market capitalisation	GR	1	0	0.003	No
Share price index	GR	1	0	**0.007**	**Yes**
	GRS	0	0	**0.004**	**Yes**
Dividend/earnings ratio	L	0	–0.10	**–0.09**	No
Cost of equity issuance	L	0	0	–0.14	**Yes**
Composite cost of financing	L	1	**–0.07**	**–0.06**	No
		3	**0.07**	**0.07**	No
	D	1	**–0.16**	**–0.08**	**Yes**
Yield curve	L	0	**0.06**	***–0.03***	**Yes**
Corporate loans	GR	4	0.02	**–0.04**	No
	GRS	0	0.07	0.01	No
		4	**–0.09**	**–0.08**	No
	RXD	4	**–124**	**–31**	**Yes**
Gross operating surplus	GR	1	0	**0.04**	**Yes**
		2	**0.01**	**0.02**	**Yes**
	GRS	4	0	**0.08**	**Yes**
Expected earnings	GR	0	***–0.01***	**0.01**	**Yes**
	GRS	0	***–0.007***	**0.01**	**Yes**

[1] Significant values are highlighted in bold. Italic cells: elasticity with wrong sign. [2] Based on the standard errors of the estimated coefficients.

4. Conclusions

The issue of possible financial constraints on a recovery in capital investment featured prominently in recent forecast discussions. This paper seeks to add to this discussion by examining the quantitative importance of financial variables in forecasts of aggregate investment. The methods used are somewhat crude and ad hoc, but the results broadly confirm prior perceptions. First, financial variables tend to be quantitatively insignificant in aggregate investment equations that include demand and cost of capital terms. On average, they help very little in improving the forecast accuracy of these equations. Second, there is some tentative evidence that the relevance of financial variables, if any, only emerges in particular periods. The results from linear specifications typically used in macroeconomic forecasting models should thus be cross-checked with the information from non-linear relationships. Overall, however, the analysis presented here suggests that, for forecasting purposes, not much is won when proceeding with aggregate investment equations that simply have indicators of financial conditions added to the set of right-hand variables. Put positively, this implies that the impact of financing conditions on investment should probably be taken into account in a more systematic and consistent way.

In principle, the quantity financial variables that interplay with expenditures on fixed capital investment can be forecast within a fully fledged flow of funds framework, in which the feedback mechanisms from the real to the financial side would be explicitly modelled through behavioural equations. Such a forecasting approach has been tested in some national central banks. The advantage is that it provides a closed and transparent system to discuss projections under different scenarios, letting forecasters monitor the different repercussions between financial and non-financial variables when changes in a position of a particular sector are rebalanced by changes in other variables along the accounting identities. In practice, however, the complexity of the behavioural relationships underlying flow of funds positions requires many restrictive assumptions and judgmental input. As a consequence, the uncertainty surrounding flow of funds forecasts is usually relatively high.

Annex 1:
Data sources

The quarterly data used in the regression analysis cover the period 1980:1 to 2003:1, with the exception of the cost of equity issuance measure, which is available as of 1987:1. For some variables, official data are only available for part of the sample period, and the missing data were compiled from the available national data.

National accounts

GDP (constant prices): Eurostat data from 1991:1, own estimates based on available national data prior to 1991:1.

GDP deflator: Eurostat data from 1991:1, own estimates based on available national data prior to 1991:1.

Non-housing investment (constant prices): Eurostat data from 1991:1, own estimates based on available national data prior to 1991:1.

Deflator for non-housing investment: Eurostat data from 1991:1, own estimates based on available national data prior to 1991:1.

Gross operating surplus (current prices): Eurostat data for total economy from 1991:1, own estimates based on available data prior to 1991:1. No breakdown in institutional sectors is available. Adjusted for income of self-employed, assuming identical per-head wage income of employees and self-employed.

Financial variables

Long-term interest rate: ECB calculation based on 10-year government bond yields or closest available bond maturity. COST used in benchmark equation is expressed as COST = log(1+LIRR*ITD/YED), where LIRR refers to 10-year government bond yields deflated by the GDP deflator. ITD/YED measures relative prices of capital goods as the ratio of the deflators for investment and GDP.

Yield curve: long-term (10-year) interest rate minus short-term (three-month) interest rate.

Stock market capitalisation and share price index: euro area overall variables computed and provided by Datastream, deflated by the GDP deflator.

Price/earnings ratio: Datastream data, calculated as total market value over total earnings, providing an earnings-weighted average of the ratios of constituents.

Dividend/yield ratio: Datastream data, calculated as total dividend amount as a percentage of the total market value for the constituents.

Dividend/earnings ratio: calculated as the product of dividend/yield and price/earnings ratios.

Expected earnings: calculated from Datastream data on price/earnings ratios and share prices, deflated by the GDP deflator.

Cost of equity issuance: ECB estimate (see Annex 2).

Composite cost of financing: ECB estimate (see Annex 2).

Debt (non-financial corporate sector): official ECB quarterly monetary and financial accounts for 1997:1 to 2003:1, prior to 1997:1 compilation based on available country data, deflated by the GDP deflator.

Loans (non-financial corporate sector): official ECB quarterly monetary and financial accounts for 1997:1 to 2003:1, prior to 1997:1 compilation based on available country data, deflated by the GDP deflator.

Annex 2: Compilation of cost of finance measures[9]

In this paper, two measures of the cost of non-financial corporations for taking up financing means are used: the cost of equity issuance and a composite cost of financing indicator.

The cost of equity issuance

While the interest payments paid on a bank loan or the coupons paid on a corporate bond can be considered as good measures of the cost of a bank loan and of issuing a corporate bond, there is no simple measure for the cost of issuing equity. The notion closest to the interest rate on a loan or a bond is the dividend yield, calculated as the ratio of current dividends per share over the price of the corporation's stock. However, dividend yields are only an imperfect measure of the cost of quoted equity, as such a measure must also take into account the fact that equities have no fixed maturity and are not subject to a systematic repayment of a fixed amount of capital at a fixed date in the future (like corporate bonds and bank loans).

The price of equity should be equal to the expected discounted sum of all future dividends paid out by the corporation. From this, it is possible to find a measure of the cost of equity that depends on the current dividend yield and on the growth rates of dividends in the future. As the chronology of future dividend growth rates is by nature unknown, two assumptions are necessary. First, it is assumed that the real average dividend growth rate for the next four years is equal to analysts' four-year-ahead real earnings growth rate expectations extracted from the monthly Thomson Financial First Call (TFFC) analysts' survey. Second, after a transition phase of eight years, the rate of growth in dividends is set to an estimate of the potential real GDP growth rate of the euro area economy, at 2.25%. This is the midpoint of the range assumed for trend potential growth in the calculation of the ECB's reference value for monetary growth. Overall, changes in the real cost of equity depend mainly on the current dividend yield and to a lesser extent on the analysts' four-year-ahead earnings growth rate expectations.

The composite cost of financing

The cost of financing of euro area non-financial corporations as used in this paper combines the marginal costs of taking up loans, market-based debt and quoted equity. The weights of the different components are based on the longer-term financing structure (in stocks) of non-financial corporations. Given data limitations, the cost of finance indicator does not address the impact of different tax regimes between financing vehicles or countries or the effect of possible non-price restrictions that non-financial corporations might face when choosing a financing means. The cost of loans, the cost of market-based debt and the cost of quoted equity have been weighted according to the shares of the notional stocks (calculated as outstanding amounts in 1997:4 extended by quarterly flows) of loans, market-based debt and quoted equity in these liabilities of non-financial corporations according to the quarterly financial accounts.

The cost of loans is measured as a composite lending rate based on short-term and long-term retail bank lending rates on loans to non-financial corporations. Due to data limitations, long-term interest rates have been estimated on a sample of euro area countries before November 1996 and back to 1990. Short-term cost and long-term cost of loans have been weighted according to the shares of the notional stocks of short-term and long-term loans in the loans of non-financial corporations.

The cost of market-based debt is obtained by aggregating yields of Merrill Lynch corporate bond indices. First, an index of the average yield of corporate bonds with a maturity greater than one year issued by euro area non-financial corporations with investment grade rating (ie BBB and better). Second, for high-yield bonds of non-financial corporations, the "total euro currency high-yield index" is

[9] Prepared by Louis Bé Duc, Stéphane Guéné and Petra Köhler.

used as a proxy. Before 1998 and back to 1990, corporate bond yields of a sample of euro area countries, weighted by GDP weights corresponding to the purchasing power parity in 2001, were used.

References

Angeloni, I, A Kashyap, B Mojon and D Terlizzese (2002): "Monetary transmission in the euro area: where do we stand?", *European Central Bank Working Paper Series*, no 114, January.

Benito, A and G Young (2002): "Financial pressure and balance sheet adjustment by UK firms", *Bank of England Working Paper Series*, no 168.

Bernanke, B, M Gertler and S Gilchrist (1998): "The financial accelerator in a quantitative business cycle framework", *NBER Working Papers*, no 6455.

Bond, S and C Meghir (1994): "Financial constraints and company investment", *Fiscal Studies*, vol 15, no 2.

Chatelain, J-B (2002): "Structural modelling of investment and financial constraints: where do we stand?", *National Bank of Belgium Working Paper* no 28.

Chirinko, R S (1993): "Business fixed investment spending: modelling strategies, empirical results, and policy implications", *Journal of Economic Literature*, vol XXXI, December.

Diebold, F and R Mariano (1995): "Comparing predictive accuracy", *Journal of Business and Economic Statistics*, vol 13.

Fagan G, J Henry and R Mestre (2001): "An area-wide model for the euro area", *European Central Bank Working Paper Series*, no 42, January.

Harvey A, S Leybourne and P Newbold (1997): "Testing the equality of prediction mean square errors", *International Journal of Forecasting*, vol 13.

Hernando, I and A Tiomo (2002): "Financial constraints and investment in France and Spain: a comparison using firm level data", *Banco de España Working Papers*, no 0214.

Hubbard, R G (1998): "Capital market imperfections and investment", *Journal of Economic Literature*, vol XXXVI, March.

Jaeger, A (2003): "Corporate balance sheet restructuring and investment in the euro area", *IMF Working Paper* no WP/03/117.

Gomes, J F (2001): "Financing investment", *American Economic Review*, vol 91, no 5.

Kaplan, S N and L Zingales (1995): "Do financing constraints explain why investment is correlated with cash flow?", *NBER Working Papers*, no 5267.

Mojon, B, F Smets and P Vermeulen (2002): "Investment and monetary policy in the euro area", *Journal of Banking and Finance*, vol 26.

Myers, S C (2001): "Capital structure", *Journal of Economic Perspectives*, vol 15, no 2.

PC-GETS: http://www.nuff.ox.ac.uk/users/doornik/pcgive/pcgets/index.html.

Schiantarelli, F (1996): "Financial constraints and investment: methodological issues and international evidence", *Oxford Review of Economic Policy*, vol 12, no 2.

Vermeulen, P (2002): "Business fixed investment: evidence of a financial accelerator in Europe", *Oxford Bulletin of Economics and Statistics*, vol 64, no 3.

Assessing the predictive power of measures of financial conditions for macroeconomic variables[1]

William English,[2] Kostas Tsatsaronis[3] and Edda Zoli[4]

Introduction

The interrelationships between the financial and real sectors are very complex. In theory, shocks to any financial market or set of financial institutions could have effects on other financial markets and institutions as well as on the real economy.[5] A great deal of research has focused on the ways in which monetary policy shocks can be transmitted to the real economy both through changes in market interest rates and also indirectly, by affecting agents' balance sheets. Such effects may provide a "financial accelerator" for monetary policy.[6] However, in recent years financial market developments not closely related to monetary policy appear to have played an increasing role in macroeconomic performance. These episodes include many instances of banking and foreign exchange crises, often with substantial real effects.[7] In addition, a number of countries have witnessed substantial booms in asset prices, often accompanied by rapid debt growth, that subsequently reversed with adverse macroeconomic consequences.[8]

Thus, when policymakers decide upon the appropriate stance of monetary policy, they must take account of the possible macroeconomic implications of developments in the financial sector. To do so, they must monitor not only risk-free interest rates and equity prices, but also risk spreads on various instruments, the financial health of businesses and households, the financial health of intermediaries, and the operation of financial markets.[9] With this information in hand, they then need to assess the likely implications of the financial developments for output and inflation.

One way to make such an assessment would be to build and estimate a large structural macroeconomic model that captured the effects of such factors. However, doing so would be difficult. Such an approach would require a structural model that included non-trivial financial markets and institutions and accounted for the effects of developments in markets and institutions on the factors influencing the spending behaviour of households and firms. Moreover, estimation of such a model would require data on the health of financial institutions, measures of risk aversion, and so on. In many cases, however, such variables are not observable, but must be judged from the behaviour of a number of possible indicator variables (such as capital ratios, profitability, asset quality, interest rate spreads and measures of debt and interest burdens).

An alternative approach that at first sight seems simpler would be to use a non-structural method, such as a VAR, to evaluate the effects of financial indicators for output and inflation. Such an

---

1 The authors would like to thank Angelika Donaubauer for her efforts in collecting and systematising the data for this study, and Maurizio Luisi for extremely able programming. All errors remain the sole responsibility of the authors. The views expressed in this paper are those of the authors and do not necessarily reflect those of the BIS, the IMF or the Federal Reserve System.

2 Board of Governors of the Federal Reserve System.

3 Bank for International Settlements.

4 International Monetary Fund.

5 See, for example, Tobin (1969).

6 For a recent treatment, see Bernanke et al (1999).

7 Kaminsky and Reinhart (1999) present evidence on the sources of dual banking and currency crises, and Hoggarth and Saporta (2001) examine the costs of such crises.

8 See Borio and Lowe (2002) for a discussion and some evidence on the possible predictability of such crises.

9 See Nelson and Passmore (2001) for one approach to such monitoring.

approach is difficult, however, because of the large number of measures that may affect the operation of financial institutions and markets and the relatively small number of degrees of freedom available. Including several lags of five, 10, or even more financial measures would quickly use up all of the degrees of freedom available. However, adding the variables one at a time to a baseline specification may give deceptive results, depending on the interrelationships among the financial indicators and the variables included in the baseline estimates.

In the light of these difficulties, the approach here follows the diffusion index method pioneered by Stock and Watson (2002).[10] Their method employs principal components to extract information from a large set of potentially informative indicator variables and then bases forecasts of the variables of interest on the principal components. Here we are interested in whether principal components based on a variety of financial variables can help to forecast output, inflation and investment.[11] To test the resulting empirical model, we compare it to an alternative model based on interest rates and spreads that are known to have forecasting power. The implicit assumption in our approach is that the key underlying factors influencing financial markets and institutions (for example, risk aversion or the financial health of intermediaries, non-financial firms and households) are well captured by the principal components, so that the inclusion of the components accounts for the bulk of the information contained in the factors.

We conduct our exercise for three countries (Germany, the United Kingdom and the United States) for which we were able to obtain data on a sufficient number of financial indicators. This cross-country approach allows us to see if the influence of various financial sector variables (as captured by the principal components) differs importantly across countries. One might expect such differences given the variation in the structure of financial markets and institutions in the different economies.

The next section describes our empirical approach and the data that we employ. The empirical results are described in Section II, and Section III provides some interpretation of the role of the factors. The final section concludes.

I. Method and data

Our approach is analogous to the diffusion index methodology proposed by Stock and Watson (2002) (hereafter referred to as SW). The method consists of extracting a set of principal components from a broad number of series that represent different aspects of the health and performance of financial markets and intermediaries, the level of financial activity, and financial market participants' assessment of future economic prospects.[12] All variables have been tested for stationarity using the Augmented Dickey-Fuller test.[13] In most cases, series for which a unit root could not be rejected at the 95% level have either been differenced or measured as percentage deviations from trend (see below for a discussion of the detrending procedure). However, in some cases - for example, some of the inflation rate and interest rate series - we chose to assume that differences were stationary rather than difference the variables a second time. As in SW, in order to avoid the possibility that measurement units and the volatility of individual series could unduly influence the estimation of the latent factors, all of the variables have been standardised (ie had their means subtracted and been divided by their standard deviations).

The SW procedure is based on the assumption that the set of predictor variables X_t and the variable to be forecast y_{t+k} can be expressed as functions of the same small set of underlying unobservable factors F_t as described by the following equations:

---

10 Bernanke and Boivin (2003) employ the same technique in developing forecasts for variables of interest to monetary policymakers.

11 We chose to forecast investment spending because it is the component of output that seems most likely to respond to financial developments.

12 A discussion of the motivation relating to the specific variables used can be found in the next subsection. A description of the complete set of series used for each country is listed in Appendix B.

13 We use five lags for the quarterly frequency variables and two lags for variables that are observed annually.

$$X_t = \Lambda F_t + e_t \text{ and } y_{t+k} = \alpha(L) y_t + \beta F_t + \eta_{t+k} \quad (1)$$

where Λ is the factor loading matrix, $\alpha(L)$ captures the autoregressive component of the variable being forecast, and β is a vector of coefficients on the financial factors. The idiosyncratic errors e_t are assumed to be weakly correlated across variables, and the forecast error η_{t+k} is assumed to be uncorrelated with the unobserved factors (ie $E[\eta_{t+k} | F_t] = 0$). SW show that asymptotically, in other words as the number of observations and the number of variables in X tend to infinity, the factors can be estimated consistently by principal components. The system (1) is potentially dynamic in the sense that X_t may contain lagged predictor variables, which will then influence the values of F_t.

Forecasting exercise

In our case, we use principal components to estimate a small set of unobserved factors that describe the systematic component of the variation in a large number of financial sector variables. We then explore the forecasting ability of the factors for three macro variables by estimating equations of the form:

$$y_{t+k} = \sum_{j=0}^{m} a_j y_{t-j} + \sum_{j=0}^{m} \beta_j \pi_{t-j} + \sum_{l \in [1,\cdots,6]} \sum_{j=0}^{n} \gamma_i F_{l,t-i} + \eta_{t+k} \quad (2a)$$

or

$$\pi_{t+k} = \sum_{j=0}^{m} a_j \pi_{t-j} + \sum_{j=0}^{m} \beta_j y_{t-j} + \sum_{l \in [1,\cdots,6]} \sum_{j=0}^{n} \gamma_i F_{l,t-i} + \eta_{t+k} \quad (2b)$$

where y is GDP or investment, and π is inflation. The choice of factors to be included in the right-hand side (among the six first principal components estimated in the previous step) and the specific lags for the factors and the variable that is being forecast are chosen by minimising the Bayesian Information Criterion (BIC). The value of the criterion declines with the goodness of fit, but it assigns penalties for lack of parsimony in the specification.[14]

To simplify our forecasting exercise, we choose to forecast the macro variables at the one- and two-year horizons (ie k of 4 or 8 with our quarterly data). These horizons seem appropriate for monetary policy decision-making. Moreover, the existing literature has documented that the forecasting ability of the term spread, a financial variable that is often found to have significant predictive ability for economic activity and inflation, is particularly strong at horizons in this range.[15] In order to reduce the effects of high-frequency noise in the variables to be forecast, we use four-quarter averages. For example, in the output regression, we use the average level of the output gap over the coming four quarters, and the average gap over the four quarters starting four quarters ahead.[16]

Horse race against standard variables

An extant body of the literature identifies a number of financial variables that have predictive ability for future macroeconomic developments. For instance, the predictive content of the term structure for future activity has been documented by Estrella and Hardouvelis (1991) and Estrella and Mishkin (1997), while Mishkin (1990a,b and 1991) has found that the term structure contains important information about future inflation. There is also evidence in the literature that stock prices contain information about future economic prospects. In order to guard against the risk that the predictive content of the principal components reflects primarily the inclusion of just a few standard variables,

---

[14] The exact formula for the criterion is: $BIC = \frac{k \log(T)}{T} + \log\left(\frac{1}{T}\sum_{t=1}^{T} u_t^2\right)$, where k is the number of variables in the regression, T is the sample size and u_t are the regression residuals.

[15] See Smets and Tsatsaronis (1997).

[16] SW also use averages, but for the eight-quarter-ahead forecasting exercises they use the eight-quarter average. We thought that the four-quarter average starting in four quarters was easier to interpret.

which might dominate our estimated factors, we run a so-called horse race. This test compares the forecasting power of the latent factors against three variables: the level of the short-term rate, the slope of the yield curve and growth in real equity prices.[17]

We perform the comparison in two ways. We first rerun equation (2) substituting the three specific variables for the set of latent factors *F*. As with the principal components, the number of lags is chosen to minimise the BIC criterion. We then compare the goodness-of-fit measures of the two models. If the latent variables do not possess superior information content then the new sets of equations should produce just as good or better fit.

The second step is a direct comparison of the two sets of variables in an "encompassing regression" framework.[18] Specifically, we add lags of the three specific financial variables to our preferred specification of the forecasting equation based on the latent financial factors. As before, the optimal lag structure is determined by using the BIC criterion. If the latent variables have any information content beyond that contained in the specific variables, then they should enter the augmented equation significantly and will improve the overall explanatory power of the model compared to either of the simpler specifications.

Data

The data we include in the derivation of the latent financial factors fall into one of the following categories: interest rates, exchange rates, risk spreads, asset prices, measures of household and business financial strength, credit aggregates, and measures of the health and performance of the banking sector.[19] Appendix B contains a detailed list of the variables used in the analysis for each country. In this subsection we will discuss the general characteristics of the financial variables we have included and their relevance for measuring the prevailing financial conditions.

The variables we have included are intended to capture aspects of the financial determinants of spending by households or businesses. Interest rates are a measure of the cost of capital and play a substantial role in models of consumption and investment spending. They also play a significant role in most empirical macroeconomic models.[20] The real exchange rate influences output through the level of net exports. Risk spreads capture the additional cost of funding for risky borrowers, and they have proved useful in the past in forecasting output.[21] Asset prices may play a number of roles. First, changes in asset prices will be reflected in the value of household wealth, and so will affect consumption spending.[22] Second, equity prices influence firms' cost of capital, and so should affect investment spending.[23] Third, increases in asset prices boost financial wealth and thereby increase the debt capacity of households and firms, facilitating further extensions of financing.[24] Similarly, measures of financial pressures on households and businesses (for example, debt burdens) could well influence credit terms and so propensities to take on additional debt to support spending. Credit aggregates and their components may play two roles, both picking up aspects of credit supply that are not captured by the interest rates and spreads included here and also reflecting demands for credit, which may be useful indicators of the economic outlook.[25] Finally, measures of the financial condition of banks are included to capture the ability and willingness of banks to provide credit to

---

[17] Since we include the inflation rate in the regression, we use the nominal short-term rate rather than the real short-term rate. It might be useful to also include a short-term credit spread in our horse race, but we do not have a short-term private rate for Germany over our sample.

[18] See, for example, Fair and Shiller (1990).

[19] Since real interest rates may matter more than nominal rates, we have also included inflation in the list of financial variables.

[20] For example, in the context of a structural model see Reifschneider et al (1999), and in the context of reduced-form models see Sims (1980a, b).

[21] The importance of risk spreads is emphasised in Bernanke (1990) and Friedman and Kuttner (1992).

[22] For a recent assessment of such effects, see Dynan and Maki (2001).

[23] Again, see Reifschneider et al (1999).

[24] This sort of effect is emphasised in the literature on the "financial accelerator". For example, see Bernanke et al (1999).

[25] Kashyap et al (1993) show that quantities can provide a useful signal of credit market effects in a forecasting context.

bank-dependent borrowers. The work on the economic effects of low levels of bank capital or the "bank credit" channel of monetary policy suggests that such effects can be substantial at times.[26]

In the case of asset prices and some of the credit variables, we have included both growth rates of the variables and their percentage deviations from a trend calculated using a Hodrick-Prescott filter. The inclusion of the deviations from trend is based on the view that such deviations will better capture the possible future effects of asset market imbalances on the macroeconomy than will the growth rates.[27] In order to avoid the possibility that future values of these variables could, by affecting the estimated trend, influence earlier measures of the deviation from trend, we calculate the trend value for each period based on data only through that period. This procedure has the added benefit that, leaving aside data revisions, one can think of the deviation from trend as available to policymakers in real time.[28]

We were not able to include the same set of variables for all countries analysed in this paper. In some cases relevant series or proxies were not available, or were only available for too short a time period. In many cases we excluded variables that were available only for a few years. In other cases, we used information available only at an annual frequency (but over a longer period), which we interpolated on the basis of their relationship to a large number of real and financial sector variables observed quarterly. For this interpolation we used an algorithm similar to that suggested by Stock and Watson (2002), but slightly modified as described in Appendix A.

II. Empirical results

This section contains the empirical results of our exercise. It first discusses the outcome of the principal components calculation, and then proceeds to describe the results from the forecasting exercises for output, investment and inflation.

The estimated factors

We apply the principal components methodology discussed above to the sets of financial variables for each country to estimate the unobserved financial factors. Table 1 gives an idea of the ability of the estimated factors to explain the overall variability of the financial measures. It shows the share of the overall variance of the financial measures used that is explained by the first 10 factors. The factors are labelled conventionally in descending order of their ability to capture the overall variance. The first factor explains the largest proportion of the variance, the second the next largest, and so on. There is surprisingly little cross-country variation in the explanatory power of these factors. There is a fairly general pattern: the first component explains about one eighth to one seventh of the common variance while the collective explanatory power of the first six factors is slightly higher than 50%. The prevalence of this pattern is especially surprising when one bears in mind that these are statistical factors, and so there is no reason why the second factor in order of importance for Germany, for example, should reflect the influence of the same set of underlying forces as in the other two countries.

The set of figures C.1-3 in Appendix C plot the time series of the first six estimated latent financial factors for each country. We will base the assessment of each factor's importance in driving business cycle developments on their ability to forecast a set of macroeconomic variables. Hence, we do not try to identify factors or select particular rotations of the factors that might render them more interpretable. Nevertheless, the movements of some of the estimated factors over time are suggestive of their close connection to developments in the financial sector. For instance, the patterns in the movement of

[26] Prominent proponents of this view include Peek and Rosengren (1995) and Bernanke and Lown (1991). For a discussion of the credit channel in the monetary transmission mechanism, see Bernanke and Blinder (1992).

[27] For a detailed argument along these lines, see Borio and Lowe (2002).

[28] The other included variables (interest rates, risk spreads and bank health measures) are not revised importantly. However, the bank data often lag substantially.

some factors resemble, at least in the sense of the timing of their peaks and troughs, the general movement of interest rates in the three countries. Other factors, however, appear to be far more volatile and have no clear link to the historical behaviour of any particular variable. Arguably, the normalisation and differencing of the variables undertaken in the construction of the components make these comparisons more difficult than one might first expect.

Table 1

The information content of the latent financial factors

Percentage of total variance explained

	United States	**Germany**	**United Kingdom**
Factor 1	14.1	13.0	13.8
Factor 2	12.0	10.9	10.5
Factor 3	9.4	8.6	8.9
Factor 4	7.1	7.7	8.6
Factor 5	5.4	6.9	6.6
Factor 6	5.3	5.8	5.8
Factor 7	4.6	5.2	5.2
Factor 8	3.9	4.8	4.4
Factor 9	3.5	4.2	4.3
Factor 10	3.1	3.6	4.1
Variance explained by first six components	53.3	53.0	54.2

Forecasting macro variables

The criterion we use for identifying the relevant latent factor structure that summarises the impact of the financial sector on the macroeconomy is based on the predictive ability of these variables for real sector developments at the one- and two-year horizons. We run a set of forecasting regressions of the form (2), where the variable to be forecast is alternatively: the output gap, the investment gap and the change in the inflation rate. Consistent with our definitions of the detrended debt and asset price series discussed earlier, we have defined the two gap variables to be the percentage difference between the actual values of GDP and private investment (less inventories) and their trend values based on a backward-looking Hodrick-Prescott filter. For the forecasting exercise, the left-hand variables are measured as the average quarterly values over the four-quarter period ending either four or eight quarters ahead. We present the results of these forecasting exercises in Tables 2.1-3, which are organised by the three variables being forecast. In each table we include the results of the exercise for the four- and eight-quarter-ahead forecasts for all the countries in our analysis.

There are three general patterns that emerge from a comparison of the results across variables, forecast horizons and countries. The first is that the latent factors do help to predict the macroeconomic variables. In all but three cases, these factors are significant at conventional levels in the forecasting equations. The performance of the financial factors is least impressive in the case of inflation, where the factors enter significantly in only four of the six equations. At least for the two gap variables, the significance of the financial factors appears to be somewhat greater at longer forecasting horizons.

The second noteworthy feature of the regressions is that the performance of the models, at least judged by the adjusted R^2, is quite good. With only two exceptions, these goodness-of-fit measures range between 40% and 85% for the two output measures. Both of the lower values relate to forecasts of investment over the eight-quarter horizon. Not surprisingly, the adjusted R^2s generally decline at the longer forecast horizon. The decline is relatively mild in a number of cases, however, perhaps suggesting that the effects of the financial variables on spending take time to emerge. By contrast, in the case of the inflation equations the performance of the models is less good, and the decline in predictive ability at the longer horizon is more pronounced.

The final regularity is that a fairly small number of the estimated financial factors generally enters the forecasting equations. Moreover, while the procedure for selecting the factors tries the first six in each equation, only the first four (in terms of their overall ability to describe the dynamics of the financial sector variables) are retained in any of the equations by the BIC. Moreover, there is relative stability in the set of selected factors across the two horizons: typically the same components appear in the forecasting models at both horizons, albeit sometimes with a difference in lag. We interpret this result as indicating that the dynamic relation between the financial sector factors and the two real sector variables is quite robust.

Table 2.1

The information content of financial factors for the output gap

	United States		**Germany**		**United Kingdom**	
	k = 4	**k = 8**	**k = 4**	**k = 8**	**k = 4**	**k = 8**
GAP_t	2.9954 [6.62]	0.2810 [0.81]	1.71873 [4.67]	0.0609 [0.20]	2.65 [8.73]	2.4883 [3.89]
GAP_{t-1}	–4.1953 [4.54]	–0.7604 [2.31]	–1.4306 [–3.93]	–0.8598 [2.77]	–2.2930 [8.00]	–2.7237 [4.63]
GAP_{t-2}	1.4292 [2.74]					
$INFL_t$	0.0011 [0.64]	0.0066 [2.98]	0.0036 [1.98]	–0.0076 [2.76]	–0.0019 [1.98]	–0.0037 [2.37]
$INFL_{t-1}$	0.0050 [2.31]	0.0055 [2.00]	8.46E-05 [0.06]	–0.0076 [2.98]	–0.0009 [0.88]	–0.0011 [0.57]
$INFL_{t-2}$	0.0054 [2.28]					
$PC1_t$	0.0003 [0.90]	0.0019 [3.68]	–0.0004 [0.72]		–0.0008 [1.87]	–0.0017 [3.27]
$PC1_{t-1}$	0.0008 [1.77]				–0.0010 [2.16]	
$PC2_t$				0.0036 [4.17]		–0.0010 [2.27]
$PC3_t$				0.0018 [3.56]	0.0009 [2.23]	
$PC3_{t-1}$					0.0003 [1.00]	
$PC4_t$				0.0015 [2.98]		
R^2 adj	57.6	40.3	45.6	45.8	84.9	66.8
RMSE						
Financial factors' significance	0.015	0.0005	0.4721	0.000	0.000	0.000

Table 2.2

The information content of financial factors for the investment gap

	United States		Germany		United Kingdom	
	k = 4	**k = 8**	**k = 4**	**k = 8**	**k = 4**	**k = 8**
Inv GAP_t	0.7612 [4.29]	–0.1960 [1.37]	0.258 [1.56]	0.0132 [0.06]	–0.4039 [4.04]	–0.0062 [0.08]
Inv GAP_{t-1}	–0.2281 [0.98]		0.0085 [0.41]	–0.4608 [2.05]		
Inv GAP_{t-2}	0.5337 [2.23]		–0.4787 [2.79]			
$INFL_t$	0.0083 [1.68]	0.0164 [2.65]	0.0014 [0.37]	–0.0151 [2.30]	0.0050 [0.82]	–0.0112 [1.87]
$INFL_{t-1}$	0.0093 [2.02]		–0.0157 [3.15]	–0.0181 [2.57]		
$INFL_{t-2}$	0.0117 [3.26]		–0.0210 [4.47]			
$PC1_t$	–0.0008 [0.75]				–0.0009 [0.56]	–0.0075 [5.86]
$PC1_{t-1}$	–0.0026 [2.93]				–0.0059 [4.55]	–0.0022 [1.67]
$PC2_t$			–0.0029 [1.70]	0.0069 [3.09]		
$PC2_{t-1}$			0.0114 [6.11]			
$PC3_t$				0.0038 [3.20]		
$PC4_t$			0.0057 [6.01]	0.0039 [2.78]		
$PC4_{t-1}$		0.0040 [3.28]	0.0053 [5.55]			
R^2 adj	49.2	19.5	58.9	28.9	58.5	55.4
RMSE						
Financial factors' significance	0.010	0.0015	0.000	0.000	0.000	0.000

Table 2.3

The information content of financial factors for inflation

	United States		Germany		United Kingdom	
	k = 4	**k = 8**	**k = 4**	**k = 8**	**k = 4**	**k = 8**
GAP_t	124.43 [4.19]	0.1488 [0.04]	2.0799 [0.41]	–6.5562 [1.31]	120.65 [3.78]	–17.16 [1.77]
GAP_{t-1}	–209.54 [3.94]				–115.37 [4.08]	
GAP_{t-2}	83.91 [2.99]					
$INFL_t$	–0.1955 [2.99]	–0.0220 [0.19]	–0.3251 [3.23]	–0.1989 [1.89]	–0.7057 [4.96]	0.2033 [2.11]
$INFL_{t-1}$	0.0904 [1.14]				0.0819 [0.75]	
$INFL_{t-2}$	0.2060 [3.15]					
$PC1_t$	0.1065 [5.09]		0.0655 [1.32]	–0.0586 [2.97]	–0.0368 [0.93]	
$PC1_{t-1}$			–0.1287 [2.65]			
$PC2_t$	0.0441 [2.88]				–0.0599 [1.48]	
$PC3_t$					–0.0761 [3.16]	0.0556 [1.80]
$PC4_t$		0.0463 [1.59]				
$PC4_{t-1}$		0.0087 [0.44]				
R^2 adj	44.7	1.8	17.6	14.0	45.5	11.5
RMSE						
Financial factors' significance	0.0000	0.2467	0.001	0.0039	0.0049	0.0777

Horse race results

To gauge the extent to which the predictive content of the estimated factors is superior to the information incorporated in more traditional financial variables, we run a set of "horse race" forecasting equations. For each model we include the short-term rate (three-month rate on government securities), the slope of the (nominal) yield curve between three months and 30 years and the growth rate in (real) stock prices as right-hand variables in addition to the estimated latent factors.

Table 3

"Horse race" against select financial variables: predicting the output gap

	United States				**Germany**				**United Kingdom**			
	k = 4		**k = 8**		**k = 4**		**k = 8**		**k = 4**		**k = 8**	
	Financial variables	**Encompassing regression**	**Financial variables**	**Encompassing regression**	**Financial variables**	**Encompassing regression**	**Financial variables**	**Encompassing regression**	**Financial variables**	**Encompassing regression**	**Financial variables**	**Encompassing regression**
$PC1_t$		–0.0000 [0.082]		0.0016 [2.695]		–0.0004 [0.735]				0.0002 [0.348]		–0.0022 [3.662]
$PC1_{t-1}$		0.0008 [2.045]		0.0003 [0.137]						–0.0018 [3.248]		
$PC2_t$								0.0035 [3.843]				–0.0004 [1.122]
$PC3_t$								0.0019 [3.346]		0.0007 [1.160]		0.0013 [3.093]
$PC3_{t-1}$										–0.0003 [0.585]		
$PC4_t$								0.0015 [2.731]				
Int rate $_t$	0.0016 [1.105]	0.00009 [0.672]	0.0003 [0.121]	0.0003 [0.137]	0.0030 [0.758]	0.0020 [0.532]	0.0084 [1.844]	–0.00001 [0.004]	–0.0020 [0.770]	–0.0012 [0.662]	–0.0018 [0.824]	–0.0056 [2.905]
Int rate $_{t-1}$									–0.0014 [0.920]	–0.0012 [0.723]		
Int rate $_{t-2}$									0.0013 [0.583]	–0.0027 [1.576]		
Term spread $_t$	0.00002 [1.607]	0.00000 [0.746]	0.00004 [1.706]	0.00003 [0.120]	0.00006 [2.389]	0.00006 [2.450]	–0.0000 [0.226]	–0.0000 [0.0341]	–0.0000 [0.220]	–0.0000 [0.239]	–0.00002 [0.740]	–0.0003 [1.975]
Term spread $_{t-1}$									–0.0000 [1.342]	–0.0000 [0.205]		
Term spread $_{t-2}$									–0.0000 [0.416]	–0.0000 [0.750]		

Table 3 (cont)

“Horse race” against select financial variables: predicting the output gap

	United States				**Germany**				**United Kingdom**			
	k = 4		**k = 8**		**k = 4**		**k = 8**		**k = 4**		**k = 8**	
	Financial variables	**Encompassing regression**	**Financial variables**	**Encompassing regression**	**Financial variables**	**Encompassing regression**	**Financial variables**	**Encompassing regression**	**Financial variables**	**Encompassing regression**	**Financial variables**	**Encompassing regression**
Equity price $_t$	0.0608 [3.397]	0.0497 [3.135]	0.0576 [2.081]	0.0192 [0.834]	0.0119 [1.579]	0.0117 [1.553]	0.0006 [0.041]	–0.0060 [0.449]	0.0365 [2.584]	0.0260 [1.969]	0.0490 [1.992]	0.0409 [3.150]
Equity price $_{t-1}$									0.0379 [3.450]	0.0348 [2.820]		
Equity price $_{t-2}$									0.0273 [2.528]	0.0153 [1.721]		
R^2 adj	57.9	61.1	27.6	41.8	49.3	49.5	17.3	43.8	83.7	90.7	47.6	74.8
Excl PCs		0.121						0.0010		0.0003		0.0001
Excl other		0.035		0.4194	0.0218	0.0106	0.1013	0.9708		0.0000	0.0362	0.0000

The results are tabulated in Table 3. For each country and each maturity, we report the outcome of two regressions: one that substitutes the three financial variables for the latent factors (left column), and one that includes both sets of variables (the encompassing regression, shown in the right column). For space considerations, we report only the coefficients for the financial variables and not those of the lags of the macroeconomic variables.

The general impression is that the estimated latent factors have greater information content than do the short-term yield, the slope of the yield curve and equity price growth. But there are important nuances across countries. In the case of the United States we find that equity prices are very good predictors of the output gap, especially at the one-year horizon. The first factor, however, maintains its significance in the encompassing regression, particularly at the two-year horizon. The results for Germany are more mixed. At the shorter horizon, the term spread is more significant than the estimated components. The opposite is true, however, in the longer-horizon forecasts, where all three components are more significant than the alternative variables. The results for the United Kingdom also point to the greater predictive ability of the latent financial factors. At both horizons, the inclusion of the estimated factors considerably increases the forecasting ability of the model, and in the encompassing framework these variables maintain their significance. However, it must be noted that the interest rate and especially equity price growth remain very significant. As was the case with the other two countries, the results are most favourable for the latent factors at the longer horizon.

Overall, we conclude that the latent financial factors contain strong and independent predictive power for the output gap. Their power is relatively stronger at the two-year horizon, suggesting that the latent factors are capturing relationships between the financial and real sectors of the economy that operate at a relatively lower frequency. This impression is reinforced by the fact that, in the case of the United States and the United Kingdom, the lagged value of the first latent factor is more significant than the contemporaneous value when forecasting at the one-year horizon.

III. Interpreting the factors

Composite factors

One can use the results of the forecasting exercise to calculate composite financial factors for the output gap, for each country. This factor is simply the linear combination of the components chosen based on the BIC. In other words, for a given country:

$$CF_t = \sum_{I \in [1,\cdots,6]}^{6} \sum_{j=0}^{n} \gamma_i F_{I,t-i} \tag{3}$$

where I sums over up to six included factors, and j sums over up to n lags. This composite factor captures the collective influence of the financial sector variables on the variable being forecast. In other words, if this combination is equal to zero, then one could argue that financial conditions are "neutral" with respect to future activity, while a positive (negative) value of the CF implies favourable (adverse) financial conditions.

Towards the construction of an FCI

This composite factor is relatively close in spirit to the monetary conditions indices or financial conditions indices (FCI) considered in the past. For example, for a time the Bank of Canada monitored a monetary conditions index that was a weighted average of the policy interest rate and the exchange rate, with weights chosen to reflect the relative effects of the two variables on output. More generally, Goldman Sachs has for some time employed a financial conditions index consisting of a weighted average of a real short-term rate, a real long-term rate, the real exchange rate and equity prices to monitor the influence of financial factors on the real economy.[29] The weights employed in the index are

[29] For the Canadian case, see Freedman (1994). For the Goldman Sachs index, see Dudley and Hatzius (1999).

chosen based on the effects of the variables in the Federal Reserve's quarterly model, as reported in Reifschneider et al (1999).

As noted by Macroeconomic Advisors, however, such indices impose the restriction that all of the financial variables included in the index are measured in the same period.[30] Thus, the lag structure of the different financial variables in any subsequent forecasting equation using the index is constrained to be the same. To avoid this problem, Macroeconomic Advisors uses a macroeconomic model to calculate the appropriate weights on the current and lagged values of a small set of financial measures to form an index that does not constrain the lag structure of the effects of the five variables to be the same. Nonetheless, this index only captures the effects of five variables: a real short-term interest rate, a real long-term interest rate, the real exchange rate, real household equity wealth and the price-earnings ratio.

By contrast, the approach taken here can potentially include many more financial variables, as well as a number of lags of those variables. Moreover, since the financial variables may enter the different factors with different weights, and the factors can enter the forecasting equation with different lags, our method imposes less structure on the effective lags employed for different financial variables. To check whether the composite indicator calculated on the basis of the forecasting regression results satisfies the condition that each component variable enters with the same lag structure, we have computed the correlation coefficients of the implied weights on these variables across different lags. These implied weights are calculated by multiplying the weights on the various financial variables in the factors by the coefficients on the factors in the forecasting equation, and then summing the resulting values separately for each lag of the financial variables.

Table 4 contains the results of these calculations for the three countries. The results, perhaps somewhat surprisingly given the discussion in Macroeconomic Advisors (1998), suggest that the lag structure does not differ as much as one might have suspected across the included variables. The correlation coefficients between the implicit weights that the variables are assigned in the composite factor across the different lags range between 66% and 99%. These relatively high correlations suggest that including current and lagged values of a single index of the financial variables at each date may not have a large effect on forecast accuracy. Indeed, we conjecture that if one averaged the individual weights across lags, and then used the average weights to construct an FCI at each date, forecasts based on that FCI would have forecasting power relatively close to that of the more general procedure used here.[31]

Table 4

Correlation coefficients across lags of individual component variable weights

		United States			Germany			United Kingdom		
		t	**t – 1**	**t – 2**	**t**	**t – 1**	**t – 2**	**t**	**t – 1**	**t – 2**
k = 4	T	1			1			1		
	t – 1	0.98	1		0.96	1		0.79	1	
	t – 2	0.88	0.94	1	..	..	..	0.66	0.66	1
k = 8	t	1			1			1		
	t – 1	0.88	1		0.91	1		0.74	1	
	t – 2	..	..	..	..	..	.	..	..	.

Note: Entries correspond to the correlation of the implicit weights on the financial variables at the lag shown in the top row with the implicit weights on the same variables at the lag shown in the first column. The implicit weights are calculated based on the weights on the variables in the estimated latent factors and the coefficients on these factors in the output gap regressions reported in Table 2.

[30] See Macroeconomic Advisors (1998).

[31] This is left for future investigation.

IV. Conclusions

This paper shows how one can use a method similar to that of Stock and Watson (2002) to incorporate a wide variety of information about financial markets and institutions into macroeconomic forecasts. The results suggest that the method has considerable promise. The financial factors captured with the principal components do a good job of forecasting future levels of output and investment. When compared to a standard set of forecasting variables, the factors generally appear to provide significant independent information. Indeed, the improvement in forecasts of output at longer horizons based on the financial factors is very substantial in some cases, suggesting that the standard variables may exclude important information about financial developments that affect output with a longer lag. By contrast, the financial factors do a much poorer job of forecasting inflation, suggesting that the main effects of financial developments are on the level of activity, with effects on inflation mostly indirect via the level of activity.

Appendices

A. The interpolation method for annual frequency series

Our objective was to base the derivation of the financial latent factors on as many variables as possible and, in particular, to include variables that contain information about the health and level of activity of financial intermediaries. To do so, we had to make use of variables that are available only at an annual frequency. As a result, we had to interpolate those variables to the quarterly frequency that we had chosen for our empirical analysis. This interpolation was done by adapting the methodology suggested by Stock and Watson (2002), which is based on a two-step procedure that is akin to the EM algorithm. In the first step, a number of factors ARE estimated on the basis of a set of series available at a quarterly frequency. These factors are then annualised and the series that are available only annually are projected on them by OLS regression. In the second step, the estimated coefficients of these regressions are used to construct quarterly series on the basis of the quarterly values of the estimated factors. Finally, we distribute the residuals from the fitted annual model to the quarterly interpolated series, so that the appropriate time aggregation of the interpolated series yields the original annual series. We have slightly modified the SW procedure to adapt it to the problem at hand. The following paragraphs detail these modifications. The interested reader is referred to the SW article for further details.

First, unlike the procedure discussed in Stock and Watson, we calculate the principal components and conduct the interpolation only once, rather than iterating on the estimation of factors and the interpolation of the annual variables until the estimated factors converge. We chose this approach because additional iterations changed the interpolated series only slightly, but they increased the volatility of the estimated latent factors considerably. We believe that this volatility may be a result of the smaller cross-section of variables used in our paper, which could lead the procedure to try to adjust the factors to better fit the interpolated series, which are in turn constructed from the factors themselves.

Second, while our main exercise employed only financial variables in the calculation of the principal components, we used both financial and real variables in the construction of the factors used for the interpolation of the annual series. We did so in order to be able to capture all the underlying forces that might influence the dynamics of the series being interpolated. We also included a one-period lag of all the quarterly financial and non-financial variables when calculating the principal components on the thought that the resulting components might better capture the dynamics in the series. The full list of real variables used is included in Appendix B.

Finally, we projected the series to be interpolated on the 20 first principal components (in other words, those that corresponded to the 20 largest eigenvalues of the covariance matrix). We used a stepwise OLS procedure to fit each of the annual frequency series onto a selected subset of the annualised series of the estimated principal components. The selection procedure resulted in the use of one to four components to fit each annual series. The estimated models for each series were then used to create the quarterly interpolated series for these variables on the basis of the quarterly values of the selected components.

B. Data tables

United States			
	Financial variables	**Frequency**	**Transformation**
1	Banks' capital and reserves/banks' total assets, sa	Quarterly	Differenced
2	Banks' credit to non-banks, sa/nominal GDP, saar	Quarterly	Differenced
3	Growth in real banks' credit to non-banks	Quarterly	None
4	Growth in nominal banks' credit to non-banks	Quarterly	None
5	Banks' credit to the private sector, sa/total banks' credit to non-banks, sa	Quarterly	Differenced
6	Banks' holdings of mortgage debt, sa/total banks' credit to non-banks, sa	Quarterly	Differenced
7	Banks' deposits from non-banks, sa/nominal GDP, saar	Quarterly	Differenced
8	Banks' deposits from non-banks, sa/bank loans to non-banks, sa	Quarterly	Differenced
9	Banks' deposits from non-banks, sa/broad money, sa	Quarterly	Differenced
10	Interbank deposits/banks' total assets, sa	Quarterly	Differenced
11	Banks' loans to non-banks, sa/nominal GDP, saar	Quarterly	Differenced
12	Growth in nominal banks' loans to non-banks	Quarterly	None
13	Growth in real banks' loans to non-banks	Quarterly	None
14	Growth in nominal commercial property price index	Quarterly	None
15	Growth in real commercial property price index	Quarterly	None
16	Total liabilities of non-fin corporations/nominal GDP, saar	Quarterly	Differenced
17	Households' total liabilities/nominal GDP, saar	Quarterly	Differenced
18	Flow of funds total debt/nominal GDP, saar	Quarterly	Differenced
19	Growth in nominal equity price index (S&P 500)	Quarterly	None
20	Growth in real equity price index (S&P 500)	Quarterly	None
21	Equity price-earnings ratio	Quarterly	Differenced
22	Growth in nominal residential house price index	Quarterly	None
23	Growth in real residential house price index	Quarterly	None
24	Yearly percentage change in CPI, sa	Quarterly	Differenced
25	Three-month commercial paper rate	Quarterly	Differenced
26	Corporate bond yields	Quarterly	Differenced
27	Ten-year government bond yields	Quarterly	Differenced
28	Three-month money market rate	Quarterly	Differenced
29	Federal funds rate	Quarterly	Differenced
30	Three-month T-bill rate	Quarterly	Differenced
31	Real long-term interest rate	Quarterly	Differenced
32	Real short-term interest rate	Quarterly	Differenced
33	Spread: three-month commercial paper rate – three-month money market rate	Quarterly	Differenced

United States (cont)

	Financial variables	Frequency	Transformation
34	Spread: three-month commercial paper rate – three-month T-bill rate	Quarterly	Differenced
35	Spread: corporate bond yields – 10-year government bond yields	Quarterly	None
36	Term spread: 10-year – three-month paper	Quarterly	None
37	Growth in real effective exchange rate	Quarterly	None
38	Total international reserves minus gold, sa/broad money, sa	Quarterly	Differenced
39	Banks' credit to non-banks/GDP gap	Quarterly	None
40	Banks' credit to the private sector/GDP gap	Quarterly	Differenced
41	Real house price index gap	Quarterly	None
42	Real commercial property price index gap	Quarterly	None
43	Real equity price index gap	Quarterly	Differenced
44	Banks' net interest income/banks' total average assets	Annual	Differenced
45	Banks' provisions on loans /banks' loans to non-banks	Annual	Differenced
46	Banks' return on assets	Annual	Differenced
47	Banks' return on equity	Annual	Differenced

United States

	Real variables	Frequency	Transformation
1	Real GDP growth, saar	Quarterly	None
2	Real GDP gap	Quarterly	None
3	Nominal private investment, saar/nominal GDP, saar	Quarterly	Differenced
4	Nominal private investment/GDP gap	Quarterly	None
5	Real private investment growth	Quarterly	None
6	Nominal government spending/nominal GDP, saar	Quarterly	Differenced
7	Growth in real government spending	Quarterly	None
8	Nominal private consumption expenditure, saar/nominal GDP, saar	Quarterly	Differenced
9	Growth in real private consumption expenditure, saar	Quarterly	None
10	Nominal total consumption, saar/nominal GDP, saar	Quarterly	Differenced
11	Unemployment rate, sa	Quarterly	Differenced
12	Current account balance, sa/nominal GDP, saar	Quarterly	Differenced

	Germany		
	Financial variables	**Frequency**	**Transformation**
1	Banks' credit to non-banks/nominal GDP, sa	Quarterly	Differenced
2	Growth in real banks' credit to non-banks	Quarterly	None
3	Growth in nominal banks' credit to non-banks	Quarterly	None
4	Banks' credit to the private sector/total banks' credit to non-banks	Quarterly	Differenced
5	Banks' deposits from non-banks, sa/nominal GDP, sa	Quarterly	Differenced
6	Banks' deposits from non-banks, sa/bank loans to non-banks, sa	Quarterly	Differenced
7	Banks' loans to non-banks, sa/nominal GDP, sa	Quarterly	Differenced
8	Growth in nominal banks' loans to non-banks	Quarterly	None
9	Growth in real banks' loans to non-banks	Quarterly	None
10	Growth in nominal equity price index (Dax index)	Quarterly	None
11	Growth in real equity price index (Dax index)	Quarterly	None
12	Equity price-earnings ratio	Quarterly	Differenced
13	Yearly percentage change in CPI, sa	Quarterly	Differenced
14	Day-to-day money rate	Quarterly	None
15	Ten-year government bond yields	Quarterly	Differenced
16	Three-month government rate	Quarterly	Differenced
17	Real long-term interest rate	Quarterly	Differenced
18	Term spread: 10-year – three-month paper	Quarterly	Differenced
19	Growth in real effective exchange rate	Quarterly	None
20	Banks' credit to non-banks/GDP gap	Quarterly	None
21	Banks' credit to the private sector/GDP gap	Quarterly	None
22	Real equity price index gap	Quarterly	None
23	Banks' capital and reserves/banks' total assets, sa	Annual	Differenced
24	Banks' net interest income/banks' total average assets	Annual	Differenced
25	Banks' provisions on loans /banks' loans to non-banks	Annual	None
26	Banks' return on assets	Annual	None
27	Banks' return on equity	Annual	None
28	Growth in nominal commercial property price index	Annual	None
29	Growth in real commercial property price index	Annual	None
30	Growth in nominal residential house price index	Annual	None
31	Growth in real residential house price index	Annual	None
32	Real house price index gap	Annual	None
33	Real commercial property price index gap	Annual	None
34	Total liabilities of non-fin corporations/nominal GDP, saar	Annual	Differenced
35	Households' total liabilities/nominal GDP, saar	Annual	Differenced

Germany			
	Real variables	**Frequency**	**Transformation**
1	Real GDP growth, sa	Quarterly	None
2	Real GDP gap	Quarterly	None
3	Nominal investment, sa/nominal GDP, sa	Quarterly	Differenced
4	Nominal investment/GDP gap	Quarterly	None
5	Real investment growth	Quarterly	None
6	Nominal government spending, sa/nominal GDP, sa	Quarterly	Differenced
7	Growth in real government spending	Quarterly	None
8	Nominal private consumption expenditure, sa/nominal GDP, sa	Quarterly	Differenced
9	Growth in real private consumption expenditure, sa	Quarterly	None
10	Nominal total consumption, sa/nominal GDP, sa	Quarterly	Differenced
11	Unemployment rate, sa	Quarterly	Differenced
12	Current account balance, sa/nominal GDP, sa	Quarterly	Differenced

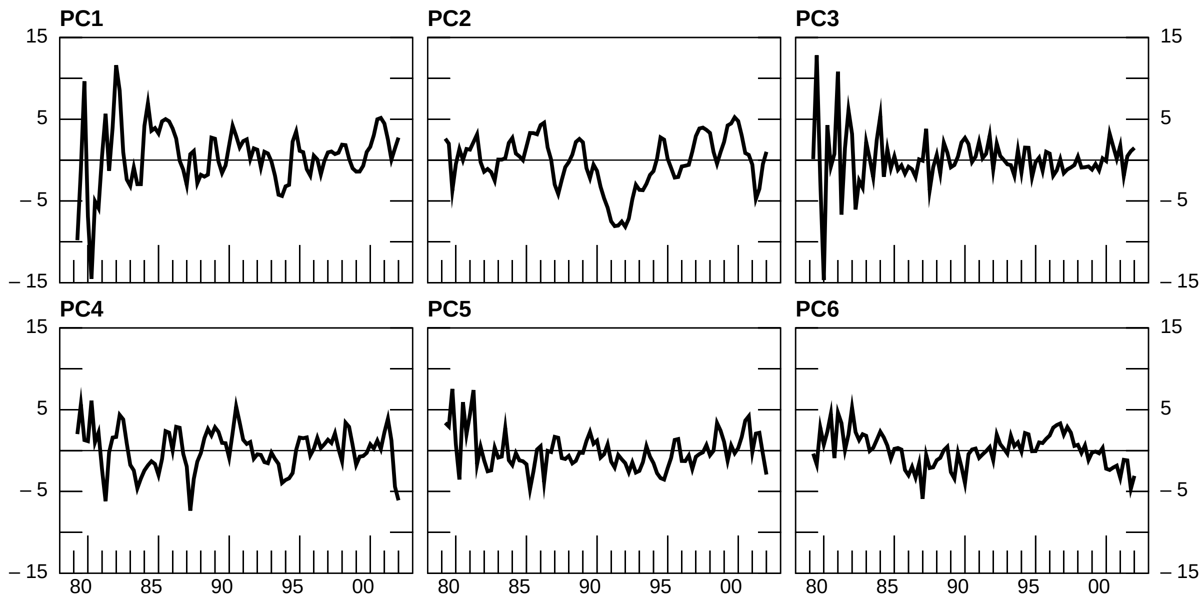
United Kingdom

	Financial variables	Frequency	Transformation
1	Banks’ credit to non-banks, sa/nominal GDP, sa	Quarterly	Differenced
2	Growth in real banks’ credit to non-banks	Quarterly	None
3	Growth in nominal banks’ credit to non-banks	Quarterly	None
4	Banks’ credit to the private sector/total banks’ credit to non-banks, sa	Quarterly	Differenced
5	Banks’ deposits from non-banks, sa/nominal GDP, sa	Quarterly	Differenced
6	Banks’ deposits from non-banks, sa/broad money, sa	Quarterly	Differenced
7	Interbank deposits/banks’ total assets, sa	Quarterly	Differenced
8	Growth in nominal equity price index (FTSE All Share)	Quarterly	None
9	Growth in real equity price index (FTSE All Share)	Quarterly	None
10	Equity price-earnings ratio	Quarterly	Differenced
11	Growth in nominal residential house price index	Quarterly	None
12	Growth in real residential house price index	Quarterly	None
13	Yearly percentage change in CPI, sa	Quarterly	Differenced
14	Policy rate: official band 1 dealing rate	Quarterly	Differenced
15	Overnight sterling interbank deposit rate	Quarterly	Differenced
16	Ten-year government bond yields	Quarterly	Differenced
17	Money market rate: three-month sterling interbank deposit rate	Quarterly	Differenced
18	Three-month government rate	Quarterly	Differenced
19	Real long-term interest rate	Quarterly	Differenced
20	Real short-term interest rate	Quarterly	None
21	Spread: three-month money market rate – policy rate	Quarterly	None
22	Term spread: 10-year – three-month paper	Quarterly	Differenced
23	Growth in real effective exchange rate	Quarterly	None
24	Total international reserves minus gold/broad money, sa	Quarterly	Differenced
25	Real house price index gap	Quarterly	None
26	Real equity price index gap	Quarterly	Differenced
27	Banks’ net interest income/banks’ total average assets	Annual	Differenced
28	Banks’ provision expenses/banks’ loans to non-banks	Annual	Differenced
29	Banks’ return on assets	Annual	Differenced
30	Banks’ return on equity	Annual	Differenced
31	Banks’ capital and reserves/banks’ total assets, sa	Annual	Differenced
32	Banks’ deposits from non-banks, sa/bank loans to non-banks	Annual	Differenced
33	Banks’ loans to non-banks/nominal GDP, sa	Annual	Differenced
34	Growth in nominal banks’ loans to non-banks	Annual	None
35	Growth in real banks’ loans to non-banks	Annual	None
36	Growth in nominal commercial property price index	Annual	None
37	Growth in real commercial property price index	Annual	None

	United Kingdom		
	Real variables	**Frequency**	**Transformation**
1	Real GDP growth, sa	Quarterly	None
2	Real GDP gap	Quarterly	None
3	Nominal investment, sa/nominal GDP, sa	Quarterly	Differenced
4	Nominal investment/GDP gap	Quarterly	None
5	Real investment growth	Quarterly	None
6	Nominal government spending, sa/nominal GDP, sa	Quarterly	Differenced
7	Growth in real government spending	Quarterly	None
8	Nominal private consumption expenditure, sa/nominal GDP, sa	Quarterly	Differenced
9	Growth in real private consumption expenditure	Quarterly	None
10	Nominal total consumption, sa/nominal GDP, sa	Quarterly	Differenced
11	Unemployment rate, sa	Quarterly	Differenced
12	Current account balance, sa/nominal GDP, sa	Quarterly	Differenced

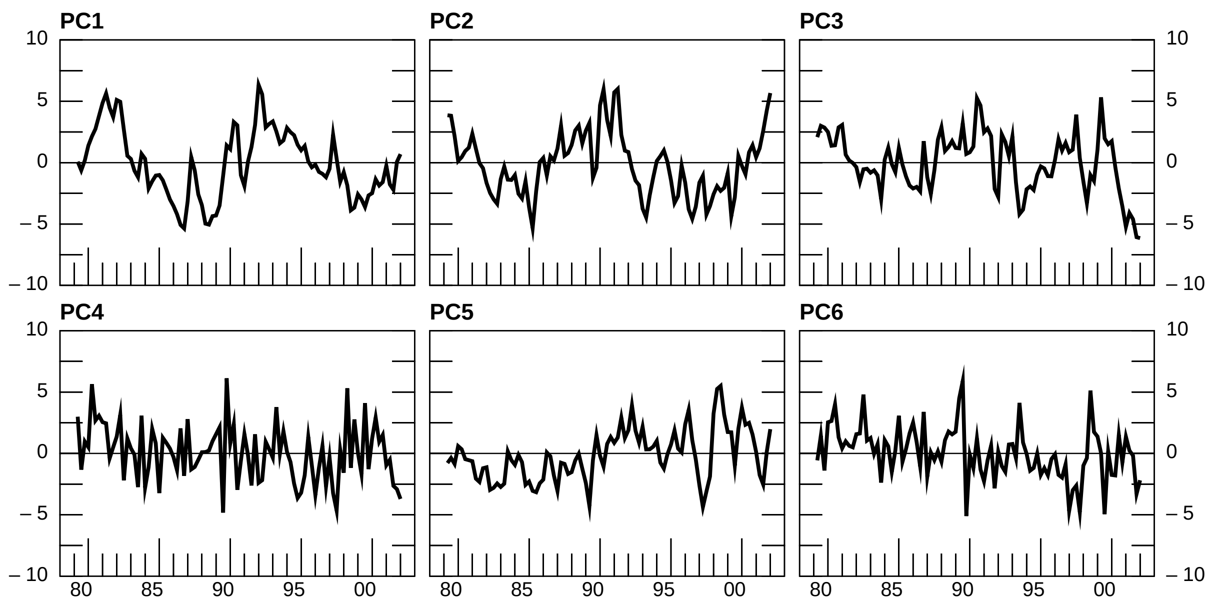
C. The estimated latent financial factors

Estimated financial factors: United States



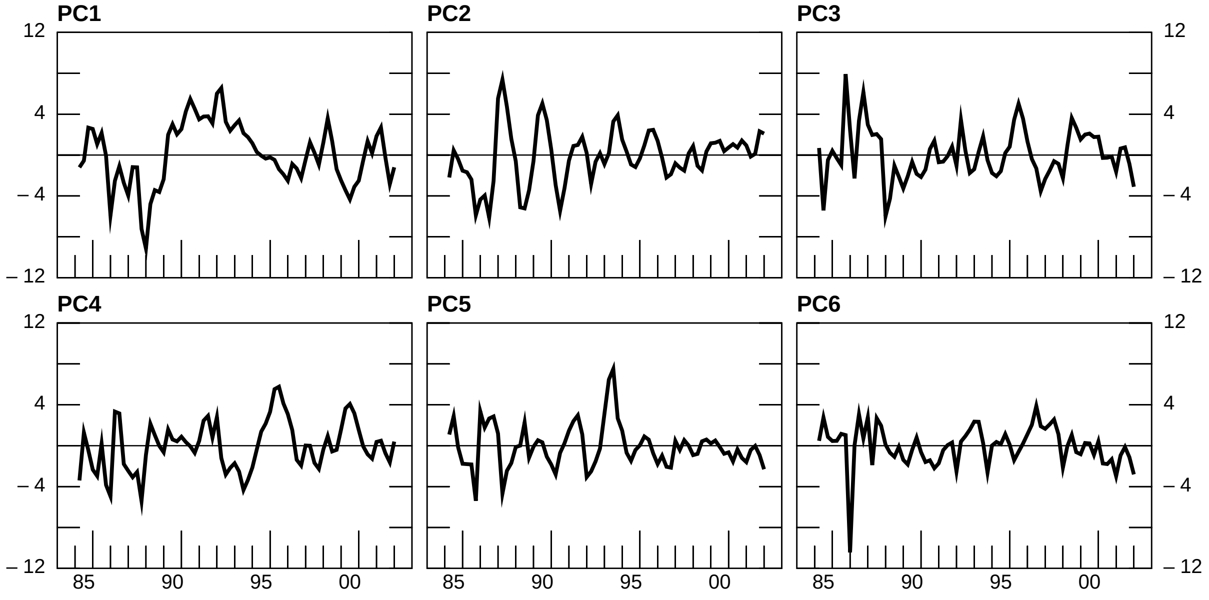
Source: BIS calculations.

Estimated financial factors: Germany



Source: BIS calculations.

Estimated financial factors: United Kingdom



Source: BIS calculations.

References

Bernanke, B (1990): “On the predictive power of interest rates and interest rate spreads”, *New England Economic Review*, Federal Reserve Bank of Boston, November-December, pp 51-68.

Bernanke, B and A Blinder (1992): “The federal funds rate and the channels of monetary transmission”, *American Economic Review*, 82(4), pp 901-21.

Bernanke, B and J Boivin (2003): “Monetary policy in a data-rich environment”, *Journal of Monetary Economics*, 50, pp 525-46.

Bernanke, B, M Gertler and S Gilchrist (1999): “The financial accelerator in a quantitative business cycle framework”, in J Taylor and M Woodford (eds), *Handbook of Macroeconomics*, Elsevier, Amsterdam.

Bernanke, B and C Lown (1991): “The credit crunch”, *Brookings Papers on Economic Activity*, 2, pp 205-39.

Borio, C and P Lowe (2002): “Asset prices, financial and monetary stability: exploring the nexus”, *BIS Working Papers*, no 114, Basel, July.

Dudley, W and J Hatzius (1999): “The Goldman Sachs financial conditions index: still accommodative after all these years”, *Global Economics Papers*, no 26, Goldman Sachs, September.

Dynan, K and D Maki (2001): “Does stock market wealth matter for consumption?”, *Finance and Economics Discussion Series*, no 2001-23, Federal Reserve Board, May.

Estrella, A and G Hardouvelis (1991): “The term structure as a predictor of real economic activity”, *Journal of Finance*, 46, pp 555-76.

Estrella, A and F S Mishkin (1997): “The predictive power of the term structure of interest rates in Europe and the United States: implications for the European Central Bank”, *European Economic Review*, 41, pp 1375-401.

Fair, R C and R J Shiller (1990): “Comparing information in forecasts from econometric models”, *American Economic Review*, 80, pp 375-89.

Freedman, C (1994): “The use of indicators and of the monetary conditions index in Canada”, in T J T Balino and C Cottarelli (eds), *Frameworks for monetary stability: policy issues and country experiences*, Chapter 18, pp 458-76, International Monetary Fund, Washington, DC.

Friedman, B and K Kuttner (1992): “Money, income, prices, and interest rates”, *American Economic Review*, 82(3), pp 472-92.

Hodrick, R J and E Prescott (1997): “Post-war US business cycles: an empirical investigation”, *Journal of Money, Credit and Banking*, 29, pp 1-16.

Hoggarth, G and V Saporta (2001): “Costs of banking system instability: some empirical evidence”, *Bank of England Financial Stability Review*, 10, Article 5.

Kaminsky, G and C Reinhart (1999): “The twin crises: the causes of banking and balance-of-payments problems”, *American Economic Review*, 89(3), pp 473-500.

Kashyap, A, J Stein and D Wilcox (1993): “Monetary policy and credit conditions: evidence from the composition of external finance”, *American Economic Review*, 83(1), pp 383-401.

Macroeconomic Advisors (1998): *Technical notes from the US economic outlook*, September.

Mishkin, F S (1990a): “The information in the longer maturity term structure about future inflation”, *The Quarterly Journal of Economics*, CV, 442, pp 815-28.

——— (1990b): “What does the term structure tell us about future inflation?”, *Journal of Monetary Economics*, 25, pp 77-95.

——— (1991): “A multi-country study of the information in the shorter maturity term structure about future inflation”, *Journal of International Money and Finance*, 10, pp 2-22.

Nelson, W and W Passmore (2001): “Pragmatic monitoring of financial stability”, *BIS Papers*, no 1, pp 367-84.

Peek, J and E Rosengren (1995): "The capital crunch: Neither a borrower nor a lender be", *Journal of Money, Credit and Banking*, August, pp 625-38.

Reifschneider, D, R Tetlow and J Williams (1999): "Aggregate disturbances, monetary policy and the macroeconomy: the FRB/US perspective", *Federal Reserve Bulletin*, January, pp 1-19.

Sims, C (1980a, b): "Macroeconomics and reality", *Econometrica*, January, pp 1-48.

——— (1980a, b): "Comparison of interwar and postwar business cycles: monetarism reconsidered", *American Economic Review*, 70(2), Papers and proceedings, pp 250-57.

Smets, F and K Tsatsaronis (1997): "Why does the yield curve predict economic activity? Dissecting the evidence for Germany and the United States", *BIS Working Papers*, no 49.

Stock, J H and M W Watson (2002): "Macroeconomic forecasting using diffusion indexes", *Journal of Business & Economic Statistics*, 20(2), pp 147-62.

Tobin, J (1969): "A general equilibrium approach to monetary theory", *Journal of Money, Credit and Banking*, 1(1), pp 15-29.

Linking real activity and financial markets: the first steps towards a small estimated model for Canada

Céline Gauthier and Fuchun Li[1]

1. Introduction

The crudest feature of many models' treatment of financial markets is that they aggregate all financial markets into only two: the market for money and the market for everything else. This aggregation allows us to summarise asset market equilibrium in a single LM curve but hides the structure needed to achieve a good understanding of how financial markets and the real economy are interrelated.

Another weakness of most models that purport to describe the transmission mechanism is their failure to pass the simple test of generating a different steady state rate of inflation in response to a series of monetary policy actions.[2] Such models with an unique steady state rate of inflation are very difficult to reconcile with the unit root test results found in the empirical literature.[3] One goal of this paper is to identify permanent shocks causing inflation to reach a new steady state rate of growth. A second goal is to model equilibrium values of financial variables through their long-run relationships with real variables in a tractable macroeconomic model.

Over the past two decades, there has been a growing interest in developing tractable macroeconomic models with transparent theoretical foundations. As written in Garratt et al (2001): "There are two main theoretical approaches to the derivation of long-run, steady state relations of a core macroeconomic model. One possibility is to start with the inter-temporal optimisation problems faced by 'representative' households and firms and solve for the long-run relations. […] An alternative approach […] is to work directly with the arbitrage conditions which provide inter-temporal links between prices and asset returns in the economy as a whole. […] The strength of the inter-temporal optimisation approach lies in the explicit identification of macroeconomic disturbances as innovations (shocks) to processes generating tastes and technology. However, this is achieved at the expense of often strong assumptions concerning the form of the underlying utility and production functions." In contrast, the approach that Garratt et al (2001) and the present paper adopt, focuses on long-run theory restrictions and leaves the short-run dynamics largely unrestricted (in the context of a VECM model), thus providing a much more flexible modelling strategy.

Our aim is to combine Garratt et al's (2001) approach with King et al's (1991) methodology allowing the identification of permanent shocks in a cointegrated system. Crowder et al (1999), Dhar et al (2000), Jacobson et al (2001) and Cassola and Morana (2002) all follow this *route* and show the degree of "structure" that may be assigned to a simple vector autoregression (VAR) framework characterised by cointegration if one embraces sufficient identifying restrictions.

The building blocks of the model consist of three cointegrating relations: (1) a money market equilibrium relation, (2) an arbitrage relation between short- and long-term bonds, and (3) a long-run relation between the stock market and real output. This last relation allows the identification of a supply shock as the only shock permanently affecting the stock market and a demand shock leading to significant transitory stock market overvaluation. We also identify a nominal shock defined as the only shock having a permanent impact on the level of inflation. In future work, we will study the

[1] Corresponding author: CGauthier@bank-banque-canada.ca. The authors wish to thank Scott Hendry, Pierre St-Amant, David Tessier and Carolyn Wilkins for helpful comments on a previous version of the paper and Alejandro Garcia for his technical assistance. The views expressed herein are those of the authors and do not necessarily reflect those of the Bank of Canada.

[2] More details on this point are made in Selody (2001).

[3] This is also a very difficult issue as inflation is expected to become stationary in a successful inflation targeting environment.

behaviour of a monetary policy reaction function consisting in reversing any identified nominal shock causing inflation to permanently deviate from the target.

Our paper is organised as follows. The theoretical foundations of the model are presented in Section 2. The cointegration analysis and specification test results are given in Section 3. Section 4 presents the econometric formulation of the core model. Section 5 analyses the impulse response functions. A conclusion follows.

2. The theoretical foundations of the model

In this section, we describe the long-run relations used as the building blocks of our model. We base our core model on Blanchard (1981), who develops a simple model of the determination of output, the stock market and the term structure of interest rates. The model is an extension of the IS-LM model. However, whereas the IS-LM model emphasises the interaction between "the interest rate" and output, Blanchard's model emphasises the interactions between output and four marketable asset values. These are shares which are titles to the physical capital, private short- and long-term bonds issued and held by individuals, and money.

Linking the real economy and the stock market

We assume that there are two main determinants of spending.[4] The first is the value of shares in the stock market. It may affect spending directly through the wealth effect on consumers, or indirectly through its impact on the borrowing capacity of consumers and investors (the credit channel effect); determining the value of capital in place relative to its replacement costs, it affects investment. The second is current income, which may affect spending independently of wealth if consumers are liquidity-constrained. Total spending is expressed as:

$$d_t = \alpha sm_t + \beta y_t;\ \ \alpha > 0;\ \ \beta > 0 \tag{1}$$

All variables are real, d denotes spending, sm is the stock market value, and y is income.[5] We can see equation (1) as a forward-looking aggregate spending curve with sm being a function of the present value of expected future profits, the latter being a function of expected future output. Hence, aggregate spending is implicitly a negative function of actual and expected interest rates and a positive function of actual and future expected output. Output adjusts to spending over time:

$$\dot{y}_t = \sigma(d_t - y_t) = \sigma(\alpha sm_t - by_t);\ \ \ \sigma > 0;\ \ b \equiv 1 - \beta \tag{2}$$

where a dot denotes a time derivative. Since output growth is a stationary variable and the level of output and the stock market price are both $I(1)$ variables, equation (2) can be seen as an error correction equation positively linking the short-run dynamics of output to deviations of the stock market from the real economy. Such a long-run relation between output and the stock market implies that transitory changes in output cannot permanently affect the level of the stock market.

Money market equilibrium

Portfolio balance is characterised by a long-run relation between money, output, interest rate and inflation:

$$M_t - p_t = cy_t - hi_t - \beta\pi_t;\ \ c > 0;\ \ h > 0;\ \ \beta > 0 \tag{3}$$

where i denotes the short-term nominal rate, y is real income, M and p denote the logarithms of nominal money and the price level, and π is the level of inflation. The parameter c is positive because an increase in output shifts the money demand for transaction purposes upwards; an increase in the

4 Blanchard also includes a balanced budget change in public spending as a third determinant of total spending.

5 No stochastic error terms are included in this section to simplify the presentation.

interest rate and an increase in inflation both increase the opportunity cost of holding money, which decreases real balance. Given that all the variables in equation (3) are better characterised as $I(1)$ variables, if deviations of real money from its determinants are transitory, then this equation represents a cointegrating relationship.

Arbitrage between short- and long-term bonds

The expectations hypothesis is perhaps the best known and most intuitive theory of the term structure of interest rates. If lr_t is the nominal yield to maturity of a discount bond and i_t is the period-t one-period rate, the expectations hypothesis in the absence of uncertainty implies that

$$(1+lr_t)^n = \prod_{i=0}^{n-1}(1+i_{t+i}) \tag{4}$$

This is an arbitrage condition ensuring that the holding-period yield on the n-period bond is equal to the yield from holding a sequence of one-period bonds. Taking logs of both sides and recalling that $\ln(1+x) \sim x$ for small x, yields a common approximation:

$$lr_t = \frac{1}{n}\sum_{i=0}^{n-1} i_{t+i} \tag{5}$$

The long-term yield is equal to the average of one-period yields. Hence, a permanent shock to the short-term yield will, in the long run, be reflected one for one in the long-term yield, once the shock is correctly perceived as permanent by the financial markets. This defines a third cointegration relationship.

3. Cointegration analysis

We estimate a monthly VECM over the 1975-2002 period with six endogenous variables and one exogenous variable and two lags.[6] The endogenous variables are the following Canadian variables: real gross domestic product (GDP) at basic prices,[7] the over 10-year marketable bond rate, the overnight rate,[8] a broad money aggregate (CPI deflated M2++),[9] the real stock market price (CPI deflated Toronto Stock Exchange (TSE)), and the CPI year-over-year inflation rate. Given the strong economic links between Canada and the United States, we incorporate as an exogenous variable the real US industrial production index, one available monthly proxy for US activity. This will allow simulation of different US scenarios. Unit root tests indicate that all variables can be treated as $I(1)$ variables. We add a dummy equalling one from 1993 onwards to capture the change in the trend of inflation after the adoption of an inflation target in 1991.

Based on the theoretical foundations of the core model described in the above section, we expect to find three cointegrating relations in the estimated VECM (as described by equations (2), (3) and (5)). The cointegration tests corrected for the presence of one exogenous variable, as proposed by Pesaran et al (2000), are presented in Table 1. The L-max test indicates the presence of three cointegration vectors, supporting our a priori expectations based on Blanchard's model, while the Trace test suggest only two vectors.

---

6 Two lags minimise the Hannan-Quinn and Schwartz information criteria and are sufficient to remove the correlation in residuals. We use monthly data because the Bank of Canada has adopted a fixed action date schedule eight times a year. A series of specification tests have been done and will be included in the next version of this paper.

7 This series has been merged with real GDP at factor cost for the period 1975-80.

8 As noted in Selody (2001), a good monetary policy instrument must be under the direct or close control of the central bank.

9 M2++ includes mutual funds, whose importance increased continuously in consumer portfolios over the 1990s, and which are relatively liquid. Using a broad aggregate like M2++ in the model avoids interpreting a precautionary portfolio adjustment from mutual funds to money as inflationary. Moreover, Longworth (2003) finds that, since 1992, both core inflation and M2++ have been remarkably stable.

Table 1

Cointegration tests[1]

L-max	Trace	H_0: r =	L-max (0.10)	Trace (0.10)
63.12	151.48	0	40.2	104.4
46.36	88.36	1	34.1	76.9
35.84	42.00	2	28.3	54.8
10.39	16.17	3	22.2	35.9
2.97	5.78	4	15.9	20.8
2.81	2.81	5	9.5	9.5

[1] The critical values corrected for the presence of one exogenous variable are taken from Table T.3 in Pesaran et al (2000).

To discriminate between our cointegration tests, we looked at the t-values of the α coefficients for the third vector, as suggested in Hendry and Juselius (2001); when these are small, say less than 3.0, then one would not lose greatly by excluding that vector as a cointegration relation in the model. Given that many of these t-values are greater than 3.0 for all three vectors and that our theoretical model also suggests three vectors, we proceed under the assumption that there are three cointegrating vectors in our model.

The Johansen (1992) procedure allows us to identify the number of cointegrating vectors. However, in the case of existence of multiple cointegrating vectors, an interesting problem arises: α and β are only determined up to the space spanned by them. Thus for any non-singular matrix ζ conformable by-product:

$$\Pi = \alpha\beta' = \alpha\zeta\zeta^{-1}\beta'$$

In other words, β and $\beta'\zeta$ are two observationally equivalent bases of the cointegrating space. The obvious implication is that before solving such an identification problem, no meaningful economic interpretation of coefficients in cointegrating space can be proposed. The solution is imposing a sufficient number of restrictions on parameters such that the matrix satisfying such restrictions in the cointegration space is unique. Such a criterion is derived in Johansen (1992) and discussed in Hamilton (1994). Our restrictions are based on Blanchard's model and suggest more than a sufficient number of constraints on the cointegration space. The overidentification restrictions can therefore be tested. The results are in Table 2.

Table 2

Testing restrictions on the cointegrating vectors[1]

The LR test, $\chi^2(10) = 7.02$, p-value = 0.72

inf	*y*	*onr*	*m*	*sm*	*lr*	y^{US}
2.41 (0.27)	–1.18 (0.08)	2.41 (0.27)	1	0	0	0
0	0	1	0	0	–1	0
0	–1	0	0	1	0	0

[1] Standard errors are given within parentheses.

The restricted core model is strongly accepted with a p-value of 0.72. These results are consistent with the theoretical foundations presented in Section 2. The first cointegrating relation corresponds to the money market equilibrium, and the second to an approximation of the pure expectations hypothesis based on an arbitrage relation between short- and long-term bonds, while the third relation links real activity with the real stock market. The coefficients of the cointegrating relation cannot usually be interpreted as elasticities even if the variables are in logs, since a shock to one variable implies a shock to all variables in the long run. Hence the coefficients do not in general allow for a ceteris paribus interpretation (see Lutkepohl (1994)). Interpreting the coefficients in the first cointegrating relation is thus meaningless. However, given that the last two cointegrating relations involve only two variables, we do not need the ceteris paribus interpretation. The second long-run relation specifies that a permanent 1% increase in the short-term interest rate is associated with the equivalent increase in the long-run interest rate. The third cointegrating relation suggests that a 1% permanent increase in output (or a 1% increase in potential output) is associated with a permanent 1% increase in the stock market. Interestingly, this last relation also implies that transitory changes in real output can only lead to transitory changes in the level of the stock market.

The economy is in a long-run equilibrium when those three cointegrating relationships are respected, that is, when there is no gap between money, output, inflation and the overnight rate (or no money gap), the overnight rate is equal to the long rate (no interest rate gap), and the stock market level is consistent with potential output (no stock market gap).

Graph 1 illustrates the money gap[10] over the sample period. The two surges in inflation, in 1981 and 1991, were preceded by an increasing money gap around two years before. It is also interesting to note that since the Bank of Canada adopted an explicit inflation target in 1991, the money gap has been much more stable, deviating only slightly from equilibrium and for short periods of time in 1995 and 2000. This is in line with the results in Longworth (2003), who reports that, since 1992, both core inflation and M2++ have been remarkably stable.

The interest rate gap is defined as the yield spread (the long minus the overnight rate), well known as a good monetary policy stance measure. With this definition of the interest rate gap, the short rate is at its neutral level, or at its long-run equilibrium value, when it is equal to the long rate. According to this definition, the Bank of Canada was restrictive at the end of the 1980s to achieve the following disinflation and was accommodative for most of the 1990s except for a short period in 1999-2001. The overnight rate was back below equilibrium at the end of 2002 by almost 2%.

The stock market gap in Graph 2 illustrates periods of "mis-valuation" of the stock market.[11] Our results show that the stock market led the 1981 and 1991 recessions and became strongly undervalued (close to 40%) after the 1981 recession. It became relatively less depressed after the 1991 recession (around 20%), but took longer to recover; the market got back to its fair value only in 1994. Graph 2 also shows that the stock market was about 20% overvalued before the 1987 crash and undershot by about 10% afterwards. The market was overvalued for most of the 1996-2000 period, except for the strong correction following the Asian crisis in 1998. By far the most significant departure from equilibrium happened at the beginning of 2000 when the stock market appeared to have been close to 60% higher than what was justified by "fundamentals". Finally, the bubble burst and the market overreacted again. Graph 2 suggests it was about 10% undervalued at the end of 2002. These results are in line with Dupuis and Tessier (2003), who estimate a three-variables VECM linking the US stock market to dividends and the long-term interest rate.

4. Econometric formulation of the core model

The three long-run equilibrium relationships can be written in the following form:

$$m_t = c_{11} + c_{12}y_t + c_{13}onr_t + c_{13}inf_t + \xi_{1t+1} \quad (6)$$

$$lr_t = c_{21} - onr_t + \xi_{2t+1} \quad (7)$$

[10] The gaps in this section are simply defined as the error correction term from the cointegrating relations. Gaps based on permanent components of the variables will be presented in Section 5.

[11] Note that the permanent components of the variables have yet to be identified before we can tell if a positive error correction term is due to the stock market being too high or output too low (or both). This is done below.

$$sm_t = c_{31} - y_t + \xi_{3t+1} \tag{8}$$

The three long-run relations of the core model, equations (6), (7) and (8), can be written as

$$\xi_t = \beta' z_{t-1} - c_0 \tag{9}$$

where $z_t = (inf_t, y_t, onr_t, m_t, sm_t, lr_t, y_t^{US})'$, $c_0 = (c_{11}, c_{21}, c_{31})'$, $\xi_t = (\xi_{1t}, \xi_{2t}, \xi_{3t})'$, and

$$\beta' = \begin{bmatrix} -c_{13} & -c_{12} & -c_{13} & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \end{bmatrix} \tag{10}$$

Let $x_t = (inf_t, y_t, onr_t, m_t, sm_t, lr_t)'$. We base our analysis on the following conditional error correction model:

$$\Delta x_t = a - \alpha \xi_t + \sum_{i=1}^{s-1} \Gamma_i \Delta z_{t-i} + b \Delta y_t^{US} + u_{xt} \tag{11}$$

where a is a 6×1 vector of fixed intercepts, α is a 6×3 matrix of error correction coefficients, b is a 6×1 vector representing the impact effects of changes in US output on Δx_t, and u_{xt} is a 6×1 vector of disturbances assumed to be $IID(0, \Sigma_x)$, with Σ_x being a positive definite matrix.

From equations (9), (10) and (11), we have

$$\Delta x_t = a_y + \alpha_y c_0 - \alpha_y \beta' z_{t-1} + \sum_{i=1}^{s-1} \Gamma_i \Delta z_{t-i} + b_y \Delta y_t^{US} + u_{yt} \tag{12}$$

where $\beta' z_{t-1}$ is a 3×1 vector of error correction terms. This specification implies the economic theory's long-run predictions by construction. The estimations of the parameters in equation (12) are obtained by using the estimation procedure of vector error correction models with exogenous $I(1)$ variables (Pesaran et al (2000)).

5. Shock analysis

The impact of a change in US industrial production

The response functions to a permanent increase of 1% in US industrial production are shown in Graph 4. Small inflation pressures are generated as output is boosted by almost 0.2% on impact. Interest rates are increased by around 25 basis points to keep demand in line with short-run supply. The Canadian stock market is negatively affected by the higher interest rate. It nevertheless increases by 0.12% in the long run, in line with the permanent increase in output.[12] Broad aggregate money is negatively affected in the short run by the slight increases in inflation and real interest rates. Only output is significantly affected in the long run.

Identification of the permanent shocks

Given the presence of three cointegrating vectors and six endogenous variables, there are three stochastic trends or permanent shocks to be identified.[13] The first permanent shock, $\varepsilon_{\pi t}$, labelled an inflation shock, is the only shock having a permanent impact on inflation. According to the "monetarist" view, the long-run money growth and inflation rate are ultimately set exogenously by monetary authorities. So the inflation shock relates to central bank monetary policy. A positive inflation shock

[12] US industrial production represents about 15% of US total GDP. Under the assumption that a permanent increase of 1% in US industrial production translates into an increase of 0.15% in US total GDP, our results suggest that a 0.15% increase in US GDP is associated with an increase of about 0.12% in Canadian GDP.

[13] Details on identification in the presence of exogenous variables will be published in a future version of this paper.

reflects the central bank's decision to permanently increase the inflation rate. Hence, the structural inflation shock is identified by assuming that the long-run system has the following recursive structure:

$$\lim_{s\to\infty}\begin{bmatrix} inf_{t+s} \\ y_{t+s} \\ onr_{t+s} \\ m_{t+s} \\ sm_{t+s} \\ lr_{t+s} \end{bmatrix} = \begin{bmatrix} \tau_{11} & 0 & 0 \\ \tau_{21} & \tau_{22} & 0 \\ \tau_{31} & \tau_{32} & \tau_{33} \\ \tau_{41} & \tau_{42} & \tau_{43} \\ \tau_{51} & \tau_{52} & \tau_{53} \\ \tau_{61} & \tau_{62} & \tau_{63} \end{bmatrix} \cdot \begin{bmatrix} \varepsilon_{\pi t} \\ \varepsilon_{yt} \\ \varepsilon_{dt} \end{bmatrix}$$

Note that τ_{ij} is the long-run response of the *i*th endogenous variable to the *j* element in the vector of structural disturbances ε_t. The restrictions $\tau_{12} = 0$ and $\tau_{13} = 0$ mean that only an inflation shock, $\varepsilon_{\pi t}$, affects the long-run level of inflation. The mainstream view would predict that the decision to change inflation permanently has no permanent impact on real variables and thus that $[\tau_{21}\ \tau_{41}\ \tau_{51}] = 0$. However, economic theory provides no clear-cut predictions on that question. In several theoretical models, the superneutrality result due to Sidrausky (1967) breaks down as inflation can have either positive or negative effects on real variables such as consumption and investment, depending on the exact assumptions concerning preferences. Additionally, in these models the real interest rate may or may not be independent of inflation in the long run (see Orphanides and Solow (1990) for a survey). Some recent empirical results (see, for example, Rapach (2003) and Gauthier et al (2003)) find support for the Mundell-Tobin effect, suggesting that an unexpected increase in inflation has a permanent negative impact on the real interest rate. We let the data talk on this point by leaving unconstrained the parameters in $[\tau_{21}\ \tau_{31}\ \tau_{41}\ \tau_{51}\ \tau_{61}]$.

Most theoretical models define supply shocks as being governed by technology innovations determining the technical capacity of the economy. We thus identify a supply shock as a shock allowed to have a permanent effect on output but not on inflation. The long-run effects on all the other real variables are left unconstrained. Note that all shocks are allowed to impact all the variables in the short run. In particular, a supply shock is expected to decrease inflation in the short run.

The third structural shock is a shock having no permanent impact either on output or on inflation. This shock is labelled a demand shock.

The inflation shock

A positive inflation shock reflects the central bank's decision to permanently increase the inflation rate.[14] Given the instrument used by the central bank, this can only be achieved by decreasing the overnight rate. Graph 5 shows that our results are consistent with this view. To achieve a typical unexpected inflation increase of around 0.3% in the long run, the central bank has to decrease the overnight rate by about 25 basis points. Given the expectations hypothesis of the term structure in our core model, the long rate is persistently depressed as well. The bank's intervention leads to a small output stimulus in the short run. The shock also significantly hurts the stock market and decreases *real* broad aggregate money in the short run.

The permanent significant negative effect of inflation on interest rates may be explained through the Mundell effect: an *unexpected* increase in inflation decreases real wealth, which increases savings. Real interest rates must then fall to restore goods market equilibrium. Our results are also consistent with the focus on stabilising output in the 1970s and the beginning of the 1980s even at the cost of higher inflation. Furthermore, they are in line with the need to persistently increase the interest rate in disinflation periods and in the first years of inflation targeting in order to gain credibility. Rapach (2003)

14 Of course, such a shock can always be reversed by a negative inflation shock of the same size, if the central bank decides to do so.

also finds that an unexpected permanent increase in inflation is associated with permanently lower long-run real interest rates in every industrialised country in a sample of 14, including Canada.[15]

When inflation is forecast to deviate permanently from the actual target of 2%, the historical estimated reaction function (the equation for the overnight rate) may be adjusted using the estimated impact over time of the typical permanent inflation shock in such a way as to eliminate the expected long-run deviation from target.

The supply shock

The typical supply shock increases the productive capacity of the economy by around 0.9% in the long run. Inflation is pushed downwards in the short run as production costs are decreased (Graph 6) but goes back to its initial level in the long run. The central bank has, over the sample, accommodated the shock by decreasing interest rates to eliminate the excess supply in the goods market and bring inflation back to target.[16] The stock market leads output and overshoots somewhat. Broad money is higher in the short run because of the accommodative stance of monetary policy and remains higher in the long run because of both higher money demand for transaction purposes and higher real value of the stock market.

A demand shock[17]

The demand shock increases inflation, output and the stock market in the short run. Short and long interest rates increase in the short run as expected. This can be seen as the result of a standard textbook open market operation with a disinflationary objective. When inflation and output turn out to be higher than expected, an inflation targeting central bank has to increase interest rates. It is interesting to note that since a demand shock has no permanent impact on output, the significant stock market surge in the first months following the shock slowly dissipates as investors realise that higher profits cannot be sustained without a permanent increase in productivity.

The permanent positive impact on the overnight rate implies that the so-called demand shock induces, on average, a higher equilibrium interest rate. This, again, is consistent with the need to persistently increase the interest rate in disinflation periods and in the first years of inflation targeting in order to gain credibility. Furthermore, as predicted by the long-run theory of growth models, any shock that persistently lowers the share of product going into investment is associated with higher real interest rates in the long run.[18] For example, fiscal shocks crowding out investment persistently will be associated with persistently higher interest rates.

Output gap

An output gap is easily obtained from our model as the difference between actual output and the historical contribution of permanent shocks to output (determining potential output). Potential output and the output gap are plotted in Graphs 8 and 9 respectively. According to these results, the Canadian economy was in excess demand before both the 1982 and the 1991 recessions and was in excess supply for most of the 1990s. The gap was closed at the end of 1999 and the economy turned

---

[15] Note that a permanent inflation shock represents an *unexpected* persistent deviation of inflation from its deterministic trend. This source of increase in inflation is associated in the long run with a decrease in interest rates. That, of course, does not mean that *expected* changes in inflation have the same effect on interest rates.

[16] In some SDGE models with adjustment costs on capital (see Neiss and Nelson (2001, p 23) for an example), productivity shocks would decrease the *neutral* rate in the short run. This provides further incentives to decrease the actual interest rate after a productivity shock.

[17] Other demand shocks having only transitory effects may also be identified.

[18] King et al (1991) estimate a significant cointegration relationship negatively linking the ratio of investment to output and the real interest rate in the United States and identify what they call a "real interest rate shock" with long-run properties very similar to our "demand" shock. They also identify what they call a "balanced-growth" shock, which is very similar to our supply shock increasing output permanently while leaving the ratios of investment and consumption to output as well as the real interest rate and inflation unchanged in the long run.

to excess demand for the following two years. The economy was back in excess supply (though close to zero) at the end of 2002. What may be more surprising is the period over which supply shocks contributed to increasing output permanently. Graph 8 suggests that it started around 1985 and lasted until 1996, the year Chairman Greenspan first talked of irrational exuberance. From 1996 until the end of 2000 and the strong stock market correction, the economy was demand-driven and potential would have been growing at a rate lower then the deterministic rate.[19] This result, in line with Dueker and Nelson (2002) and the latest economic developments, casts some doubts on the purported "new economy" in the second half of the 1990s.

6. Conclusion

We have estimated a small monthly VECM to study the interactions between the real and financial sectors of the Canadian economy. To take into account the high degree of economic integration between Canada and the United States, the US industrial production index has been included as an exogenous variable. Identification of permanent shocks in a VECM with exogenous variables represents a technical contribution to the literature.

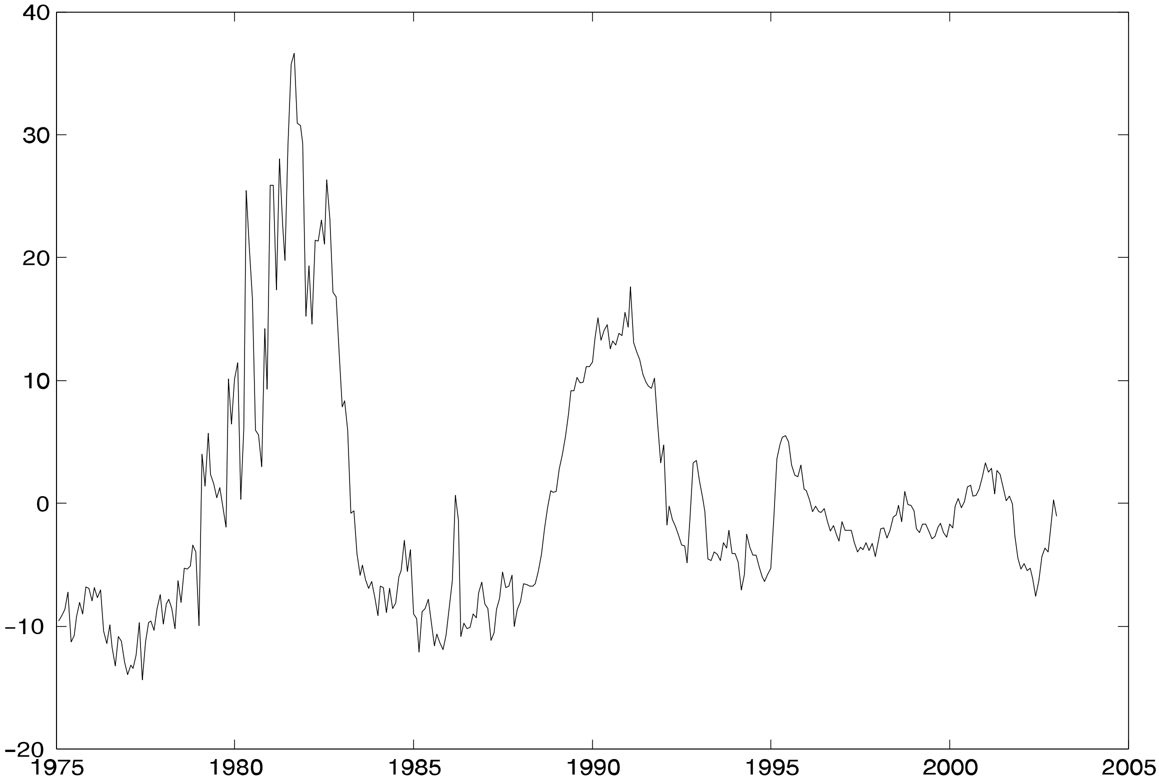
Our principal results are: (1) the identification of a long-run relation between the stock market and real output which allows the identification of a supply shock as the only shock permanently affecting the stock market and a demand shock leading to significant transitory stock market overvaluation; (2) the money gap defined as the error correction term from the first cointegrating relation has been much more stable since the adoption of inflation targets in Canada.

The next step in this project is to study the behaviour of a reaction function that would reverse any identified nominal shock causing inflation to persistently deviate from the target. The model could also be used to build a financial condition index for Canada using the stock market and money gaps from the core model together with the deviation of the actual real interest rate from the neutral interest rate recommended by the proposed reaction function. This index could also possibly be completed with the deviation of the Canadian exchange rate from equilibrium provided in Gauthier and Tessier (2002) and tested against those proposed in Gauthier et al (2003). This is left for future research.

[19] It should be noted, however, that a shift in the deterministic trend in output is estimated in 1993. Hence, the growth of potential in the second half of the 1990s is lower compared with a relatively higher growth in trend. Depending on our judgment on the source of this shift, the story can be completely different. If the higher deterministic output growth is attributed to supply shocks, then potential output would have increased continuously in the 1990s and the Canadian economy would currently be in considerable excess supply. Nevertheless, given the deterministic nature of this shift and the recent economic developments, we proceed under the assumption that this change in trend should be considered as demand-driven, implying that potential output and the output gap are well approximated by Graphs 8 and 9. The fact that potential has been below the higher growth trend for the last seven years is also an indication that the higher trend should be seen as transitory.

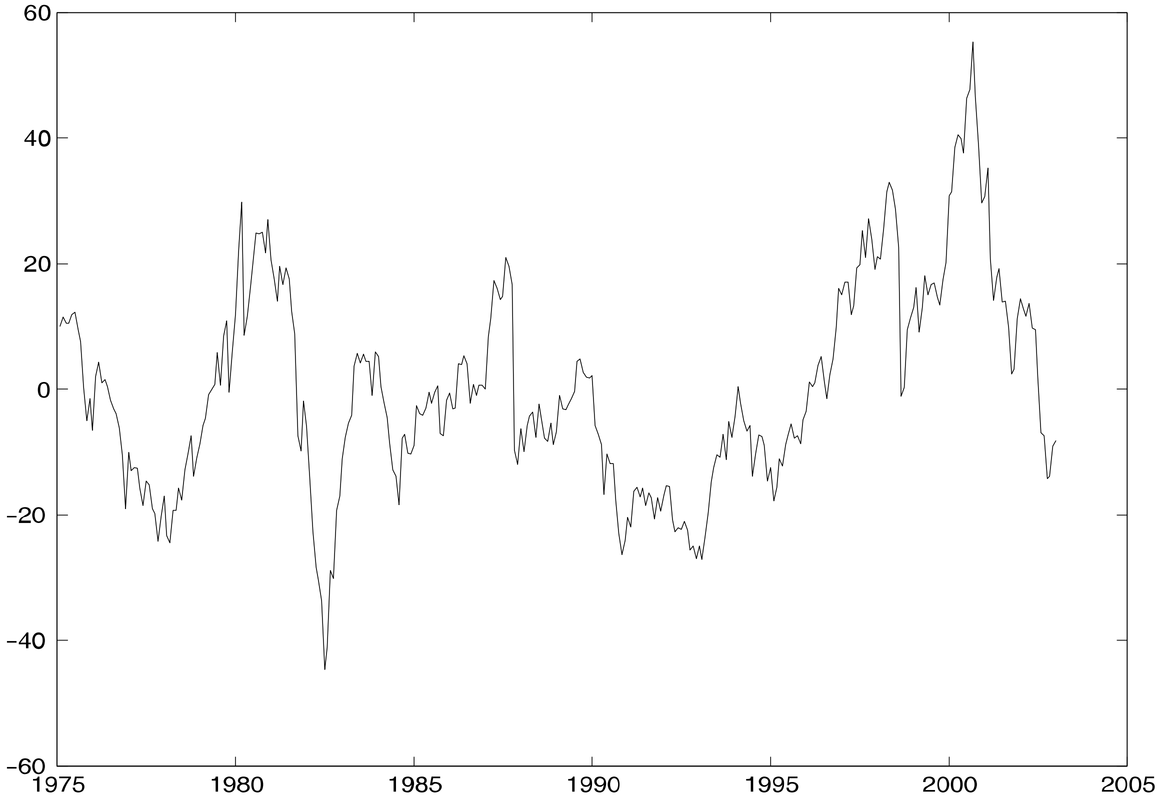
Graph 1

Money gap



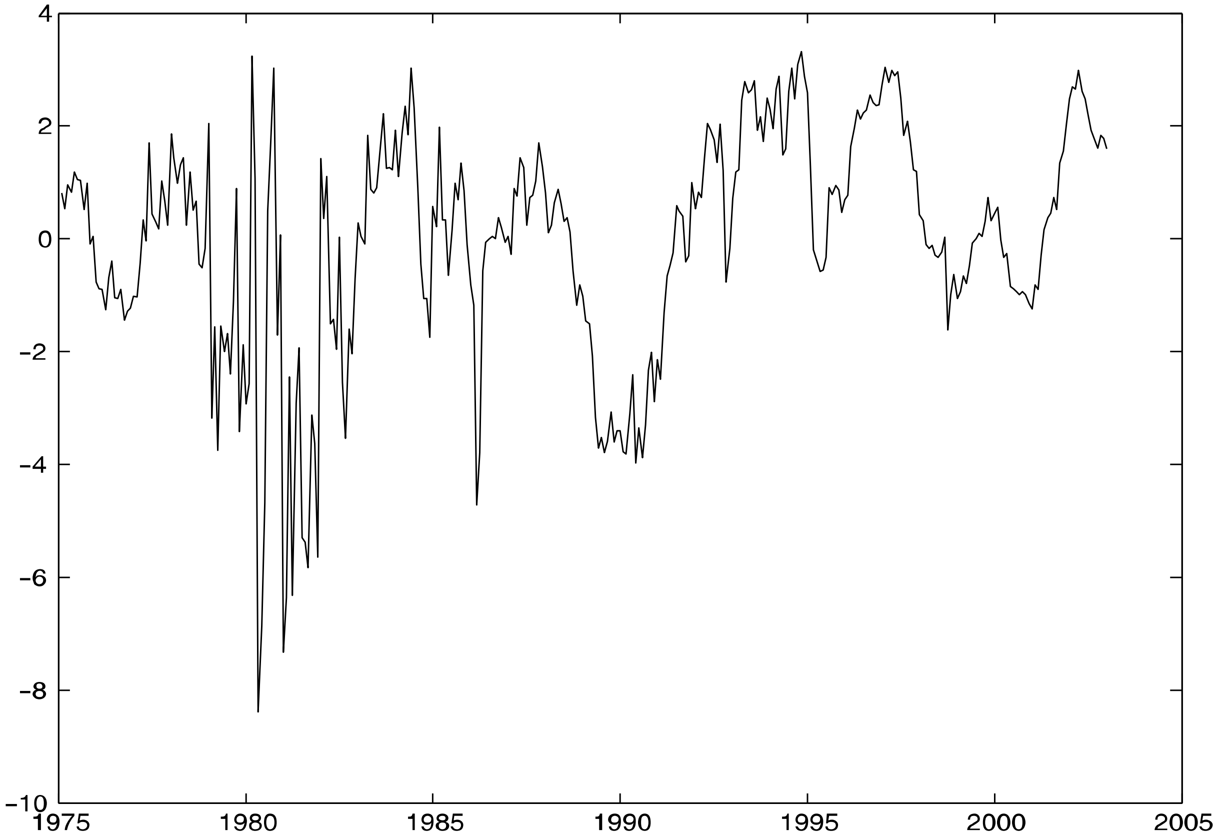
Graph 2

Stock market gap



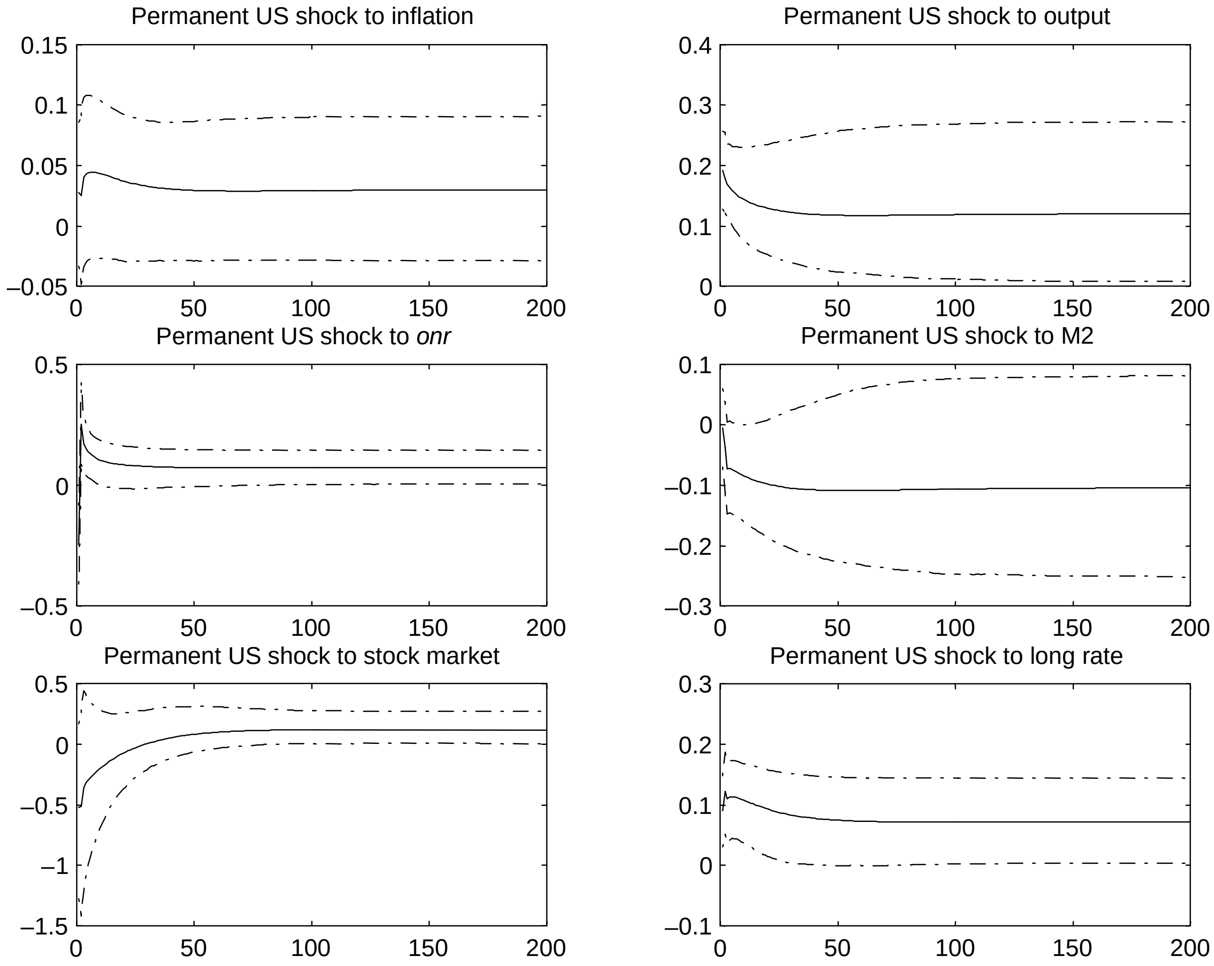
Graph 3

Interest rate gap



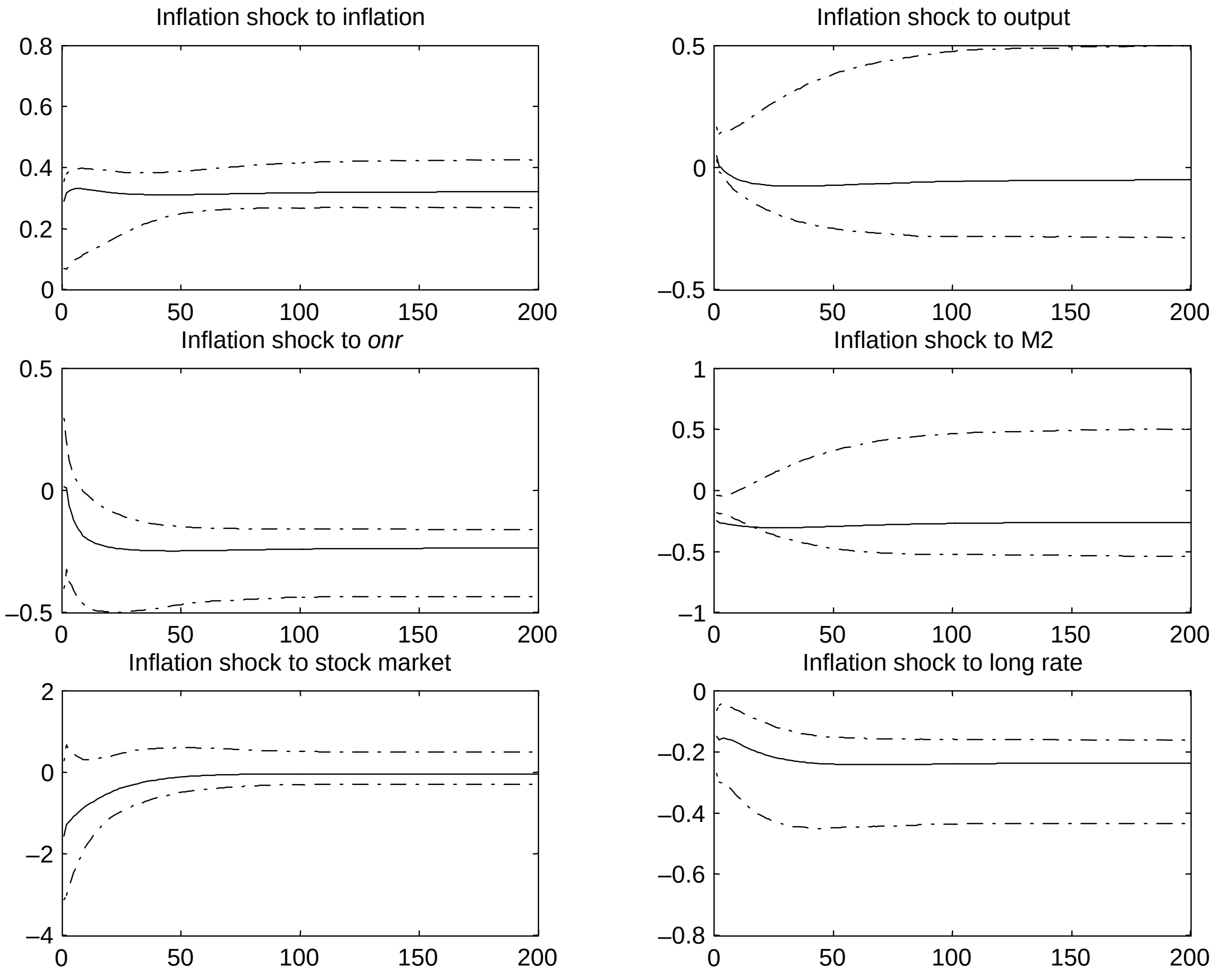
Graph 4

Responses to a permanent increase in US industrial production



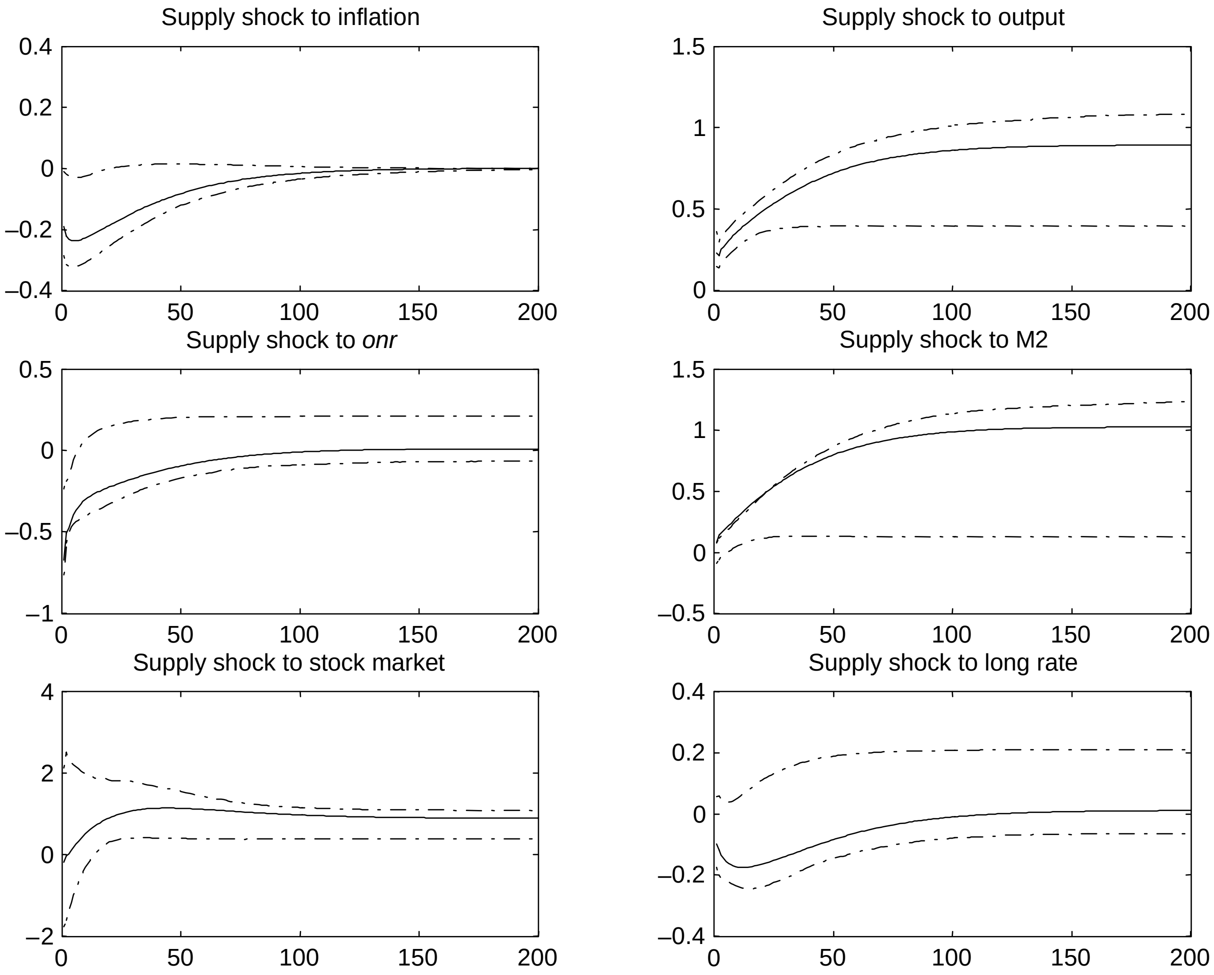
Graph 5

Impulse responses to an inflation shock



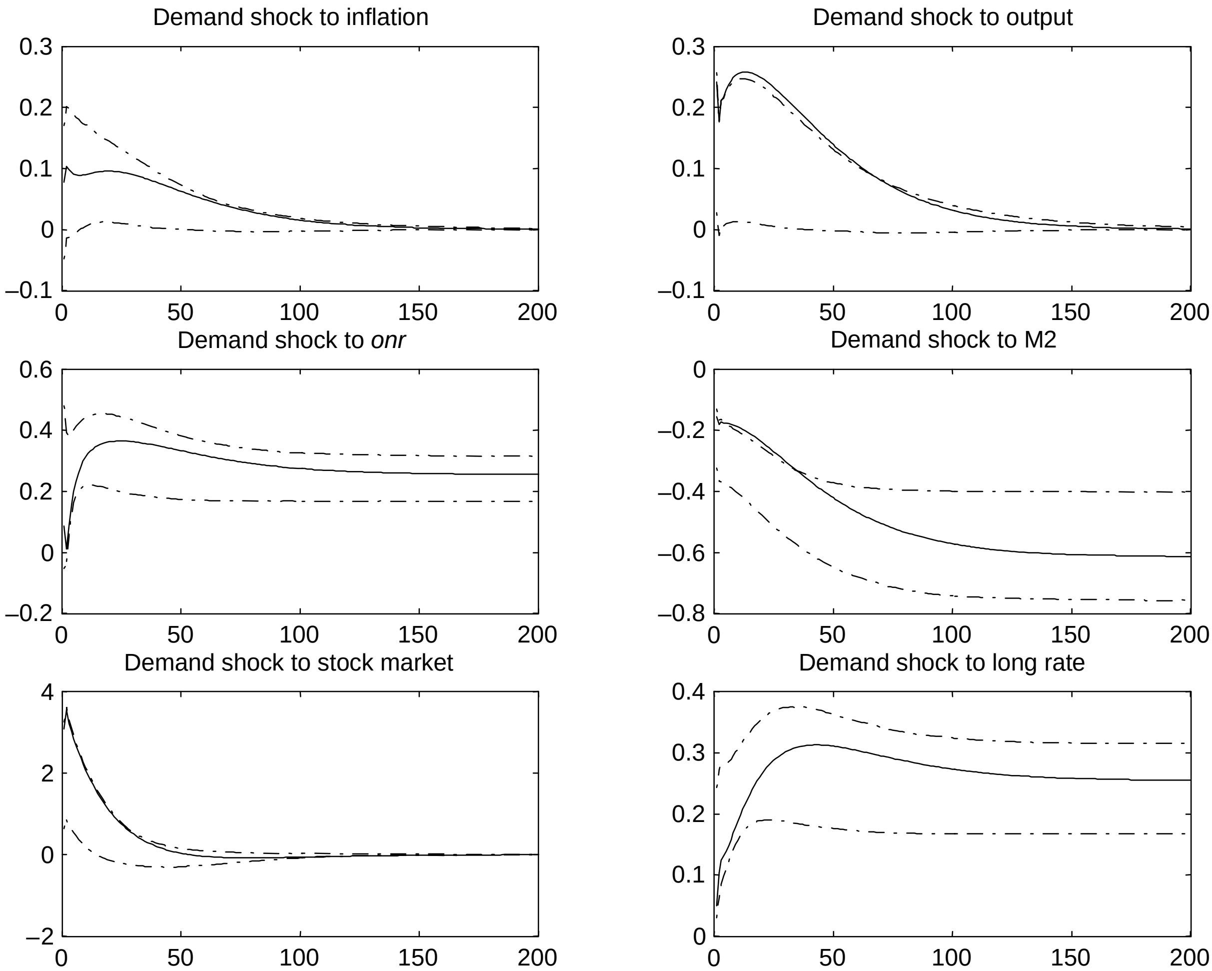
Graph 6

Impulse responses to a supply shock



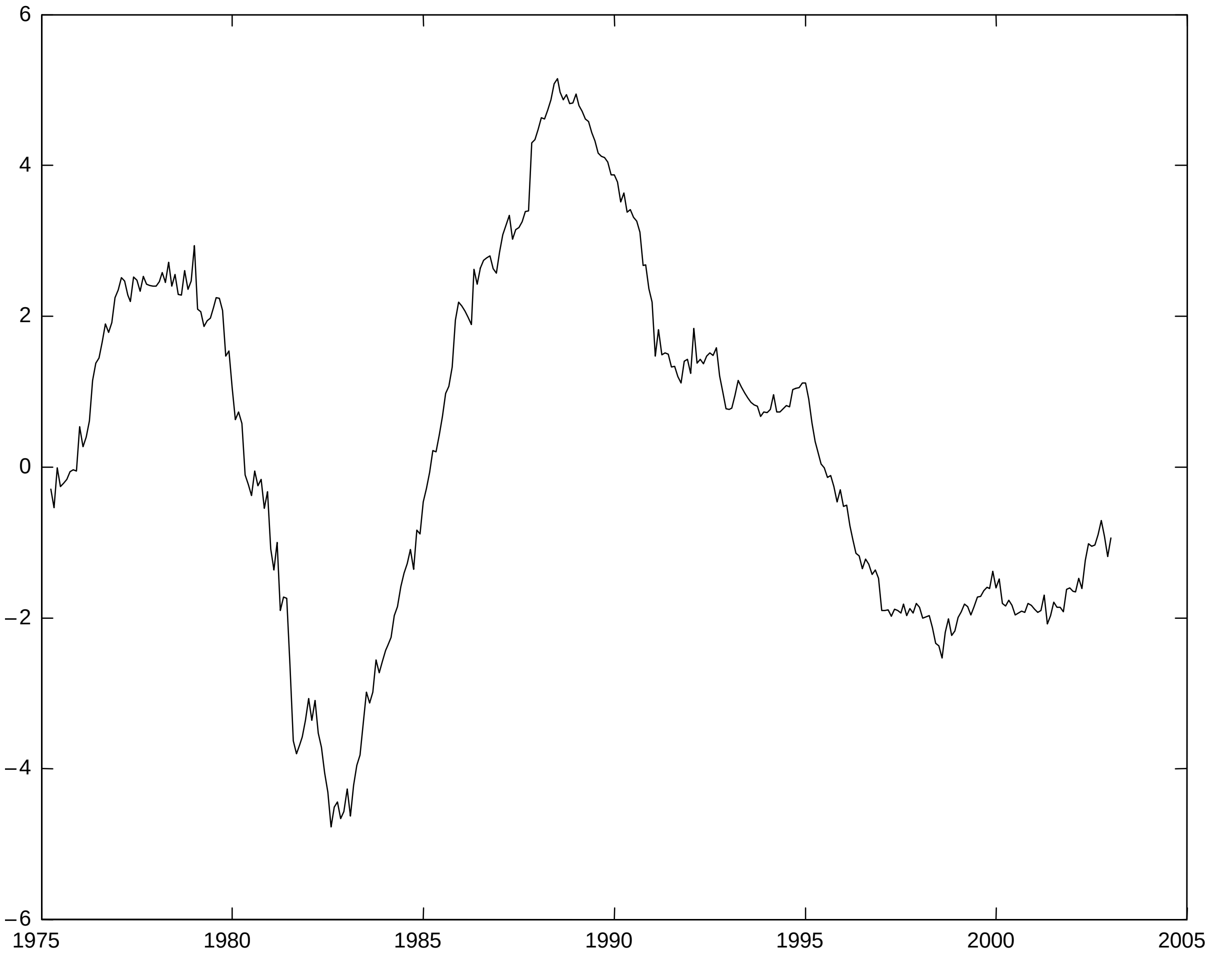
Graph 7

Impulse response to a demand shock



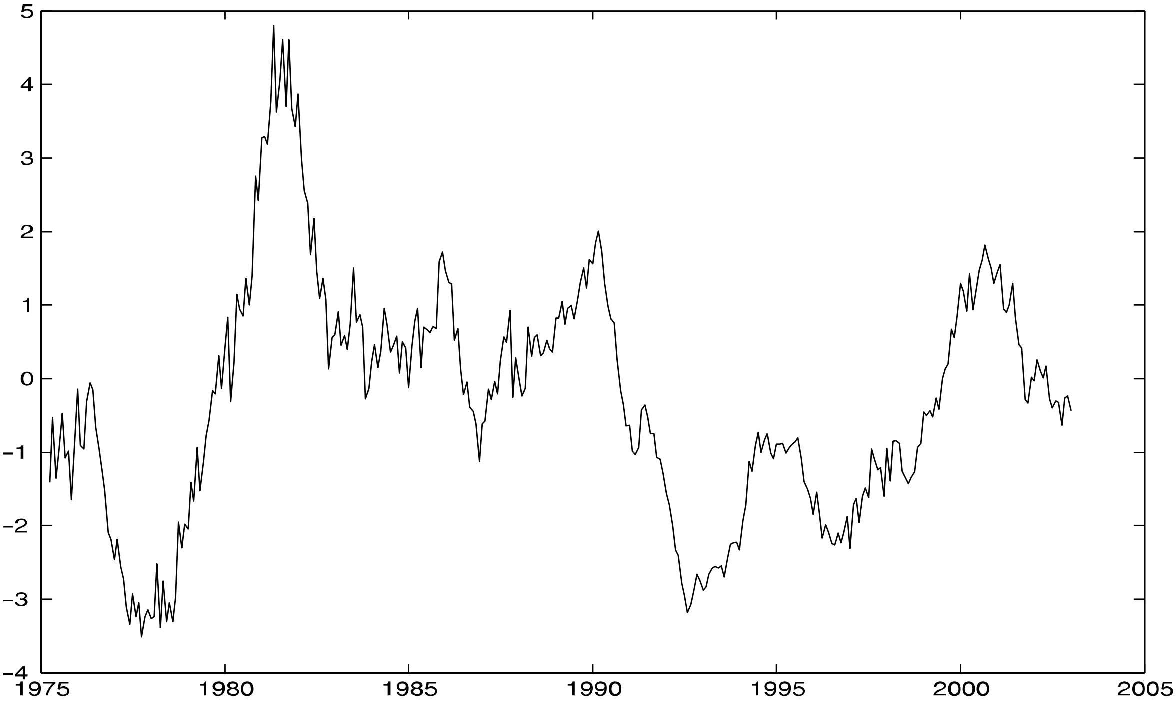
Graph 8

Potential output



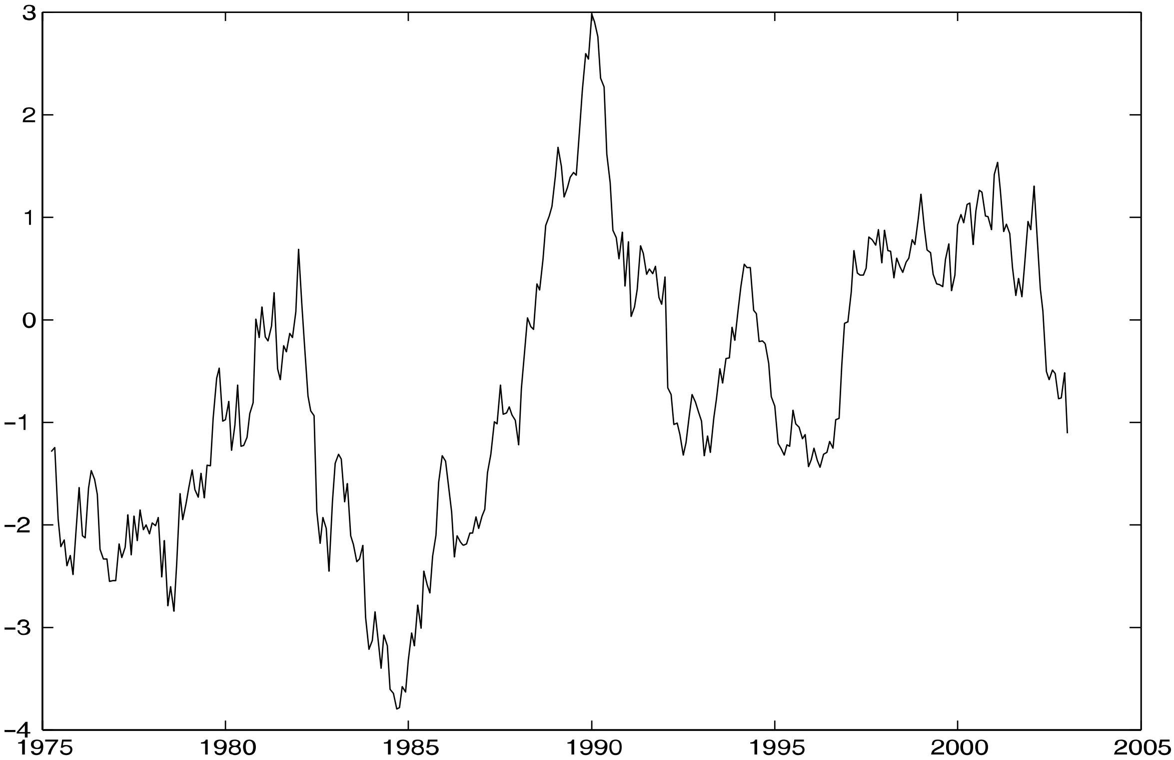
Graph 9

Output gap



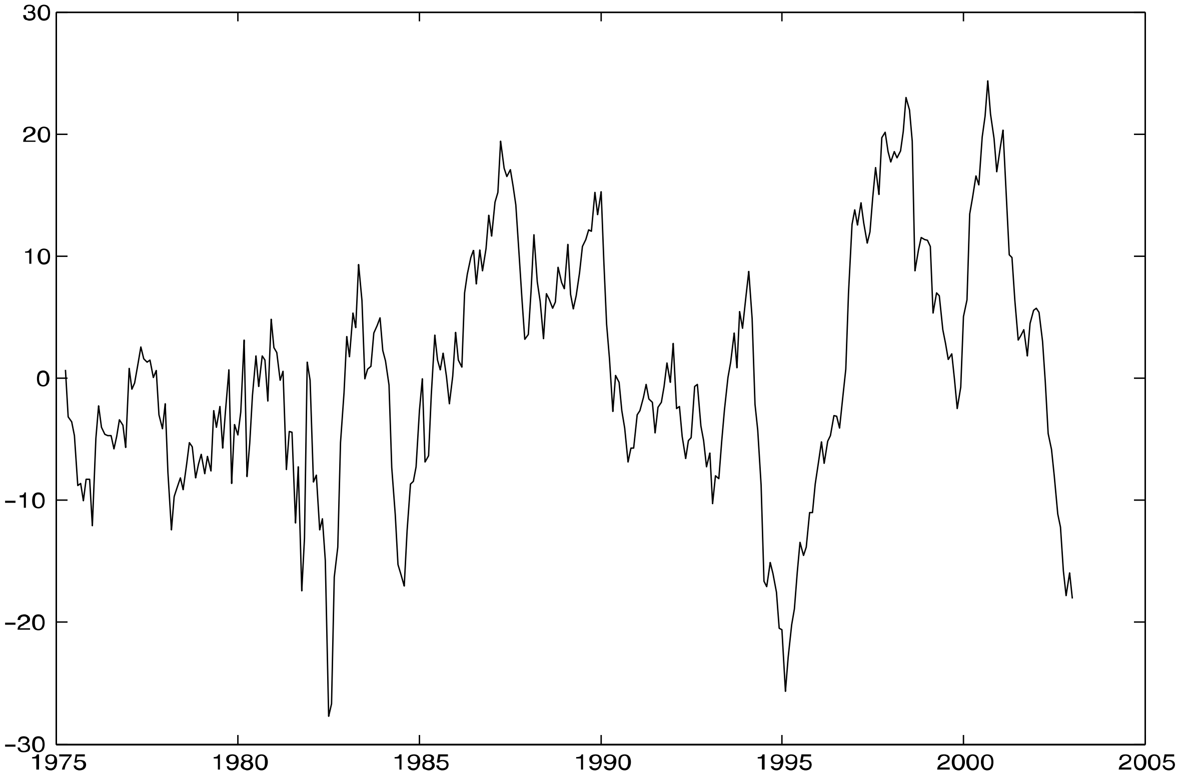
Graph 10

Money gap (based on permanent components)



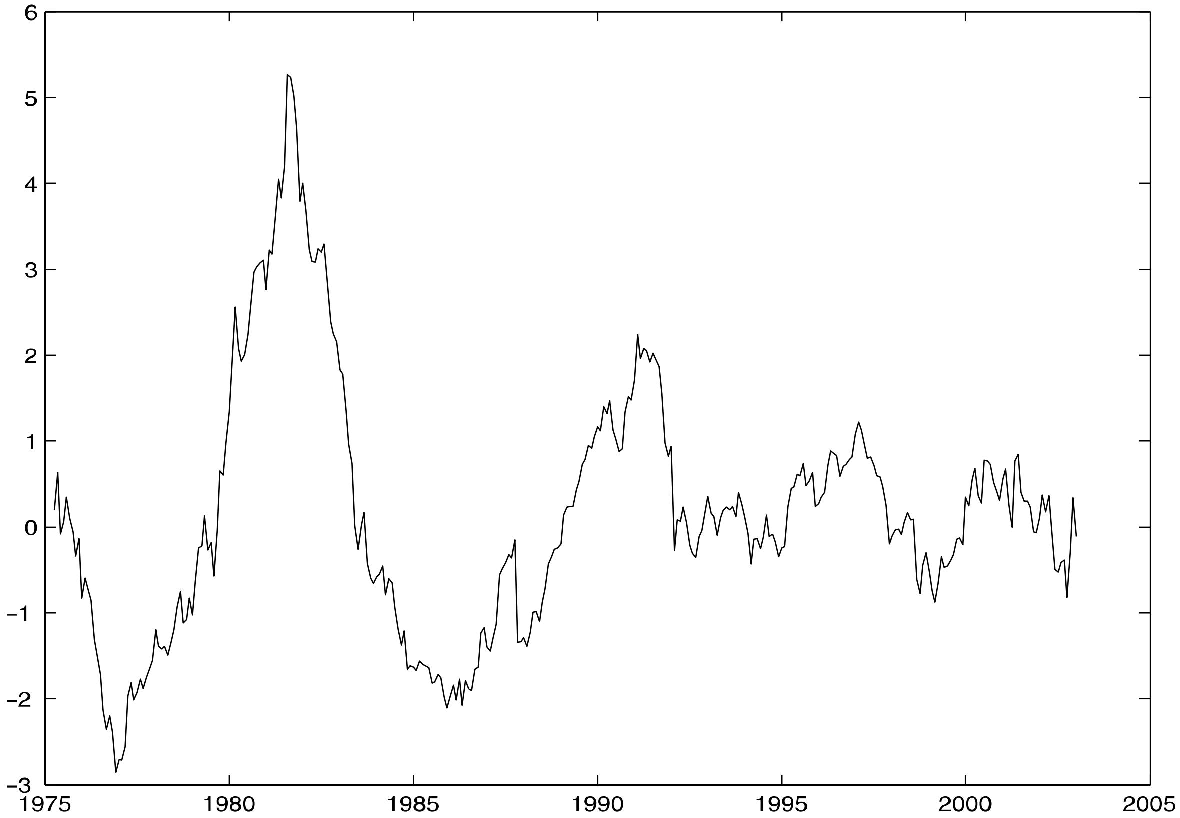
Graph 11

Stock market gap (based on permanent components)



Graph 12

Transitory component of inflation



References

Ahmed, S and J Rogers (2000): "Inflation and the great ratios: long-term evidence from the US", *Journal of Monetary Economics*, 45, pp 3-35.

Angeloni, I, G Coenen and F Smets (2003): "Persistence, the transmission mechanism and robust monetary policy", *European Central Bank Working Paper Series*, 250.

Blanchard, O J (1981): "Output, the stock market, and interest rates", *American Economic Review*, 71, pp 132-43.

Blanchard, O J and D Quah (1989): "The dynamic effects of aggregate demand and supply disturbances", *American Economic Review*, 79, pp 1146-64.

Cassola, N and C Morana (2002): "Monetary policy and the stock market in the euro area", *European Central Bank Working Paper Series*, 119.

Coenen, G (2002): *Inflation persistence and robust monetary policy design*, European Central Bank, mimeo.

Cogley, T and T Sargent (2001): "Evolving post-World War II inflation dynamics", *NBER Macroeconomics Annual*, 16.

Crowder, J, D L Hoffman and R H Rasche (1999): "Identification, long-run relations, and fundamental innovations in a simple cointegrated system", *The Review of Economics and Statistics*, vol 81, pp 109-21.

Dhar, S, D Pain and R Thomas (2000): "A small structural empirical model of the UK monetary transmission mechanism", *Bank of England Working Paper Series*.

Dueker, M J and C R Nelson (2002): "Business cycle detrending of macroeconomic data via a latent business cycle index", *Federal Reserve Bank of St Louis Working Paper* 2002-025A.

Dupuis, D and D Tessier (2003): "The US stock market and fundamentals: an historical decomposition", *Bank of Canada Working Papers*, 2003-20.

Garratt, A, K Lee, M H Pesaran and Y Shin (2003): "A long-run structural macroeconometric model of the UK", *The Economic Journal*, 113, pp 412-55.

Gauthier, C, C Graham and Y Liu (2003): *Financial condition indexes for Canada*, Bank of Canada, mimeo.

Gauthier, C, F Pelgrin and D Schweisguth (2003): *The long-run relationship between inflation and real stock prices: is there neutrality?*, Bank of Canada, mimeo.

Gauthier, C and D Tessier, (2002): "Supply shocks and real exchange rate dynamics: Canadian evidence", *Bank of Canada Working Papers,* 2002-31.

Goodfriend, M (1993): "Interest rate policy and the inflation scare problem: 1979-1992", *Federal Reserve Bank of Richmond Economic Quarterly*, Winter, pp 1-24.

Hamilton, James D (1994): *Time series analysis*, Princeton University Press, Princeton.

Hendry, David F and Katarina Juselius (2001): "Explaining cointegration analysis: part II", *Energy Journal*, vol 22, issue 1, pp 75-120.

Jacobson, T, P Jansson, A Vredin and A Warne (2001): "Monetary policy analysis and inflation targeting in a small open economy: a VAR approach", *Journal of Applied Econometrics*, 16, pp 487-520.

Johansen, S (1992): "Cointegration in partial systems and the efficiency of single-equation analysis", *Journal of Econometrics*, 52, pp 389-402.

King, R G, C I Plosser, J H Stock and M W Watson (1991): "Stochastic trends and economic fluctuations", *American Economic Review*, 81, pp 819-40.

Longworth, D (2003): "Money in the Bank (of Canada)", *Bank of Canada Technical Reports*, 93.

Lucas, R E (1976): "Econometric policy evaluation: a critique", in K Brunner and A Meltzer (eds), *The Phillips curve and labor markets*, North-Holland, Amsterdam.

Lutkepohl, H (1991): *Introduction to multiple time series analysis*, Springer-Verlag.

——— (1994): "Interpretations of cointegrating relations", *Econometric Reviews*, 13, pp 391-4.

Lutkepohl, H and H-E Reimers (1992): "Impulse response analysis of cointegrated systems", *Journal of Economic Dynamics and Control*, 12, pp 53-78.

Neiss, K S and E Nelson (2001): "The real interest rate gap as an inflation indicator", *Bank of England Working Paper Series*.

Orphanides, Athanasios and Robert M Solow (1990): "Money, inflation and growth", *Handbook of monetary economics*, vol 1, North-Holland, Amsterdam, Oxford and Tokyo, distributed in the United States and Canada by Elsevier Science, New York, pp 223-61.

Pesaran, M H, Y Shin and R J Smith (2000): "Structural analysis of vector error correction models with exogenous *I*(1) variables", *Journal of Econometrics*, 97, pp 293-343.

Rapach, D E (2003): "International evidence on the long-run impact of inflation", *Journal of Money, Credit and Banking*, 35, pp 23-48.

Selody, J (2001): *Principles for building models of the monetary policy transmission mechanism*, Bank of Canada, mimeo.

Sidrausky (1967): "Inflation and economic growth", *Journal of Political Economy*, 75, pp 796-810.

Sims, C (1980): "Macroeconomics and reality", *Econometrica*, 48, pp 1-48.

Stock, J (2001): "Comment", *NBER Macroeconomics Annual*, 16.

Wickens, M R (1996): "Interpreting cointegrated vectors and common stochastic trends", *Journal of Econometrics*, 74, pp 255-71.

Wickens, M R and R Motto (2001): "Estimating shocks and impulse response functions", *Journal of Applied Econometrics*, 16, pp 371-87.

Yang, M (1998): "On identifying permanent and transitory shocks in VAR models", *Economics Letters*, 58, pp 171-5.

Interactions between business cycles, financial cycles and monetary policy: stylised facts[1]

Sanvi Avouyi-Dovi and Julien Matheron[2]

Introduction

The spectacular rise in asset prices up to 2000 in most developed countries has attracted a great deal of attention and reopened the debate over whether these prices should be targeted in monetary policy strategies. Some observers see asset price developments, in particular those of stock prices, as being inconsistent with developments in economic fundamentals, ie a speculative bubble. This interpretation carries with it a range of serious consequences arising from the bursting of this bubble: scarcity of financing opportunities, a general decline in investment, a fall in output, and finally a protracted contraction in real activity. Other observers believe that stock prices are likely to have an impact on goods and services prices and thus affect economic activity and inflation.

These theories are currently at the centre of the debate on whether asset prices should be taken into account in the conduct of monetary policy, ie as a target or as an instrument.[3] However, the empirical link between asset prices and economic activity on the one hand, and the relationship between economic activity and interest rates or between stock prices and interest rates on the other, are not established facts. This study therefore sets out to identify a number of stylised facts that characterise this link, using a statistical analysis of these data (economic activity indicators, stock prices and interest rates).

More specifically, we study the co-movements between stock market indices, real activity and interest rates over the business cycle. Assuming that there is no single definition of the business cycle, we adopt an agnostic approach in our methodology.

The traditional approach characterises the cycle as a series of phases of expansion and contraction. Formally, expansion phases are defined as the periods of time separating a trough from a peak; conversely, contraction phases correspond to periods separating a peak from a trough. In this respect, it is vital to define and accurately identify peaks and troughs.

Although this view of the cycle fell out of fashion after the 1970s, it has recently come back into focus thanks to a number of studies, in particular by Harding and Pagan (2002a,b), who proposed a simple method for analysing the concordance between macroeconomic variables.[4] By definition, the concordance index represents the average number of periods in which two variables (eg GDP and a stock market index) coincide at the same phase of the cycle.

The traditional approach defines the business cycle directly by analysing changes in the level of a variable, eg GDP. The modern approach (mentioned above), using the appropriate statistical filtering techniques, enables us to split a variable into two components, one cyclical or short-term, and the other structural or permanent. As its name suggests, the cyclical component has no trend and can be associated with the business cycle. Consequently, we can calculate the correlations between the cyclical components of two variables in order to study the degree of their co-movement (ie the similarity of their profile). However, we show that the structural component of a variable is driven by a

---

1 A non-technical version of this paper is published in the *Revue de la stabilité financière*, no 3, 2003. The views expressed herein are those of the authors and do not necessarily reflect those of the Bank of France.

2 Correspondence: DGEI-DEER, Centre de recherche (41-1391), 31 rue croix des petits champs, 750049 Paris Cedex 01. E-mail addresses: sanvi.avouyi-dovi@banque-France.fr and julien.matheron@banque-France.fr.

3 A large amount of theoretical literature has recently been published on this subject. See Bernanke and Gertler (2001), Bullard and Schalling (2002), Filardo (2000) and other references cited in these papers.

4 For a recent application on euro area data, see Artis et al (2003).

trend. Hence, to avoid spurious relationships, we study the growth rate of the structural components. We can also calculate the correlations between the growth rate of the structural components of two variables in order to study their co-movement.

As the notions of concordance and correlation do not have an identical scope, it is useful to use both of these tools when attempting to characterise the stylised facts relating to the business cycle.

The first part of this study is devoted to the empirical analysis of the concordance indicator; the second part starts off by describing changes in the variables studied (real activity, stock prices and interest rates) by separating the cyclical (or short-term) components from the structural (or long-term) components, and then compares the variables using the dynamic correlations of their corresponding components (ie cyclical/cyclical and structural/structural).

In both parts, we compare the results obtained on the business and stock market cycles to the monetary policies applied over the period studied: first, we analyse the behaviour of short-term interest rates over the phases of expansion and contraction of real activity and stock prices; and second, we calculate the correlations between the cyclical components of real activity, stock prices and interest rates on the one hand, and the correlations between the structural components of these variables on the other.

1. Concordance between business cycles and stock market cycles: an empirical analysis

As a concordance indicator, we use a descriptive statistic recently developed by Harding and Pagan (2002a,b) and implemented at the IMF by Cashin et al (1999) and McDermott and Scott (2000). Cashin et al applied this method to an analysis of the concordance of goods prices while McDermott and Scott used it to study the concordance of business cycles in major OECD countries.

The underlying method is based on studies by the National Bureau of Economic Research (NBER) and consists in dating the turning points in cycles. On the basis of these points, we can associate a contraction period with the lapse of time that separates a high point (peak) from a low point (trough). We follow the procedure advocated by Harding and Pagan (2002a,b) to identify turning points. This procedure states that a peak/trough has been reached at t when the value of the studied series at date t is superior/inferior to the previous k values and to the following k values, where k is a positive integer that varies according to the type of series studied and its sampling frequency. A procedure is then implemented to ensure that peaks and troughs alternate, by selecting the highest/lowest consecutive peaks/troughs. Additional censoring rules are implemented, which, for example, restrict the minimal phase and cycle durations.[5]

1.1 The concordance index

We can now define the contraction and expansion phases for one or more variables and thus define the concordance statistic that indicates the (standardised) average number of periods in which two variables (eg GDP and a stock market index) coincide at the same phase of the cycle. There is a perfect concordance between the series (perfect juxtaposition of expansions and contractions) if the index is equal to 1 and perfect disconcordance (a marked lag or out of phase) if the index is equal to 0.

Once the turning points of a variable y have been identified, we can define the binary variable $s_{y,t}$ such that:

$$s_{y,t} = \begin{cases} 1 \text{ if } y \text{ is in expansion at } t \\ 0 \text{ otherwise} \end{cases}.$$

[5] See Appendix A for further details on the determination of business cycle dates.

We proceed in the same fashion with x, by defining $s_{x,t}$. The concordance index between x and y, c_{xy}, is then defined as the average number of periods where x and y are identified simultaneously in the same phase, and is expressed as follows:

$$c_{xy} = \frac{1}{T}\sum_{t=1}^{T}\left[s_{x,t}s_{y,t} + \left(1-s_{x,t}\right)\left(1-s_{y,t}\right)\right]$$

Thus, c_{xy} is equal to 1 if x and y are always in the same phase and to 0 if x and y are always in opposite phases. A value of 0.5 indicates the lack of any systematic relationship in the dynamics of the two variables.

As McDermott and Scott (2000) observed, it is only possible to compute analytically the statistical properties of c_{xy} in a handful of particular cases. For example, if the processes x and y are independently drawn from the same Brownian motion, assuming that no censoring rules have been enforced in defining the turning points, then c_{xy} has mean 1/2 and variance $1/[4(T–1)]$.

Note that if T is very large, the variance of c_{xy} converges to 0 (c_{xy} is asymptotically constant).

However, in general, the distribution properties of c_{xy} are unknown, especially when the censoring rules have been enforced. In order to calculate the degrees of significance of these indices, we use the method suggested by Harding and Pagan (2002b) given below. Let μ_{s_i} and σ_{s_i}, $i = x,y$, denote the empirical average and the empirical standard deviation of $s_{i,t}$. If ρ_s denotes the empirical correlation between $s_{x,t}$ and $s_{y,t}$, it can be shown that the concordance index obeys:

$$c_{xy} = 1 + 2\rho_s\sigma_{s_x}\sigma_{s_y} + 2\mu_{s_x}\mu_{s_y} - \mu_{s_x} - \mu_{s_y} \tag{1.1}$$

According to equation (1.1), c_{xy} and ρ_s are linked in such a way that either of these two statistics can be studied to the same effect. In order to calculate ρ_s, Harding and Pagan estimate the linear relationship:

$$\left(\frac{s_{y,t}}{\sigma_{s_y}}\right) = \eta + \rho_s\left(\frac{s_{x,t}}{\sigma_{s_x}}\right) + u_t \tag{1.2}$$

where η is a constant and u_t an error term.

The estimation procedure of equation (1.2) must be robust to possible serial correlation in the residuals, as u_t inherits the serial correlation properties of $s_{y,t}$ under the null hypothesis $\rho_s = 0$. The ordinary least squares method augmented by the HAC procedure is therefore used here.

Note that equation (1.1) makes it clear that it is difficult to assess a priori the significance of c_{xy} relative to 0.5. Indeed, in the case of independent, driftless Brownian motions, $\rho_s = 0$, and $\mu_{s_x} = \mu_{s_y} = 0.5$, so that $c_{xy} = 0.5$. Now, assume that x and y are drawn from the same Brownian motion, though characterised by drifts, so that $\mu_{s_x} = \mu_{s_y} = 0.9$. In this case, using equation (1.1), it must be the case that $c_{xy} = 0.82$. However, x and y have been sampled independently, and should not be characterised by a high degree of concordance. Thus, a high value for c_{xy} relative to 1/2 is not synonymous with a high degree of concordance.

1.2 Presentation of the data

We set out to study the relationship between business cycles and stock market cycles in France, Germany, Italy, the United Kingdom and the United States.

Stock prices are obtained from composite indices calculated by Morgan Stanley (MSCI), deflated by the consumer price index. These variables are available at a quarterly and a monthly frequency. We use three variables to define the business cycle: at the quarterly frequency, market GDP and household consumption (these variables are taken from the OECD database over the study period from the first quarter of 1978 to the third quarter of 2002); and at the monthly frequency, retail sales (in volume terms, over the period January 1978-December 2002). This series is only available as of 1990 for Italy. We therefore do not take this country into account in our analysis of monthly data. Moreover, the monthly sales index displays a highly erratic pattern that could conceal some turning points. We strip out the most erratic parts of these series by prefiltering and focus the analysis on an adjusted version of these variables.[6]

The data sources are detailed below:

- *Financial data*: Morgan Stanley Capital International (MSCI) indices obtained from Datastream. In order to calculate excess returns, we use the nominal interest rate on government bonds (annualised) for France, the United Kingdom and the United States, the interbank rate for Germany and the money market rate for Italy. For all of these countries, we use the three-month money market rates as indicators of monetary policy. These data are obtained from the IMF database.
- *Real data*: real market GDP and real private consumption are expressed in 1995 prices. Real sales are obtained from the real retail sales index (1995 base year). These data are obtained from the OECD database. We also use the consumer price index from the same database to deflate the stock market indices.

1.3 Results

The turning points in real GDP, real consumption and MSCI indices are shown in Graphs 1, 2 and 3, respectively. Those for the retail sales index and the MSCI indices at the monthly frequency are given in Graphs 4 and 5, respectively.

At the quarterly frequency, results derived from the graphs relating to real activity variables (Graphs 1 and 2) are compatible overall and consistent with the analysis of McDermott and Scott (2000) and with that of Artis et al (2003). Naturally, we do not detect a perfect identity between the cycles described by GDP and real consumption. In France, for example, a short contraction can be observed in 1995 when we study private consumption data, whereas the French economy was in a phase of expansion according to GDP data. When studying the turning points observed in stock markets, we note in particular that they are more frequent than in the real economy, irrespective of the country considered in our sample. The long phase of expansion in the 1990s is clearly visible in all countries. Some pronounced lags are observed between the phases of the business and stock market cycles, in particular in Europe, and especially at the start of the 2000s.

We note that the retail sales index is a more or less reliable indicator of private consumption and is more volatile than the latter. Nevertheless, these are the two indicators that must be compared. We therefore compare the turning points derived from the analysis of these two variables. Overall, in sales indices we observe the same marked contractions as in consumption, as well as more occasional contractions, consistent with the high volatility of these indices. We can carry out the same analysis on stock market indices at two frequencies: all pronounced contractions at a quarterly frequency can also be observed at a monthly frequency; here, too, more contractions are detected at the monthly frequency.

These initial findings obtained from analysing the graphs naturally call for a more in-depth study of the co-movements of real economy and stock market variables. Table 1 lists the intra-country index of concordance between the MSCI indices and the three real activity indicators used.

The United States appears to be characterised by a significant concordance between the level of real activity and stock prices. Indeed, this is the case for all three real activity indicators used, which is not surprising in view of the role of stock markets in the investment and financing behaviour of US economic agents. However, the same is not true of the other countries in the sample.

Stock market and business cycles do not occur at the same frequencies and furthermore may be uncorrelated, with the exception of the United States. Indeed, an analysis of Graphs 1 (or 2) and 3 shows that the duration of a stock market expansion is generally shorter than one in GDP or consumption. This difference naturally contributes to reducing the degree of concordance between real activity and stock markets. Nevertheless, the lack of significant concordance in most countries under review does not necessarily mean that business and stock market cycles are different or

[6] See Watson (1994).

uncorrelated phenomena. The result obtained simply highlights the fact that the periods of expansion and contraction of GDP and stock prices (for example) do not coincide.

We observe that the start of US stock market contractions (ie the dates of peaks) precede contractions in real activity measured by real GDP.[7] The lag oscillates between one and four quarters. However, we also note that not all stock market contractions result in contractions in real activity. In particular, when they are very short (like in 1987), they do not seem to spill over into real activity. A similar phenomenon can be detected in European countries such as France and Italy. Like in the United States, but to a lesser degree, GDP contractions are preceded by stock market contractions, although most stock market contractions in these two countries do not lead to contractions in real activity. However, this rule does not apply to Germany and the United Kingdom. Stock market contractions may precede or follow contractions in real activity by more than a year. Therefore, contrary to received wisdom, it does not always appear relevant to use negative turning points in stock markets as leading indicators of the start of a contraction phase of GDP or consumption.

Turning now to the relationship between monetary policy and business and stock market cycles, we observe a relative decoupling between certain contraction periods of real activity or stock markets and money market rate developments, used here as indicators of monetary policy (Graph 6). No clear rule emerges from a comparison between stock markets and money markets: for the business cycle, a decline in rates more or less coincides with a contraction but, here too, it is difficult to establish a general rule. This graph suggests that the reaction of money market rates to turnarounds in real activity or stock markets is not systematic or correlated in the countries studied. This corresponds in theory to the mandate of monetary authorities as well as to the way we have modelled monetary policy rules in recent macroeconomic studies.[8]

Concordance indices have enabled us to measure the degree of “juxtaposition” between two chronological series, without having to consider whether there is a trend in the variables (non-stationarity). It should nevertheless be noted that only one aspect of the notion of cycles is taken into account here.

It could therefore be useful to broaden the study by retaining the concepts of phase and duration, but without limiting ourselves to such restrictive indicators as concordance indices. To do this, in Part two we decompose the different series studied in order to isolate the long-term (or structural) and the short-term (or cyclical) components; the latter correspond to the business cycle concept put forward by the NBER.

2. Correlation of cyclical and structural components

On the basis of NBER studies, we identify business cycles with all movements whose recurrence period is between six and 32 quarters.[9] This corresponds to the frequency of business cycles. Furthering this approach, it has become common in macroeconomics to split a variable (y_t) according to the frequencies band over which its components are concentrated. The one corresponding to the business cycle is determined as the residual obtained after stripping out long movements, imputable to structural economic factors (τ_t).[10] By construction, the residual variables ($y_t - \tau_t$) obtained by robust statistical techniques (filtering) are detrended (stationary). We can thus calculate the correlations between the corresponding components of the series in the hope of isolating a set of statistical regularities or stylised facts that characterise the business cycle.

The analysis of these components is based on the assumption that it is possible to isolate them from each other. To this end, we use two complementary non-parametric methods. First, we take the band pass filter recently put forward by Christiano and Fitzgerald (2003) (CF filter). For each country and

7 To date, statistics for testing the significance of these lags do not exist.

8 See, in particular, studies in the collective work edited by Taylor (1999).

9 Estrella (2003) uses a slightly different definition of business cycle frequencies.

10 This is the approach generally adopted following Kydland and Prescott (1982).

each variable (y_t), we thus define the short-term (or cyclical, y_t^{st}) components and the long-term (or structural, y_t^{lt}) components and calculate the correlations between the corresponding components. Second, we compute the dynamic correlations between the studied variables, following the work by Croux et al (2001).

The following section briefly reviews the methodological tools used.

2.1 A brief review of spectral analysis

2.1.1 The band pass filter

The ideal band pass filter used to isolate cyclical movements whose recurrence periods are in the interval $[b_l, b_u]$, is defined by the following equation:

$$y_t^{st} = B(L)y_t, \qquad B(L) = \sum_{k=-\infty}^{k=+\infty} B_k L^k, \qquad L^k y_t = y_{t-k},$$

where B_k is expressed as:

$$B_k = \frac{\sin(2k\pi / b_l) - \sin(2k\pi / b_u)}{\pi k}.$$

In order to interpret the role played by the filter, we introduce the concept of *spectral density*. The spectral density of the stationary stochastic process y_t, denoted $S_y(\omega)$, is interpreted as the decomposition of the variance of y_t in the frequency domain. As y_t can be decomposed into a sum of orthogonal cyclical movements that each appear at a different frequency, we can interpret $S_y(\omega)$ as the variance of y_t explained by the cyclical movements operating at frequency ω.

A classic result of spectral analysis shows us that, under certain conditions, the equation $y_t^{st} = B(L)y_t$ implies that the spectral density of the process y_t^{st}, $S_{y^{st}}(\omega)$, is deduced from that of y_t, $S_y(\omega)$, using the formula:

$$S_{y^{stt}}(\omega) = \left\| B(e^{-i\omega}) \right\|^2 S_y(\omega),$$

where $\left\| B(e^{-i\omega}) \right\|^2$ is the squared modulus of $B(e^{-i\omega})$. Given the definition of B_k, a direct calculation shows that:

$$B(e^{-i\omega}) = \begin{cases} 1 & \text{for} \quad \omega \in]2\pi / b_u, 2\pi / b_l[\cup] - 2\pi / b_l, -2\pi / b_u[\\ 0 & \text{otherwise} \end{cases}.$$

From this formula it can be observed that the spectral density of y_t is not 0 on the frequency band $]2\pi/b_l, 2\pi/b_u[\cup]-2\pi/b_l, -2\pi/b_u[$, and 0 everywhere else in the interval $]-\pi, \pi[$. In other words, all the variance of y_t^{ct} is explained by cyclical movements whose recurrence periods are between b_l and b_u.

The definition of the filter $B(L)$ imposes a major limitation, as it requires a data set of infinite length. In practice, we work with a finite sample and must therefore make an appropriate approximation of $B(L)$. Starting from a finite number of observations $\{y_1, ..., y_T\}$ of the stochastic process y_t, Christiano and Fitzgerald (2003) define the optimal linear approximation $\hat{y}_t^{st}$ of y_t^{st} as the solution to the problem:

$$\min E\left[\left(y_t^{st} - \hat{y}_t^{st} \right)^2 \middle| \{ y_1, \ldots, y_T \} \right] \tag{2.1}$$

The method therefore consists in minimising the mathematical expectation of the square error between the ideally filtered series and the approximately filtered series, where the expectation is conditioned on all the available data.

2.1.2 Dynamic correlation

Consider a bivariate stationary stochastic process $(x_t, y_t)'$. The classical notion of correlation is a static measure of the linear relation between x_t and y_t. In contrast, the dynamic correlation between x_t and y_t, denoted $\rho_{xy}(\omega)$, permits us to decompose the correlation between these series in the frequency domain. In particular, it allows us to quantify the amount of covariation between the cyclical components of x_t and y_t at frequency ω.

Let us define formally the notion of dynamic correlation. Let $S(\omega)$ denote the spectral density of $(x_t, y_t)'$:

$$S(\omega) = \begin{pmatrix} S_x(\omega) & S_{xy}(\omega) \\ S_{yx}(\omega) & S_y(\omega) \end{pmatrix}, \quad \omega \in [-\pi, \pi],$$

where the cross-spectrum $S_{xy}(\omega)$ is a complex number, such that $S_{xy}(\omega) = S_{yx}(\omega)'$ (where " ' " denotes the transpose-conjugate operation). The dynamic correlation $\rho_{xy}(\omega)$ associated with $(x_t, y_t)'$ is defined by the relation:

$$\rho_{xy}(\omega) = \frac{C_{xy}(\omega)}{\sqrt{S_x(\omega) S_y(\omega)}}, \quad \omega \in [0, \pi[,$$

where $C_{xy}(\omega)$ is the real part of $S_{xy}(\omega)$. Thus, this statistic is nothing more than the correlation coefficient between real waves of frequency ω appearing in the spectral decomposition of $(x_t, y_t)'$.

To estimate $\rho_{xy}(\omega)$ we first estimate $S(\omega)$ through the well known relation:

$$S(\omega) = \sum_{k=-\infty}^{k=+\infty} \Gamma_k e^{-i\omega k}, \quad \omega \in [-\pi, \pi].$$

Here, $\Gamma_k = E z_t z'_{t-k}$ is the k-th autocovariance of $(x_t, y_t)'$. In practice, the Γ_k are not known and are replaced by their sample counterparts:

$$\hat{\Gamma}_k = \frac{1}{T} \sum_{t=k+1}^{T} z_t z'_{t-k},$$

where T is the sample size. Finally, $S(\omega)$ is replaced by its empirical estimate, denoted $\hat{S}(\omega)$, which is obtained by smoothing the empirical covariogram with a Bartlett window of width q:

$$\hat{S}(\omega) = \frac{1}{2\pi} \left[\hat{\Gamma}_0 + \sum_{k=1}^{q} \left(1 - \frac{k}{q+1}\right) \left(\hat{\Gamma}_k e^{-i\omega k} + \hat{\Gamma}'_k e^{i\omega k}\right) \right].$$

Finally, to compute the confidence intervals reported below, we used a traditional block-bootstrap approach.

2.2 Empirical results

Here, the analysis is limited to quarterly frequencies. The different real activity indicators are logarithms of real market GDP and private consumption; for the financial sphere, we consider the excess returns on stocks relative to the risk-free interest rate.[11]

---

11 Excess returns are defined as the difference between the nominal interest returns on stocks and on three-month government bonds.

We propose two applications. First, for each country, we calculate the correlation between the cyclical (short-term) components of the variables studied and the correlation between the structural (long-term) components. In the latter case, we do not deal with real activity indicators and measures of returns in the same way. Indeed, real activity indicators are characterised by trends and therefore do not have the required statistical properties (they are not stationary) for calculating the correlations.

We show that their long-term components are non-stationary too. Consequently, we focus on the growth rate of the structural components that are, in general, stationary (in particular, they are not characterised by a trend). Conversely, the excess returns on stocks relative to the risk-free interest rate and their components are stationary. We can therefore study these variables in level form.

In order to determine the cyclical components, we adopt the traditional definition of the cycle presented above. For all the variables studied, the business cycle is identified with all movements whose recurrence period is between six and 32 quarters. In order to isolate the structural components, we apply the CF filter so as to strip out the cyclical movements with a recurrence period of less than 32 quarters. We then calculate the difference between the initial series and the filtered series to obtain the structural component.

Let y_t denote the log of real GDP at t, and x_t the excess return at t. For each country i (i = France, Germany, Italy, the United Kingdom and the United States), we calculate the following correlations:

– the correlation between the cyclical component of GDP and excess returns, $y_{t+k}^{st}(i)$ and $x_t^{st}(i)$, for k = –3,...,3;

– the correlation between the growth rate of the structural component of GDP, $\Delta y_{t+k}^{lt}(i)$, and the structural component of excess returns, $x_t^{lt}(i)$, for k = –3,...,3;

where Δ is the first difference operator ($\Delta a_t = a_t - a_{t-1}$). We establish k as ranging from –3 to 3 as is the usual practice in studies of US data. For the purposes of symmetry, we adopt the same horizon for the other countries. As mentioned above, the exponent *st* denotes the short-term component and the exponent *lt* denotes the long-term component. We estimate these correlations using the Generalised Method of Moments (GMM) completed with the HAC procedure developed by Andrews and Monahan (1992). We use the same methods for real private consumption, replacing y_t by c_t, the logarithm of consumption.

Second, for each country, we calculate the dynamic correlation between excess returns and either GDP growth or consumption growth. We decide to study growth rates of trending variables for the same reasons as those outlined above. Thus, it is important to keep in mind that the dynamic correlation between output growth and excess returns at low frequencies does not exactly cover the same phenomenon as the simple correlation between the structural component of excess returns and the growth rate of the structural component of output.

From Tables 2 and 3, we cannot conclude that there is a strong link between the cyclical components of GDP or consumption and those of excess returns in the different countries reviewed.

However, in France, Germany and the United States, the correlation between y_{t+k}^{st} and x_t^{st} is significantly positive for k = 2 or 3 quarters. This means that a positive variation of the cyclical component of GDP at $t+2$ or at $t+3$ is associated with a positive variation of the cyclical component of excess returns at t. In other words, a positive variation of the cyclical component of GDP follows an increase in the cyclical component of excess returns with a lag of two or three quarters.[12]

Even though the share of equities in household wealth differs across the Atlantic[13] the reactions of the three economies display a certain convergence. A similar link is observed for the cyclical component of consumption, although the lag in the correlation appears to be closer to three quarters.

[12] This result must, however, be considered with caution as the sign of the correlation coefficient sometimes changes with k in some countries (see the line corresponding to the United States).

[13] See Odonnat and Rieu (2003).

However, the correlations between the growth rate of the structural component of GDP and the structural component of excess returns are significantly positive for all countries, at a fairly short horizon (Tables 4 and 5). The structural determinants of excess returns appear to covary positively with those of real activity. This result is borne out overall when consumption is used as a real activity indicator, at least for short horizons.[14]

The previous results are partly confirmed by the dynamic correlation analysis. Figure 7 reports the dynamic correlation between GDP growth and excess returns. The graph clearly shows that, in most countries, this correlation is significantly positive at low frequencies while not always significantly different from 0 at higher frequencies. This confirms our analysis: excess returns and real activity are strongly linked at low frequencies, because they share possibly common structural determinants; conversely, at shorter horizons, the determinants of these variables can differ. Graph 8 reports the dynamic correlation between consumption growth and excess returns. Once again, we obtain similar results, even though the dynamic correlation appears to be higher at higher frequencies for some countries.

If we compare the cyclical and structural components of the real activity indicator, stock prices and interest rates, we see that in most countries studied (Table 6), with the notable exception of France, the correlation between the cyclical component of GDP and that of the nominal interest rate is positive for negative *k* and negative for positive *k*. These results seem to point to a stabilising monetary policy: temporary rises in the level of real activity are followed by temporary increases in the money market rate, which precede a decline in the cyclical component of GDP. The difference in the French case may be due, inter alia, to the implementation of the "strong franc" policy at the start of the 1980s, which introduced a break.

We do not, however, detect a significant relationship between the cyclical component of excess returns and that of money market rates (Table 7), except in the United Kingdom: overall, short-term fluctuations in excess returns appear in some respects to be independent of those in money market rates. If we use these rates to represent monetary policy, this analysis does not rule out the possibility that monetary authorities may have reacted to some stock market events, but it indicates that, in general, stock price fluctuations do not play a determining role in the conduct of their policy. In results not reported here, we obtain confirmation of this conclusion with the dynamic correlation approach. The latter is not found statistically significant at business cycle frequencies.

Table 8 suggests that there is a negative relationship between the long-term component of the money market rate and that of real GDP in the United States, France and Germany (where we observe a lag).[15]

This relationship means that a lasting rise in the money market rate results in a fall in the growth rate of the long-term component of GDP. We could enhance the interpretation of this result by comparing the long-term components of real activity with those of real interest rates, calculated ex ante, in keeping with economic theory. However, this exercise is not easy because no simple and reliable measurement of this interest rate is available.

Lastly, we do not detect a significant link between the long-term component of the money market rate and that of excess returns (Table 9), except in the United Kingdom and to a lesser extent in the United States. The long-term component of interest rates therefore does not appear to react to the structural component of excess returns, except in the United Kingdom and the United States, no doubt owing to the weight of equities in household wealth that characterises these countries.

[14] We can compare these conclusions with those of Daniel and Marshall (1998). These authors show that it is not possible to reject the augmented C-CAPM models when consumption and excess returns have been stripped of their short-term cyclical movements.

[15] Once again, we obtain similar results with the dynamic correlation approach.

Conclusion

In order to understand the link between business cycles and stock market cycles and use it to improve the conduct of monetary policy, it is first necessary to identify the stylised facts underlying this relationship.

In practice, we set out to study the links between business and stock market cycles by using two complementary approaches that enable us to measure the co-movements between these phenomena.

First, in the tradition of the NBER, we defined the business cycle as a succession of phases of expansion and contraction in order to compare the cycles based on two variables by calculating their concordance index. Above all, this exercise allowed us to identify significant concordance between the business and stock market cycles in the United States.

Second, using the predominant methodology in applied macroeconomics, we analysed this link by decomposing the variables studied into short- and long-term components and by calculating the correlations between corresponding components (ie cyclical/cyclical and structural/structural).

We draw two conclusions from the various analyses carried out: (i) there does not seem to be a strong dependence link between stock prices and the level of real activity at business cycle frequencies, except in the United States; and (ii) in the longer term, it appears that real activity and stock prices share the same determinants. At any rate, we cannot clearly identify an impact of asset prices on three-month interest rates, used to represent monetary policy in the countries studied. In general, we do not detect a significant relationship between the cyclical components of excess returns and money market rates, nor do we observe a significant link between the structural components of these same variables.

These conclusions appear to be robust. However, it may be useful to further investigate the dichotomy between the short and long term using an approach based on a behavioural analysis of agents (or a microeconomic analysis of markets). In particular, we will attempt to identify the transmission mechanisms that enable us to detect links between business and stock market cycles.

Table 1

Concordance between real and financial cycles

	United States	France	Germany	United Kingdom	Italy
GDP	0.68687*	0.61616	0.62626	0.58586	0.54545*
Consumption	0.64646*	0.60606	0.66667*	0.59596	0.53535
Sales	0.73874*	0.54655	0.56456	0.62462*	...

Note: A star denotes a coefficient significant at the 5% level. These levels are determined according to the method advocated by Harding and Pagan (2002b).

Table 2

Short-run correlation, GDP-stock prices

k	**–3**	**–2**	**–1**	**0**	**1**	**2**	**3**
United States	–0.0097	–0.1872	–0.2940	–0.2835	–0.1528*	0.0493	0.2461*
France	–0.0020	0.1015	0.2178	0.2884	0.2729*	0.1789*	0.0377
Germany	–0.1131	–0.1129	–0.0438	0.0656	0.1666*	0.2357*	0.2625*
United Kingdom	0.1215	0.1276	0.0875	0.0070	–0.0675	–0.1023	–0.0938
Italy	0.1279	0.1631	0.1647	0.1381	0.0997	0.0769	0.0731

Note: Correlation between $y_{t+k}^{st}(i)$ and $x_t^{st}(i)$, where *i* is the country in the first column.

Table 3

Short-run correlation, consumption-stock prices

k	**–3**	**–2**	**–1**	**0**	**1**	**2**	**3**
United States	–0.1076	–0.1958	–0.2181	–0.1530	–0.0165	0.1352	0.2368*
France	–0.2315	–0.0839	0.0949	0.2280	0.2929*	0.2659*	0.1707
Germany	–0.1902	–0.2442	–0.2528	–0.2024	–0.0995	0.0502	0.2125*
United Kingdom	0.0208	–0.0262	–0.0816	–0.0975	–0.0609	0.012	0.0248
Italy	–0.0323	0.0018	0.0369	0.0793	0.1251	0.1830*	0.2362

Note: Correlation between $c_{t+k}^{st}(i)$ and $x_t^{st}(i)$.

Table 4

Long-run correlation, GDP-stock prices

k	**–3**	**–2**	**–1**	**0**	**1**	**2**	**3**
United States	0.6243*	0.6528*	0.6665*	0.6653*	0.6415*	0.6073*	0.5641*
France	0.1872*	0.3062*	0.4179*	0.5197*	0.5997*	0.6650*	0.7143*
Germany	0.0622	0.1381	0.2128	0.2845	0.3265*	0.3663*	0.4029*
United Kingdom	0.6161*	0.6242*	0.6175*	0.5965*	0.5586*	0.5093*	0.4501*
Italy	0.4909*	0.5735*	0.6424*	0.6959*	0.7254	0.7423	0.7462

Note: Correlation between $\Delta y_{t+k}^{st}(i)$ and $x_t^{st}(i)$.

Table 5

Long-run correlation, consumption-stock prices

k	**–3**	**–2**	**–1**	**0**	**1**	**2**	**3**
United States	0.3898	0.4041	0.4091*	0.4054*	0.4060	0.3889*	0.3850*
France	0.0629	0.1698*	0.2714*	0.3653*	0.4580*	0.5369*	0.6006*
Germany	0.0974	0.1675	0.2362	0.3019	0.3425*	0.3804*	0.4149*
United Kingdom	0.3423	0.3855	0.4175	0.4380	0.4556*	0.4602*	0.4522*
Italy	0.3377*	0.4391*	0.5305*	0.6098*	0.6598*	0.6991*	0.7266*

Note: Correlation between $\Delta c_{t+k}^{st}(i)$ and $x_t^{st}(i)$.

Table 6

Short-run correlation, GDP-money market rates

k	**–3**	**–2**	**–1**	**0**	**1**	**2**	**3**
United States	0.5341*	0.6218*	0.6334*	0.5430*	0.3629*	0.1096	–0.1750*
France	0.1775	0.1996	0.1827	0.1188	0.0219	–0.0801	–0.1720
Germany	0.7303*	0.7233*	0.6299*	0.4475*	0.2020*	–0.0585*	–0.2846*
United Kingdom	0.5535*	0.5172*	0.3870*	0.1663	–0.0904	–0.3187*	–0.4740*
Italy	0.5129*	0.5983*	0.5702*	0.4524*	0.2644	0.0973	–0.0137

Table 7

Short-run correlation, excess returns-money market rates

k	–3	–2	–1	0	1	2	3
United States	–0.0115	–0.1372	–0.22137*	–0.2298	–0.1842	–0.1009	–0.0007
France	–0.1078	–0.1159	–0.0643	–0.0195	–0.0058	–0.0222	–0.0417
Germany	0.0796	0.0778	0.0580	0.0235	–0.0111	–0.0231	–0.0007
United Kingdom	–0.1632	–0.729	0.1482	0.3792*	0.4989*	0.4289*	0.2083*
Italy	–0.0950	–0.0931	–0.0750	–0.0301	0.0367	0.1051	0.1381*

Table 8

Long-run correlation, GDP-money market rates

k	–3	–2	–1	0	1	2	3
United States	–0.2332	–0.2493	–0.2600*	–0.2646*	–0.2761*	–0.2776*	–0.2685*
France	–0.2404	–0.2906*	–0.3363*	–0.3764*	–0.4187	–0.4549	–0.4835
Germany	0.1101	0.0233	–0.0612	–0.1417	–0.2272	–0.3044*	–0.3715*
United Kingdom	–0.3266	–0.3582	–0.3824	–0.3986	–0.4026	–0.3929	–0.3691
Italy	0.1183	0.0932*	0.0732	0.0587	0.0309	0.0086	–0.0077

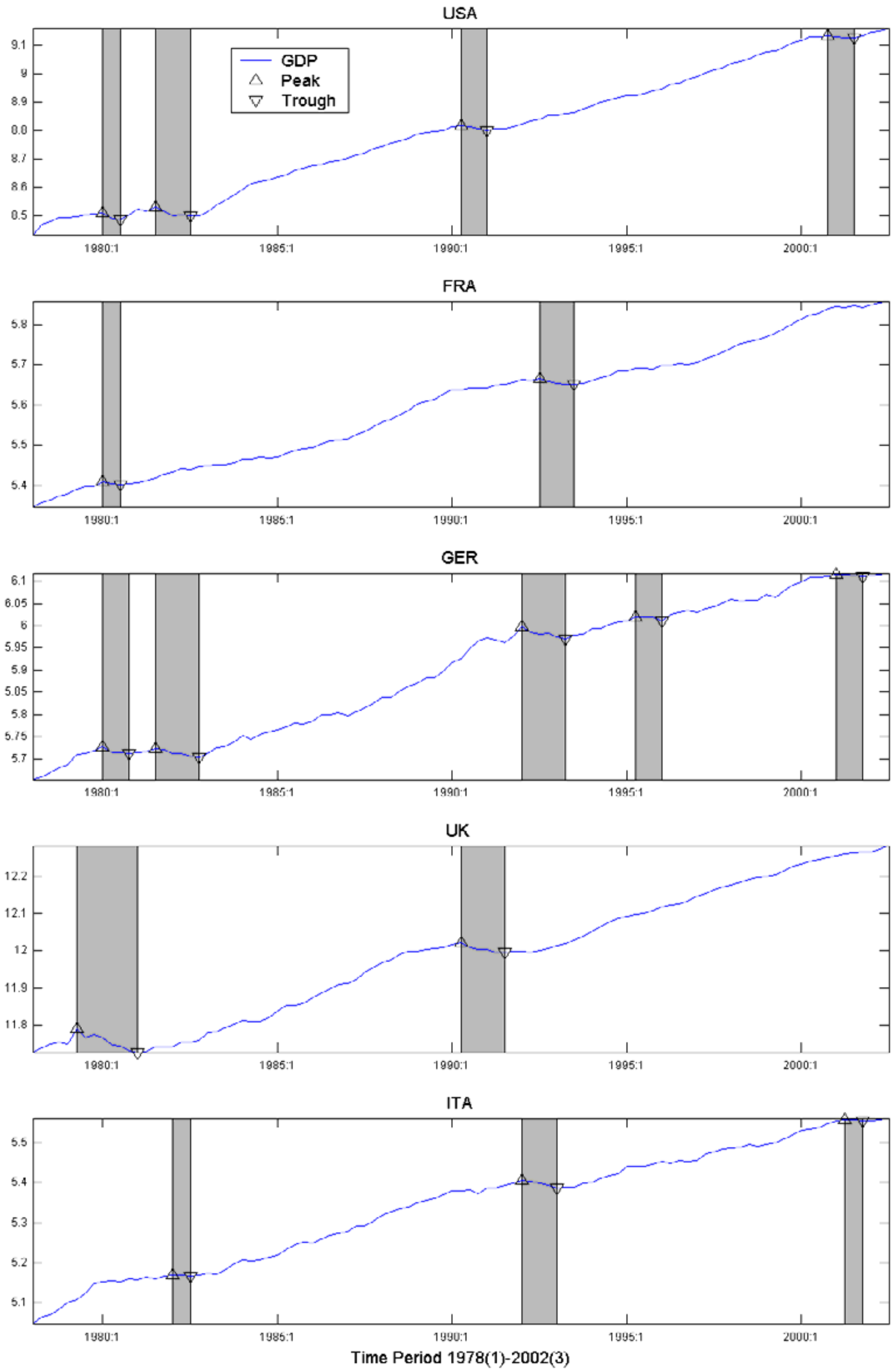
Table 9

Long-run correlation, excess returns-money market rates

k	–3	–2	–1	0	1	2	3
United States	0.0312	0.0615	0.0895	0.1155*	0.0606	0.0112	–0.0316
France	–0.167	–0.1386	–0.0995	–0.0497	–0.0618	–0.0630	–0.0528
Germany	–0.2636	–0.2238	–0.1724	–0.1097	–0.1036	–0.0860	–0.0571
United Kingdom	0.2013*	–0.2068*	0.2163*	0.2305*	0.1796	0.1347	0.0971
Italy	0.0489	0.147	0.1693	0.2421	0.2326	0.2276	0.2270

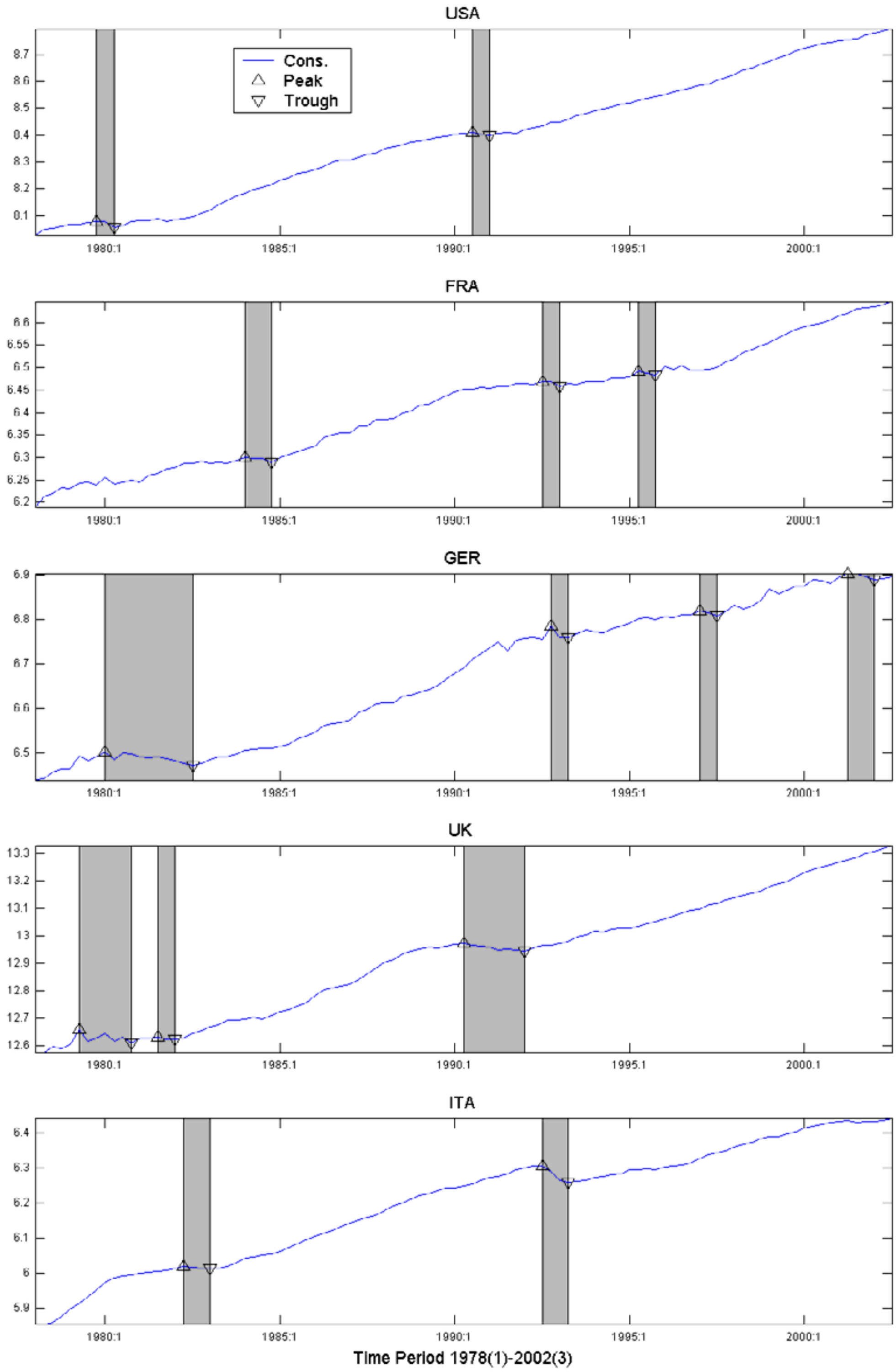
Graph 1

Turning points of real GDP



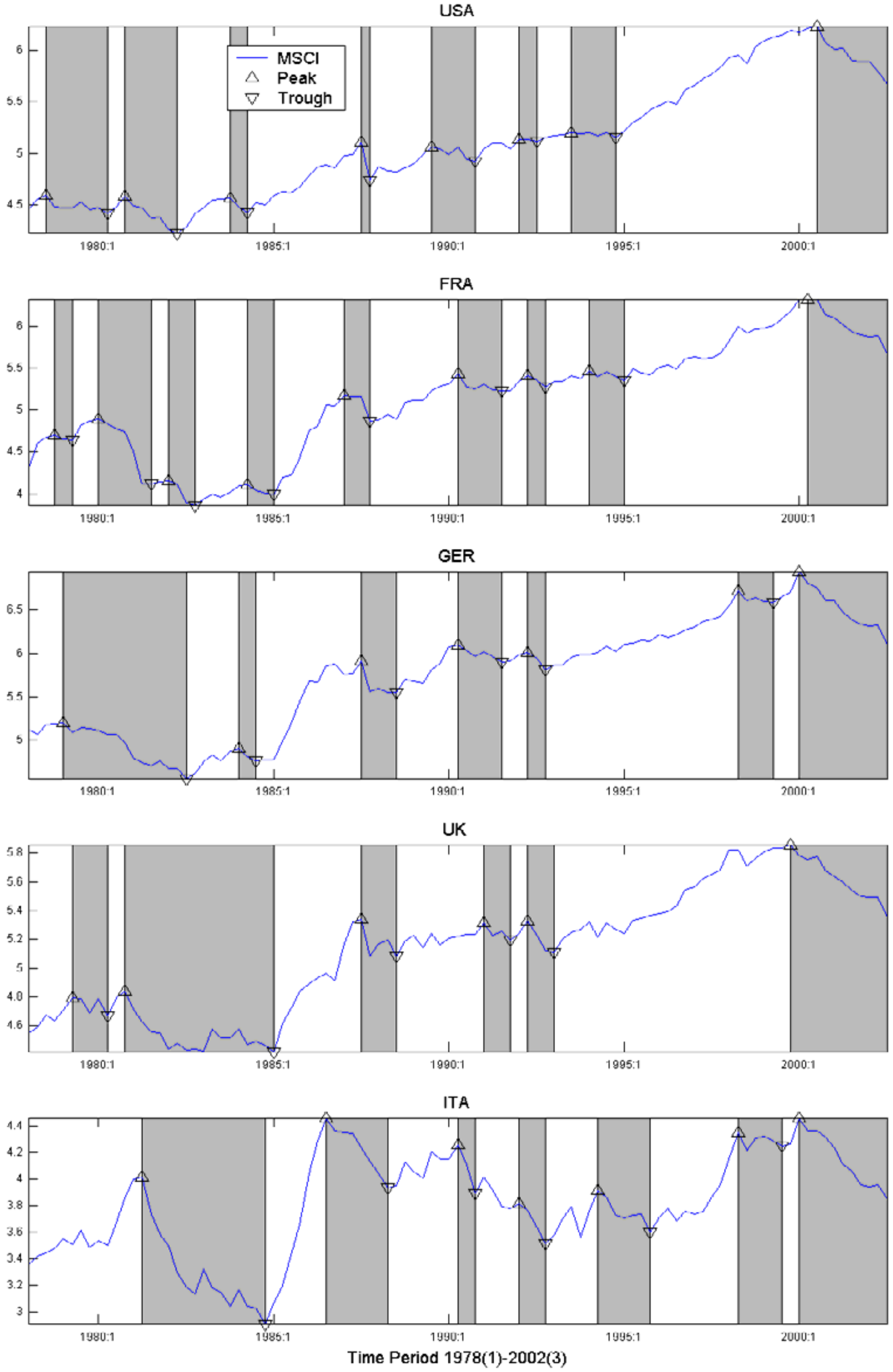
Graph 2

Turning points of real private consumption



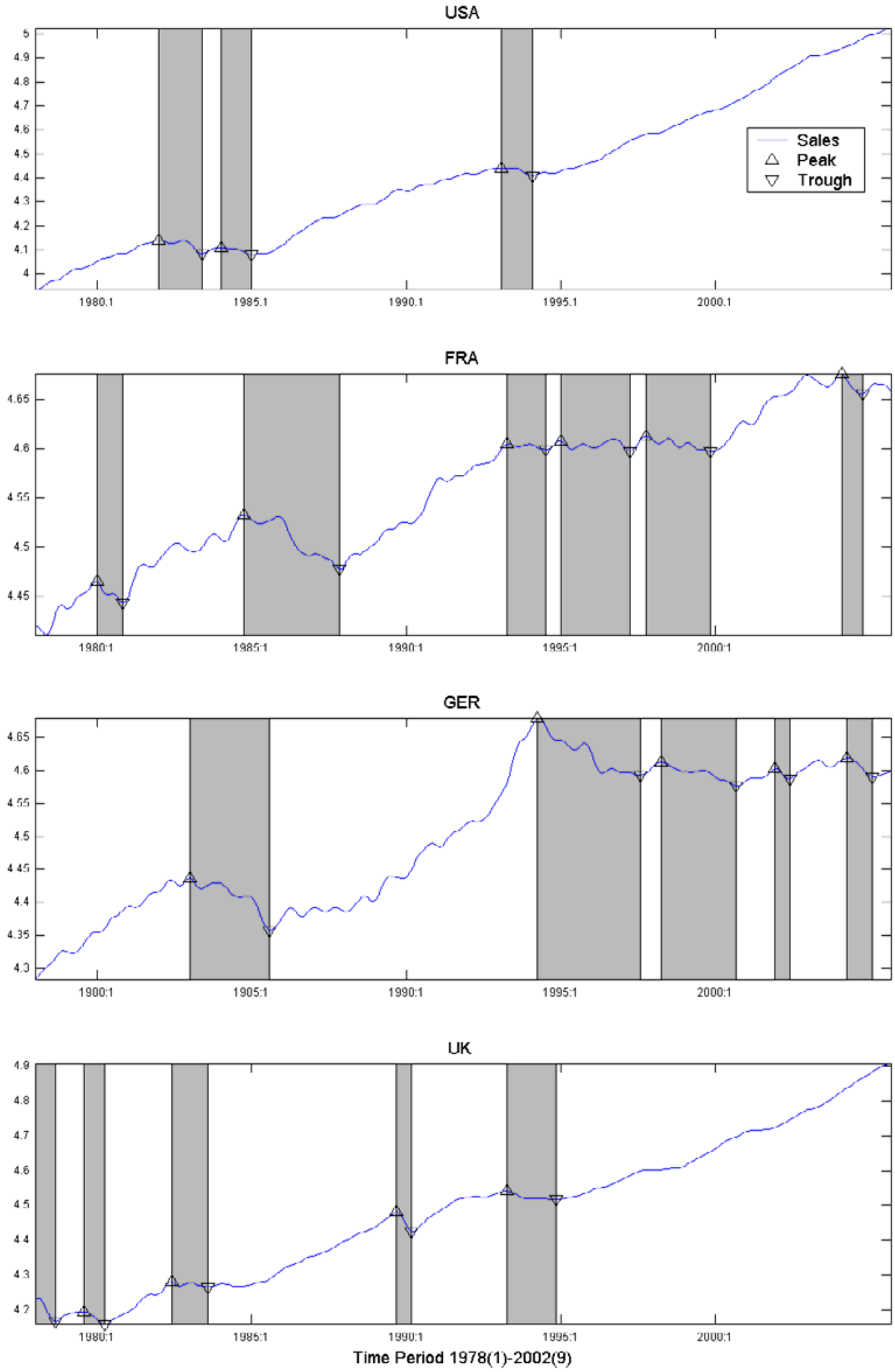
Graph 3

Turning points of MSCI return indices



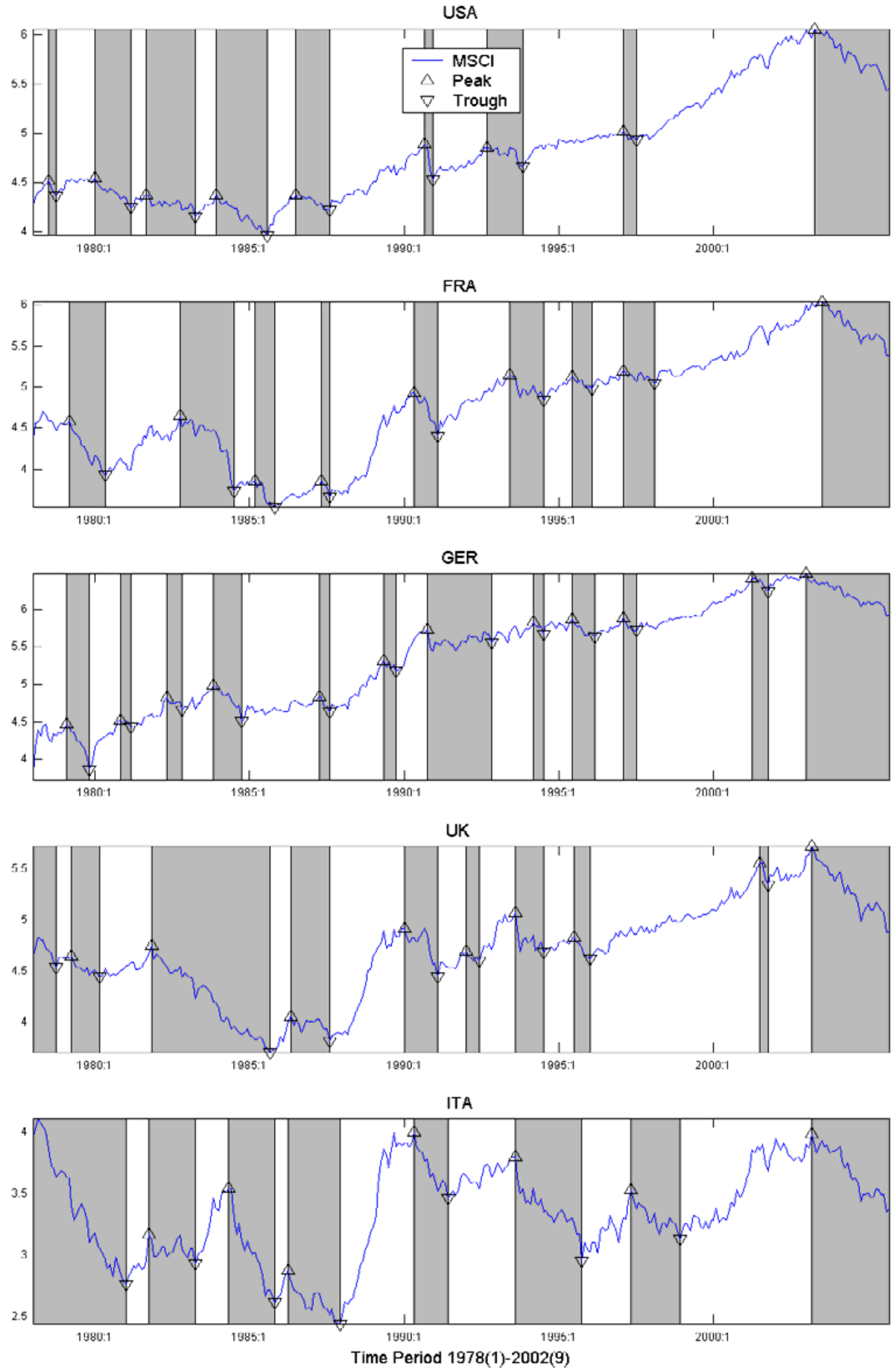
Graph 4

Turning points of real retail sales index (filtered)



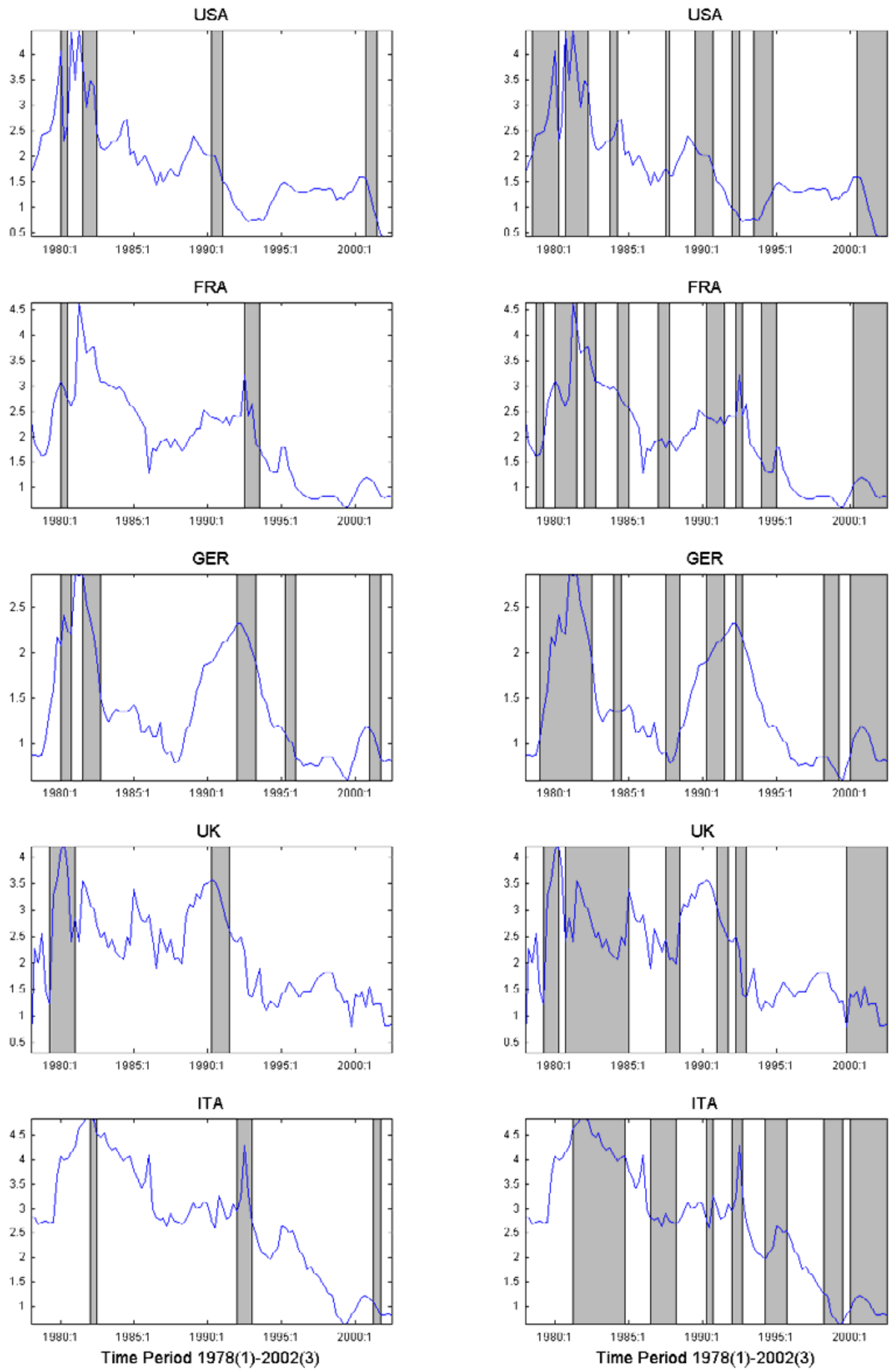
Graph 5

Turning points of monthly MSCI indices



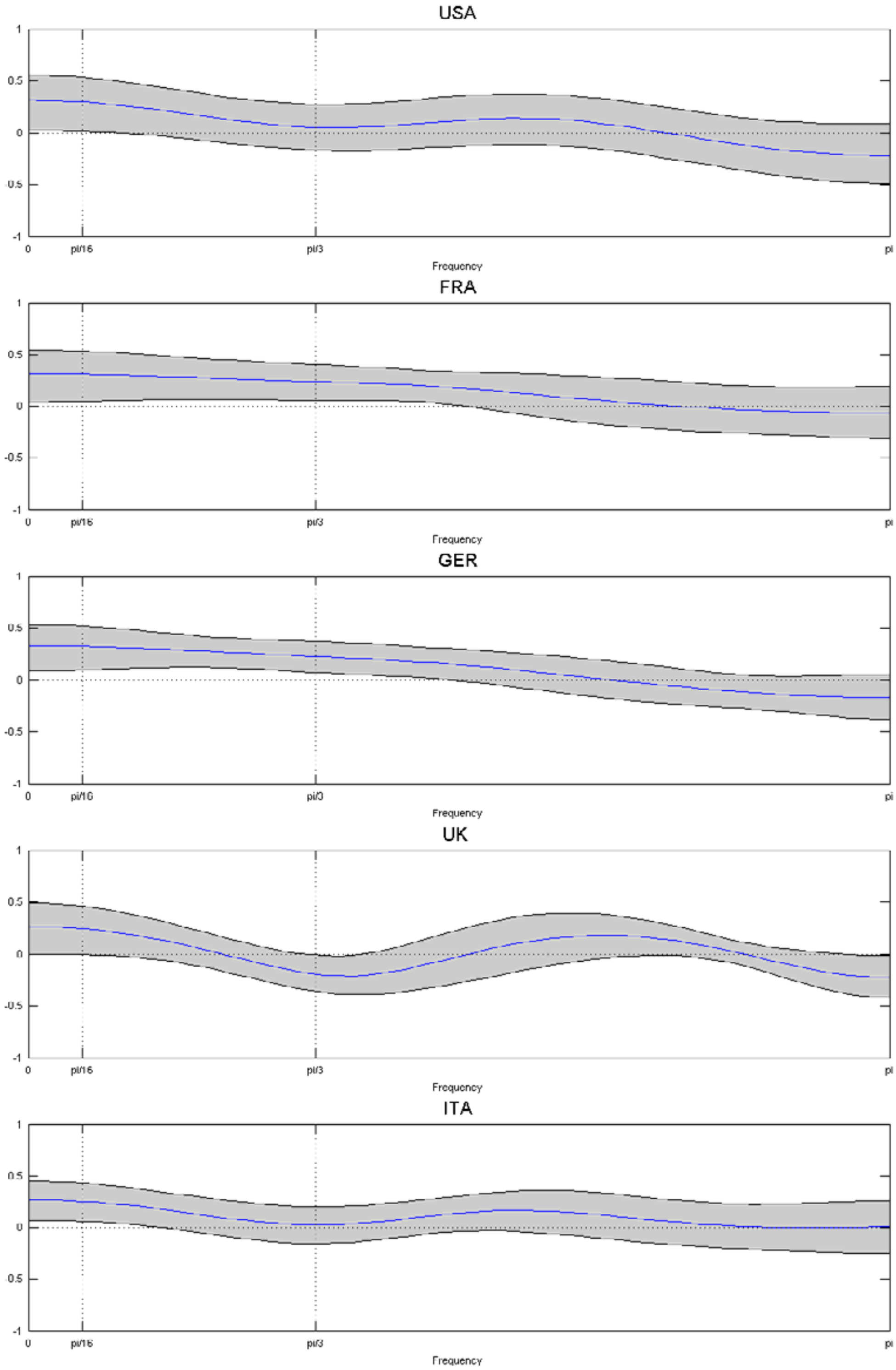
Graph 6

Money rates and GDP turning points (left-hand column) and return index turning points (right-hand column)



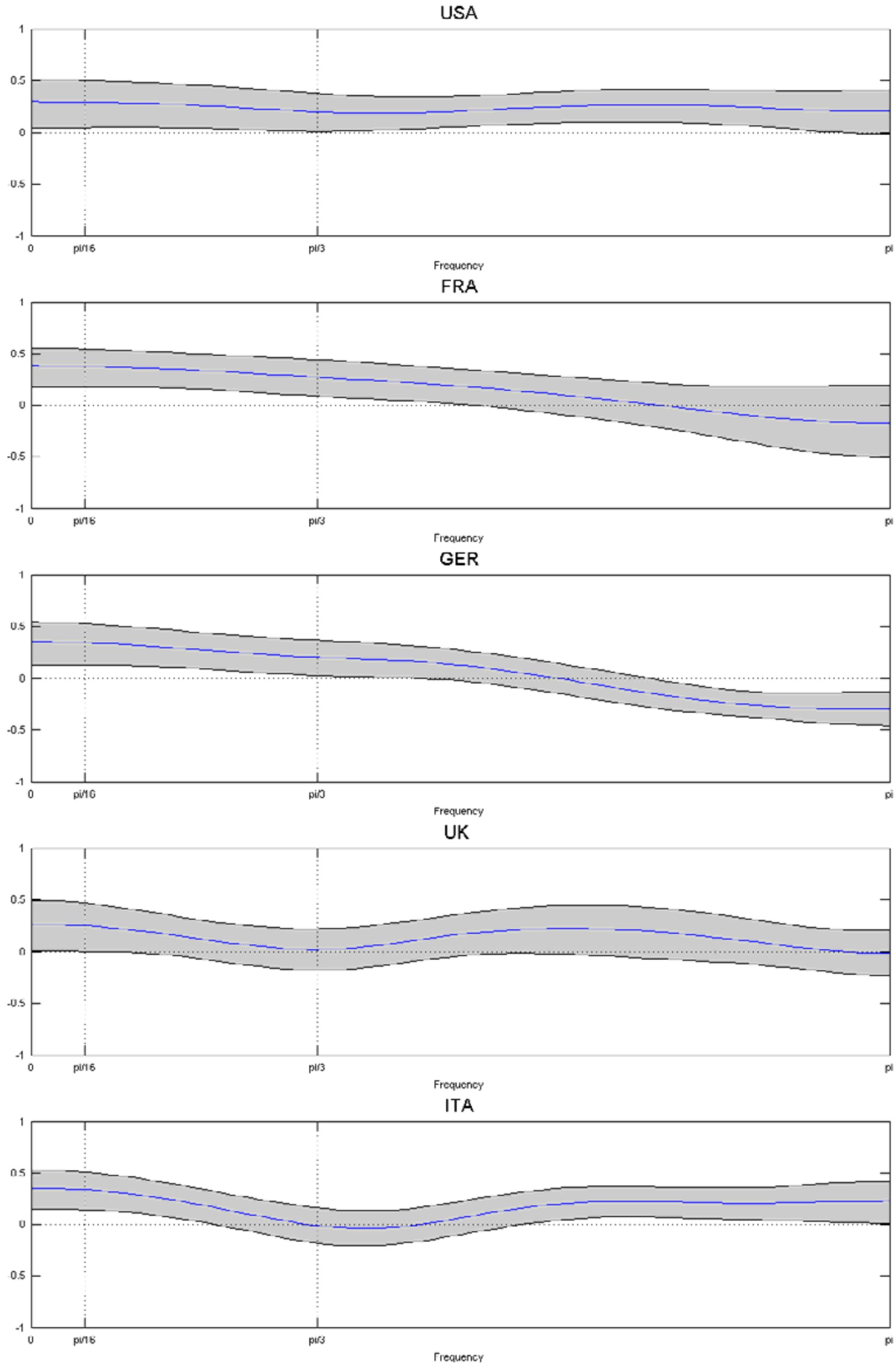
Graph 7

Dynamic correlation between GDP growth and excess returns



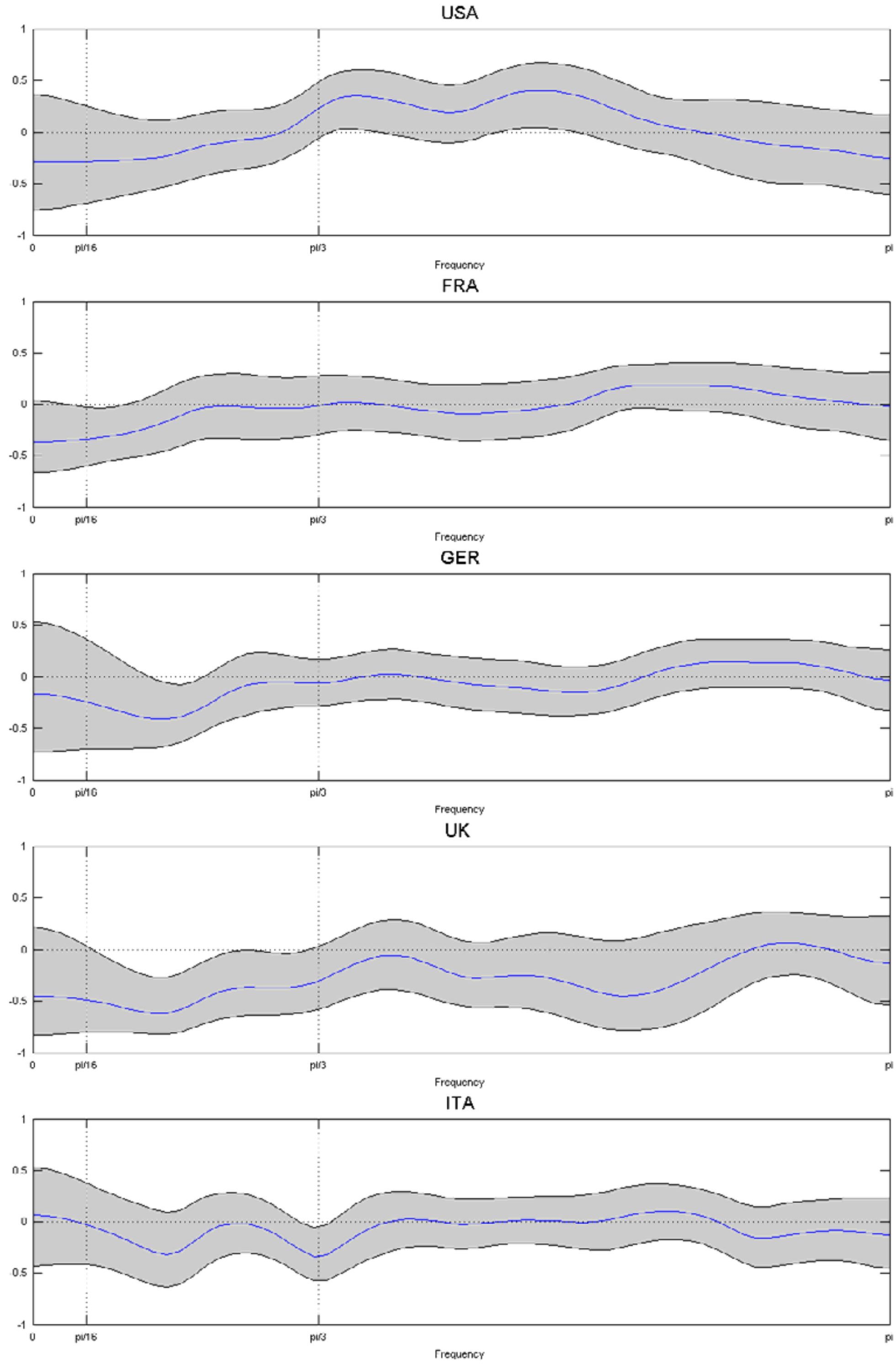
Graph 8

Dynamic correlation between consumption growth and excess returns



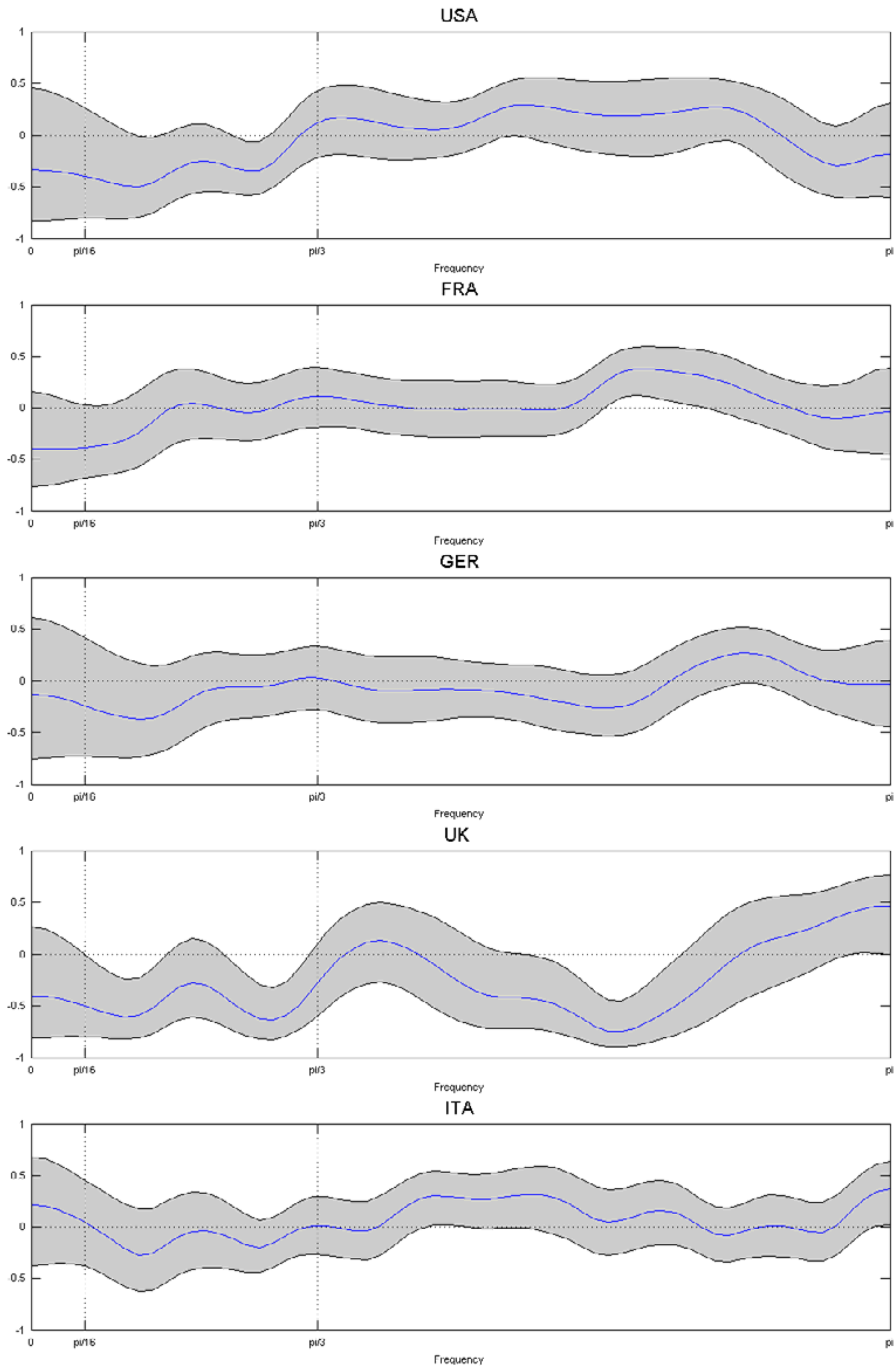
Graph 9

Dynamic correlation between GDP growth and money market rates



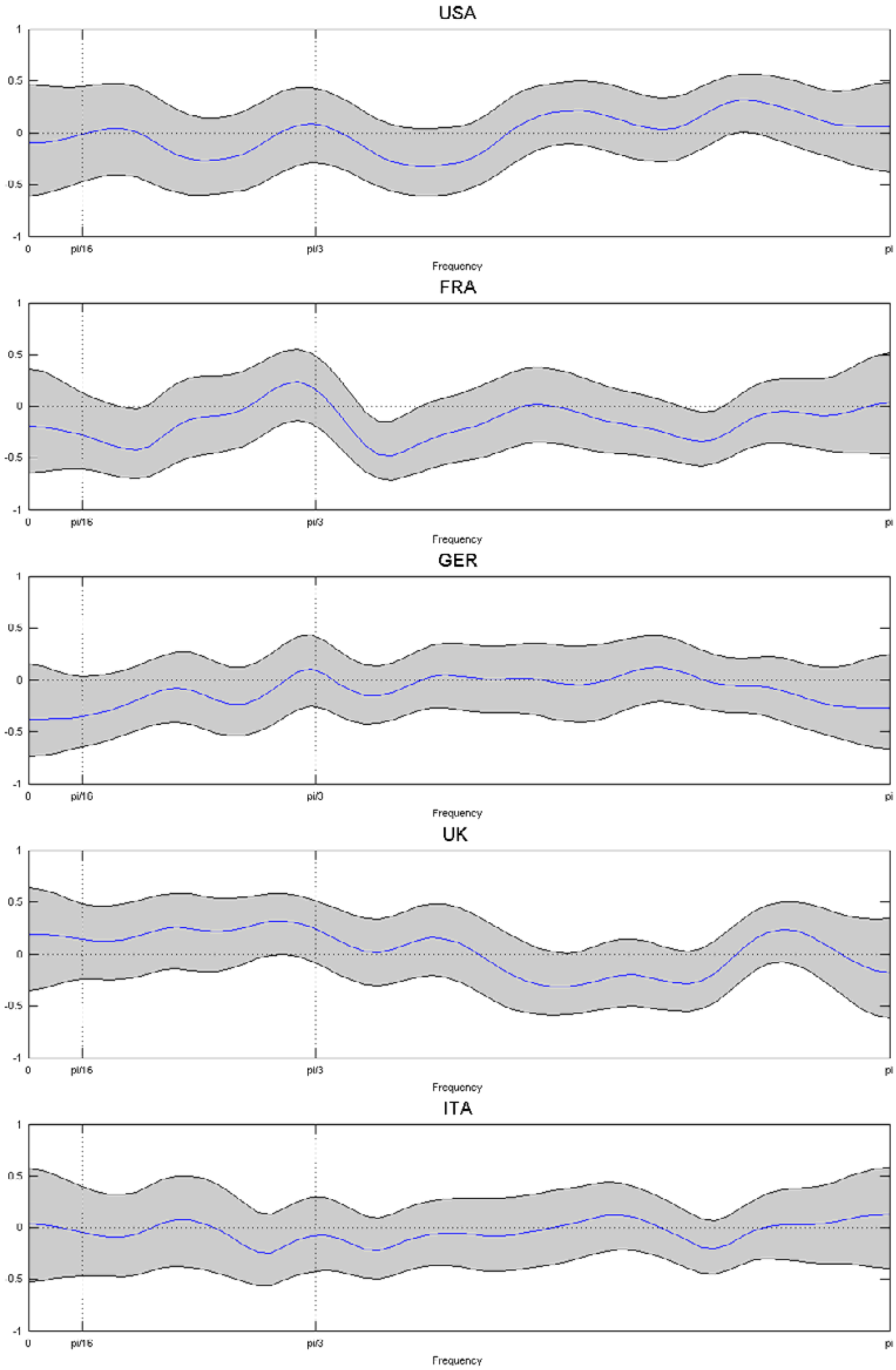
Graph 10

Dynamic correlation between consumption growth and money market rates



Graph 11

Dynamic correlation between excess returns and money market rates



Appendix A
Turning points and concordance

Bry and Boschan (1971) determined an algorithm that made it possible to replicate the contraction start dates identified by a committee of experts from the NBER. We used a variation of this algorithm, developed by Harding and Pagan (2002a,b), whose steps are as follows:

1. A peak/trough is reached at t if the value of the series at date t is superior/inferior to previous k values and to the following k values, where k is a positive integer that varies according to the type of series studied and its sampling frequency.[16]

2. A procedure is implemented to ensure that peaks and troughs alternate, by selecting the highest/lowest consecutive peaks/troughs.[17]

3. Cycles whose duration is shorter than the minimum time m are stripped out, as are cycles whose complete recurrence period (number of periods separating a peak from a peak or a trough from a trough) is lower than the prespecified number of periods M.

4. Complementary rules are applied:

   (a) the first peak/trough cannot be lower/higher than the first point in the series, and the last peak/trough cannot be lower/higher than the last point in the series;

   (b) the first/last peak/trough cannot be positioned at less than e periods from the first/last point in the series.

The monthly sales index is prefiltered using a Spencer curve, in accordance with the usual procedure adopted in the literature. The latter defines the filtered series $\tilde{x}_t$ from the raw series x_t according to:

$$\tilde{x}_t = \sum_{i=-7}^{7} s_i x_{t+i}, \quad s_i = s_{-i} \quad \textit{for} \quad i = 1,\ldots,7$$

$$s_0 = \frac{74}{320},\ s_1 = \frac{67}{320},\ s_2 = \frac{46}{320},\ s_3 = \frac{21}{320},\ s_1 = \frac{3}{320},\ s_1 = -\frac{5}{320},\ s_1 = -\frac{6}{320},\ s_1 = -\frac{3}{320}.$$

Note that, like Pagan and Sossounov (2003), we do not prefilter the monthly financial series. Moreover, in the latter case, imposing a minimum phase m may be restrictive. Pagan and Sossounov therefore propose relaxing the constraint on the minimum phase when a fall or a rise in excess of 20% is present in a period. We adopt this procedure here.

A contraction/expansion phase is thus defined as the time separating a peak/trough from a trough/peak, when the sequence of peaks and troughs meets all the identification rules listed above.

Note that the identification of turning points is very sensitive to the choice of parameters k, e, m and M: if the latter are set to small values, almost all absolute declines in the level of a series will be identified as troughs, all the more so as the original variable is not too smooth. On the other hand, if these are set to large values, the procedure will come up with almost no turning points.

The choice of k, e, m and M depends upon the series under consideration and their sampling frequency. For example, if y denotes logged real quarterly GDP, one generally sets $k = 2$, $e = 2$, $m = 2$ and $M = 5$. These values allow us to replicate the NBER business cycle dates.

[16] In this method used for identifying turning points, it is not necessary to assume that the series studied is stationary.

[17] This criterion is not always adopted in the literature (see Canova (1999)).

References

Andrews, D and C Monahan (1992): "An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator", *Econometrica*, vol 60, pp 953-66.

Artis, M, M Marcellino and T Proietti (2003): "Dating the euro area business cycle", *CEPR Discussion Paper no 3696*, CEPR.

Bernanke, B and M Gertler (2001): "Should central banks respond to movements in asset prices?", *American Economic Review*, vol 91, pp 253-57.

Bry, G and C Boschan (1971): "Cyclical analysis of time series: selected procedures and computer programs", *Technical Papers*, no 20, NBER.

Bullard, J and E Schalling (2002): "Why the Fed should ignore the stock market", *Economic Review of the Federal Reserve Bank of St. Louis*, vol 84, pp 35-41.

Canova, F (1999): "Does detrending matter for the determination of the reference cycle and the selection of turning points?", *Economic Journal*, vol 109, pp 126-50.

Cashin, P, J McDermott and A Scott (1999): "The myth of comoving commodity prices", *IMF Working Papers*, no 99/169.

Christiano, L and T Fitzgerald (2003): "The band pass filter", *International Economic Review*, vol 44, pp 435-65.

Croux, C, M Forni and L Reichlin (2001): "A measure of comovement for economic variables: theory and empirics", *Review of Economics and Statistics*, vol 83, pp 232-41.

Daniel, K and D Marshall (1998): "Consumption-based modeling of long-horizon returns", *Working Papers*, WP-98-18, Federal Reserve Bank of Chicago.

Estrella, A (2003): "Productivity, monetary policy and financial indicators", *working paper*, Federal Reserve Bank of New York.

Filardo, A (2000): "Monetary policy and asset prices", *Economic Review of the Federal Reserve Bank of Kansas City*, third quarter, pp 11-37.

Harding, D and A Pagan (2002a): "Dissecting the cycle: a methodological investigation", *Journal of Monetary Economics*, vol 49, pp 365-81.

——— (2002b): "Synchronisation of cycles", *working paper*, Melbourne University.

Kydland, F and E Prescott (1982): "Time-to-build and aggregate fluctuations", *Econometrica*, vol 50, pp 1345-70.

McDermott, J and A Scott (2000): "Concordance in business cycles", *IMF Working Papers*, no 0/37.

Odonnat, I and A-M Rieu (2003): "Dans quelle mesure la situation financière des ménages conditionne-t-elle leur consommation", *Revue de la stabilité financière*, no 3.

Pagan, A and K Sossounov (2003): "A simple framework for analysing bull and bear markets", *Journal of Applied Econometrics*, vol 18, pp 23-46.

Taylor, J (1999): *Monetary policy rules*, NBER and Chicago University Press.

Watson, M (1994): "Business-cycle durations and postwar stabilization of the U.S. economy", *American Economic Review*, vol 84, pp 24-46.

Business cycle, credit risk and economic capital determination by commercial banks

Alexis Derviz and Narcisa Kadlčáková[1]
Czech National Bank

1. Introduction

Regular assessments of the default risk of bank clients and estimations of credit risk at the portfolio level are becoming a necessity for banks in their daily operations. The design of optimal lending contracts and the need to conform to new regulatory trends constitute at least two reasons why banks have to pay closer attention to quantitative methods for assessing the credit risk of their clients. While primarily designed for use in commercial banks, credit risk models have recently started to attract the attention of other groups of economic professionals. It is the supervisory function of central banks that is mostly triggering the interest in examining credit risk models in this environment. In addition, an overall assessment of the creditworthiness of domestic firms has implications for the conduct of monetary policy. These and other reasons have prompted several central banks in Europe to develop and implement their own models for monitoring the financial situation of domestic firms and the lending performance of domestic banks.[2]

The objective of this paper is to develop an assessment technique for analysing the impact of different risk-based capital requirement rules on the potential needs for capital in the Czech banking sector. For this purpose, we apply these methods to an artificially constructed risky loan portfolio. The latter reflects a number of prominent features of Czech non-financial borrowers.

When defining the creditworthiness characteristics of the loan portfolio, we apply the Moody's KMV method for rating private firms. To determine capital requirements for this portfolio, we use the New Basel Capital Accord (NBCA) and the CreditMetrics and CreditRisk+ models. In the context of CreditMetrics, we are able to conduct stress testing to gauge the impact of interest rate uncertainty (eg caused by changes in monetary policy and different reactions of the yield curve to these changes) on economic capital calculations. In addition, we describe an independent debt valuation model similar to that of KMV and outline the techniques for its numerical implementation. The proposed model has a substantial advantage over the previously mentioned ones in that it addresses three key problems of credit risk modelling. Namely, this model, although remaining in the KMV line of analysis:

- incorporates macroeconomic systemic factors, such as position in the business cycle, interest rate and exchange rate volatility and the monetary policy stance, when deriving a valuation of bank lending risks;
- combines the features of structural and reduced-form models of debt valuation;
- offers a framework for assessing the influence of market risk factors on credit risk in a bank loan portfolio.

In the latter respect, our model advances towards an *integrated financial risk assessment methodology*, which has recently been called for in the risk analysis literature (see, for instance, Barnhill and Maxwell (2002), or Hou (2002)).

The principal feature of the paper is a comparative analysis of the predictions of these models when applied to an artificially created loan portfolio constructed using Czech data. Another important

---

1 Corresponding author: Alexis Derviz, Czech National Bank, Monetary and Statistics Dept, International Economic Analyses Division, Na Prikope 28, CZ 115 03 Praha 1, Czech Republic.

2 Rating systems and creditworthiness assessment models for firms have been developed, among others, by the central banks of Austria, France, Germany, Italy and the United Kingdom.

contribution is a demonstration, even if in an incipient manner, of the way in which scarce and usually unavailable variables can be estimated or proxied to obtain the inputs required by the credit risk models. One oft-mentioned drawback of credit risk modelling is the difficulty with which the credit risk analyst can access the required input data. This problem was also present in our case. Although overcome, the data problem has had negative implications for the robustness of our results. Thus, from this perspective we have to look at the paper's findings with caution. However, the insight into credit risk modelling that is offered here can be extended at a later stage when more data is available. We also hope that our findings may be of use to banking supervisors when these issues become a matter of regulatory practice.

Although credit risk models often prove useful for other purposes, their main merit rests in estimating the capital level that banks have to maintain over the given risk horizon. The outcome is called regulatory capital in regulatory terms and economic capital in terms of credit risk modelling. Both regulatory and economic capital are supposed to cover unexpected losses resulting from banks' lending operations to clients with different levels of default risk. Whereas holding regulatory capital is compulsory as a part of adherence to prudential regulations, holding economic capital beyond the minimum required level is the banks' own choice. Worth mentioning, however, is the regulatory tendency to come closer to credit risk modelling and to allow banks to develop their own models for determining the amount of regulatory capital to hold. These models will most probably adopt and synthesise many features of the credit risk models already in use. This is one reason why comparing regulatory and economic capital today is becoming an insightful exercise for the regulatory decisions of the future.

1.1 Literature review

In June 1999, the Basel Committee on Banking Supervision released a proposal to replace the 1988 Basel Capital Accord with a more risk-sensitive framework. A concrete proposal in the form of a consultative document, the New Basel Capital Accord (NBCA), was presented in January 2001. This document proposed new regulatory rules for banks' capital adequacy evaluations. The main innovations related to credit and operational risk. In terms of credit risk, the NBCA revised the 1988 Accord by proposing a more risk-sensitive methodology for assessing the default risk of banks' clients. The risk inputs entering the final capital adequacy computations were closely related to the risk characteristics of individual bank clients. In this sense, the proposed methodology opted for the adoption of ratings (developed by external agencies or by banks themselves) in quantifying and signalling to the bank the default risk of individual borrowers. In a simpler version of the methodology (the standardised approach), ratings are directly associated with risk weights (for example, an A-rated asset would be assigned a risk weight of 50%, a BBB-rated asset would be assigned a risk weight of 100%, and so on). In the more advanced internal ratings-based (IRB) approach, ratings represent the basis for computing the probability of an obligor's default. Default probabilities and other risk characteristics (loss-given-default, exposure at default) enter more complicated formulas for determining the risk weights of individual assets in regulatory capital estimations.

In the banking industry, credit risk modelling has also been explored and extended since the release of the four major credit risk models at the end of the last decade.[3] In this paper we consider only two such models, CreditMetrics and CreditRisk+, which utilise, respectively, the structural and the reduced-form approach to modelling default risk (see Duffie and Singleton (1998)). In the structural approach, it is assumed that default is triggered when an unobserved variable (obligor's firm asset value) falls below a certain threshold level (firm's outstanding debt). CreditMetrics extends this reasoning to rating downgrades by defining rating class-specific threshold levels that mark the switch from one rating class to another in the event that the firm's standardised asset returns cross these threshold values. In the reduced-form literature, default is modelled as an autonomous stochastic process that is not driven by any variable linked to the obligor firm's capital structure or asset value. Particular formulations for the default process were considered in Jarrow and Turnbull (1995; exponential distribution), Jarrow et al (1997; a continuous Markov chain) and Duffie and Singleton (1998; a stochastic hazard rate process). CreditRisk+ represents the reduced-form approach by

[3] We refer to JPMorgan's CreditMetrics/Credit Manager model, Credit Suisse Financial Products' CreditRisk+, KMV Corporation's KMV model, and McKinsey's CreditPortfolioView.

assuming that the average number of defaults in each homogeneous class of obligors follows a Poisson distribution. The unifying element of the CreditMetrics and CreditRisk+ models is the value-at-risk (VaR) methodology used in quantifying and provisioning for credit risk at the portfolio level. Even though CreditMetrics derives the portfolio value distribution and CreditRisk+ the portfolio loss distribution at the end of the risk horizon, both models estimate economic capital such that unexpected losses are covered by the estimated economic capital within an acceptable confidence level.

The KMV model represents another step towards market-based derivation of economic capital. Similarly to CreditMetrics, it uses the obligor's equity price statistics to derive the value distribution of a given loan. Correlations are obtained automatically from the risk factors that determine the obligor firm value (equity). However, this method requires the assumption of complete markets, the validity of risk neutral asset valuation and tradability of both the obligors' equities and their debt in the bank portfolio. The KMV team offers unspecified remedies in cases where one of these preconditions is not satisfied, but open sources of credit risk literature offer no general solution of these problems.

This paper proposes a way around the said difficulties in the KMV approach by resorting to the so-called pricing kernel methods of asset pricing (comprehensive expositions can be found in, for instance, Campbell et al (1997) and Cochrane (2001)). Asset tradability and market completeness are no longer necessary, and there are numerous possibilities for modelling default events that depend on systemic and idiosyncratic risk factors. Numerical approaches to calculating pricing kernel-based asset values have also been developed in recent years (see, for instance, Ait-Sahalia and Lo (2000), or Rosenberg and Engle (2002)).

1.2 Methodology

Three pillars make up the main structure of our analysis. First, a tested bank loan portfolio is constructed in such a way as to reflect with some degree of realism the rating distribution of a pool of Czech bank clients. Second, we take into account the random nature of interest rates and other economic fundamentals that enter the loan valuation. Among other things, this means that market risk factors (interest rates and exchange rates) were an integral part of the capital calculations as far as each of the tested approaches allowed. Third, when conducting model-based economic capital calculations, we follow the market loan pricing point of view wherever possible (ie when the corresponding model allows it either explicitly or implicitly). This is done because we want to identify those elements of capital requirements which may be seen differently from the credit risk modelling and regulatory perspectives.

Our analysis utilises a hypothetical portfolio containing 30 loans. This simplified portfolio mirrors the rating structure of a real loan portfolio obtained on the basis of a pool of corporate customers of six Czech banks.[4] Since ratings are the key input in many credit risk approaches, a simplified version of Moody's rating methodology for private firms has been applied to obtain ratings in our real sample of bank clients. Estimates of other inputs required by credit risk modelling which were not available in the real bank data set were obtained using aggregate data from Czech National Bank (CNB) databases.

In an earlier paper (Derviz et al (2003)), we examined and compared the predictions of the NBCA with those delivered by the CreditMetrics and CreditRisk+ models. Following the January 2001 consultative version of the NBCA guidelines we found that in our particular example the standardised approach of the NBCA predicted approximately the same level of capital as the credit risk models at the 95% confidence level. At the 99% confidence level, the internal credit risk models predicted a higher level of economic capital than the NBCA standardised approach, but these estimates were still lower than the estimates of the NBCA IRB approach. We obtained different results when applying the NBCA guidelines as formulated by the third quantitative impact survey (QIS 3, October 2002). Here, the results of both NBCA approaches (standardised and IRB) were more similar to each other, with the IRB requirement being slightly lower than that of the standardised approach. The results of both regulatory approaches were even lower than the level of capital required by the various credit risk models. For ease of reference, we reproduce the regulatory capital results along with the modelled economic capital ones at the end of the present paper (Table 8).

4 We would like to thank Alena Buchtíková for making this data set available for our research purposes.

In the context of the CreditMetrics model, we have extended the analysis of economic capital by allowing both the bank lending rates and the forward zero coupon rates used as discount factors in asset valuations to become random variables (a form of stress testing; see the details in Appendix A). Floating lending rates and random changes in the forward curves were all implemented using Monte Carlo simulations. As expected, more uncertainty associated with the evolution of these variables required more economic capital to be held by banks. However, the proposed changes in forward zero curves did not impose significantly different levels of credit risk-related economic capital as compared with the case of stable forward yield curves (but maintaining floating interest rates in both cases). Downward movements in the forward zero curves required higher levels of capital at all confidence levels. This is an inconvenient consequence of the existing credit models (CreditMetrics and KMV in particular) which we strive to overcome by proposing a model of our own.

Our approach to modelling financial and real uncertainties is similar to Ang and Piazzesi (2003), although we do not orient our state-space estimation on fitting the observed yield curve. Instead, we estimate the pricing kernel parameters that fit the returns of a number of basic infinite maturity assets. The reason for this is that we need a direct connection between macroeconomic risk factors, asset prices and bank loan values. Looking for this connection through the prism of yield curve dynamics would be too circumspect for our purposes, since extracting business cycle information from the yield curve is a misspecification error-laden process in itself. In contrast, by allowing the model to reflect a one-to-one correspondence between a vector of basic assets and another vector of unobserved factors, we are likely to capture a latent principal component responsible for the economic activity. We believe that this estimated model, which we later use for simulations, contains less noise in the identified business cycle position of the economy than most multifactor yield curve models in the literature.

The paper is structured as follows. In Section 2 we briefly describe the main characteristics of the real bank and test portfolios and their estimation. We also mention the reasons why the models could not be implemented entirely on the basis of real Czech bank data. In Section 3 we outline the methodologies proposed by two popular credit risk models (CreditMetrics and CreditRisk+) and present their economic capital estimations. Section 4 outlines our own model of risky debt, its valuation and the resulting economic capital requirements, going along the structural lines of the original KMV model. Section 5 discusses estimation procedures and outcomes. Section 6 concludes.

2. The test portfolio

Our bank data set contains the balance sheets and profit and loss accounts of non-financial firms that were granted bank loans between 1994 and 2000. The CZ-NACE classification,[5] legal form and CNB loan classification[6] (from 1997) were also available for each bank customer. Six Czech banks provided the data to the CNB from 1994 until 1999, of which two banks terminated cooperation in 2000. The banks reported only a fraction of their corporate portfolios. The exact selection procedure used by banks to choose particular firms is not known. Also unknown is the proportion of reported versus unreported clients satisfying certain criteria. In this sense, we observed a certain bias of the data providers towards non-reporting of loans in the last two categories (4 and 5) but were not able to assess the direction and magnitude of this sampling bias in our results.

Since our main goal was to assign ratings to banks' corporate clients, we primarily focused on their default behaviour. Default was defined as a credit event in which the loan classification of a certain company migrated from the first or second category to any of the third, fourth or fifth categories over the considered risk horizon. Due to the short time length of our data set we had to focus on annual default rates. The largest number of defaults occurring over a one-year period was recorded between

[5] CZ-NACE (Czech abbreviation: OKEČ) represents the industry classification of economic activity in the Czech Republic.

[6] The CNB's loan classification ranges from 1 to 5, with category 1 meaning standard, 2 watch, 3 non-standard, 4 doubtful and 5 loss loans.

1997 and 1998, representing 8% of all firms in the sample.[7] The sample-based annual default rates were 0.07% between 1998 and 1999 and 0% between 1999 and 2000. These low default rates may be partially explained by the Czech economic recovery and by more prudent bank lending behaviour during 1999-2000. Nevertheless, we think that the main reason is insufficient default reporting by banks. Therefore, we preferred to restrict the reference data set only to the accounting information collected in 1997 and the default events observed in 1998, assuming that default reporting by banks in that period was closer to the reality. While analysing a longer time period would have been highly valuable, we considered that the sample-based information over 1998-2000 painted a biased picture about corporate default and, consequently, it was not used in modelling the rating structure of the test portfolio.

The annual default rate of 8% in the reference data set was significantly higher than the average value of 1.5% usually used by Moody's in the context of western European economies. However, volume-based information about the loan defaults of individuals and corporates in the entire banking sector revealed an annual default rate of approximately 20%. We considered that neither 1.5% nor 20% would be the appropriate annual default rate for our artificial portfolio[8] and, in general, for a typical Czech bank portfolio of corporate loans. Without any other more reliable source of information, we used the 8% default rate as indicated by our sample bank data to calibrate the probit model.

To assign ratings to each firm we used the calculated default rates and the tables containing cumulative default rates published by different rating agencies. Even though Moody's rating methodology was used, we preferred to calibrate our results to Standard & Poor's (S&P) ratings. This was done since (a) the NBCA assigned risk weights based on S&P ratings and (b) our inputs into the credit risk models were, to a great extent, based on S&P data. For calibration purposes we used the 1996 S&P cumulative one-year default rate matrix shown in Table 1.

Table 1

S&P's rating class-specific one-year default rates

Rating	One-year default rate (%)	Cutoff values for defining ratings in the bank portfolio (%)
AAA	0	–
AA	0	–
A	0.06	0.03
BBB	0.18	0.12
BB	1.06	0.62
B	5.2	3.13
CCC	19.79	12.495

Source: CreditMetrics - Technical Document.

The probabilities given in the second column are rating class-specific default probabilities published by S&P. The third column contains the probabilities that mark the transition from one rating class to another in our model. They are the midpoints in the intervals determined by the one-year default probabilities given in the second column. For example, if the estimated default probability of a certain firm belonged to the interval [0, 0.03), an AA rating grade was assigned to that firm. If the estimated

[7] The reference sample included only bank clients that were present both in 1997 and 1998 (663 firms). To examine only one-year default behaviour, those enterprises which were already in default in 1997 were also eliminated. In the end we obtained a data set containing 606 firms.

[8] These figures reflected assumptions that were not applicable in our case. For example, the 1.5% level was based on the western experience, while the 20% level was volume-based and represented both firms' and individuals' default behaviour.

default probability fell into the interval [0.03, 0.12), then an A grade was assigned and so on. Based on this mapping procedure, each firm present in the 2000 data set was marked with a certain rating grade. The resulting rating structure and the loan classification of the pooled bank portfolio for 2000 are presented in Table 2.

Table 2

The pooled bank portfolio structure in 2000 according to loan classification and ratings (number of firms/percentage)

	AA	**A**	**BBB**	**BB**	**B**	**CCC**	**Total**
1	15/1.45	5/0.48	26/2.51	87/8.41	580/56.04	89/8.60	802/77.49
2	1/0.10	0	1/0.10	19/1.84	150/14.49	23/2.22	194/18.74
3	0	0	0	0	12/1.16	11/1.06	23/2.22
4	0	0	0	0	3/0.29	3/0.29	6/0.58
5	0	0	0	0	1/0.10	90.87	10/0.97
Total	16/1.55	5/0.48	27/2.61	06/10.24	746/72.08	135/13.04	**1,035/100**

Next, we constructed an artificial portfolio incorporating as much real information as possible. We could not perform our risk capital estimations on the pooled bank portfolio because (a) this would have been extremely time-consuming and (b) many inputs required by the credit risk models were not available for the real bank customers. For instance, outside the ratings, the bank data set did not contain information regarding loan volumes or maturities, charged interest rates and borrower asset returns or recovery rates. Since these parameters represent required inputs into many regulatory and internal credit risk models, we constructed proxy variables based on data available at the macro level or obtained them as random drawings from known distributions. While in a reduced portfolio (like our testing one) the construction of the proxy variables is easily done, this construction would have been more difficult to produce based on a portfolio of 1,035 bank clients. In what follows we describe the manner in which these inputs were generated. The main information source was the CNB supervisory database, which contains yearly data on residual maturity of Czech bank loans, their category and the borrower CZ-NACE code. It also categorises loans according to the charged interest rate.

Ratings and exposures

The rating structure displayed in Table 2 was adjusted to reflect the following changes:

- All bank clients in loan category 5 (loss loans) were eliminated. Such loans are usually covered by provisions created in the current period. Moreover, the fifth category is an absorbing state: a loan falling into this category has a negligible probability of recovery. These loans pose a vacuous problem from the risk management perspective, since their future status is not associated with any uncertainty.
- The 8.6% of firms with a CCC rating were removed from category 1 and added to categories 3 and 4. We assumed that the 8.6% outcome reflected the imperfections of our model. Czech banks monitor the creditworthiness of their clients, thus loans falling in the first category are unlikely to be granted a CCC grade.
- The rating structure was adjusted to resemble the loan volume configuration at the end of December 2000 as closely as possible (as shown in Table 3).

Now the rating structure of our test portfolio takes the form shown in Table 4. To have a fair representation of all ratings in each loan category, we needed a minimum of 30 assets.

The exposure of an asset in a certain loan category represents the ratio of the total loan volume to the number of assets in that category. For example, all assets belonging to category 1 (21 in the test portfolio) would have an exposure of CZK 607.235 billion/21 = CZK 28.915 billion.

Table 3

Loan volumes by category granted by Czech banks, as of end-2000

Category	Volume (CZK bn)	Proportion (%)
1	607.235	68.58
2	85.811	9.69
3	54.577	6.16
4	26.982	3.05
5	110.834	12.52

Source: Czech National Bank.

Table 4

Rating structure of the artificial portfolio: number of assets in each loan category and rating class

Loan category	AA	A	BBB	BB	B	CCC	Total
1	1	1	2	3	14	0	21
2	0	0	0	1	2	1	4
3	0	0	0	0	1	2	3
4	0	0	0	0	0	2	2
Total	1	1	2	4	17	5	30

Maturities, lending and recovery rates

Loans with a maturity exceeding five years are sparsely represented in the Czech bank portfolios. For this reason, we considered maturities that ranged from one to five years only. Maturity was assigned to individual assets by drawing random numbers from the interval [1, 5] according to the uniform distribution and then rounding these numbers to the nearest integer.

We computed the mean and standard deviation of the lending rates for each loan category (using the SUD data set). To assign lending rates to assets in our portfolio, we randomly drew numbers from normal distributions described by the estimated means. Standard deviations were in general reduced to prevent interest rates from deviating too much from these mean values.

We also generated collateral and recovery rates (see Derviz et al (2003) for details). The main characteristics of the artificial test portfolio are shown in Table 5.

Asset return correlations

Firms in the bank data set were grouped according to ratings and loan classification. Then we found the CZ-NACE category that was the most frequently represented in each group. For example, in the group of firms with rating AA and loan classification 1 the largest number of firms belonged to CZ-NACE 51. If in a certain group no dominating CZ-NACE could be found, we randomly selected the representative figure from those that were present in that group. The resulting CZ-NACE structure was mapped to the test portfolio. Having assigned a CZ-NACE label to each asset in the test portfolio, we used the price index characteristic of the corresponding branch as a proxy variable for that asset's returns.[9] Asset return correlations were determined by computing correlations among price indices.

Table 5

Portfolio composition and individual loan characteristics

Loan	CZ-NACE	Regulatory loan class	Rating	Loan volume[1]	Maturity	Lending rate (%)	Type of collateral	Recovery rate (%)
1	51	1	AA	28.916	3	6.5	6	0
2	36	1	A	28.916	1	7.3	6	0
3	74	1	BBB	28.916	3	7.5	6	0
4	31	1	BBB	28.916	5	7.6	6	0
5	74	1	BB	28.916	5	8.2	4	71.43
6	20	1	BB	28.916	5	8.5	1	94.34
7	51	1	BB	28.916	1	8.7	6	0
8	28	1	B	28.916	3	8.8	2	100
9	15	1	B	28.916	4	9.5	6	0
10	51	1	B	28.916	2	10.5	1	94.34
11	52	1	B	28.916	2	11.0	5	50
12	29	1	B	28.916	1	11.1	6	0
13	70	1	B	28.916	1	11.3	5	50
14	74	1	B	28.916	2	11.6	4	71.43
15	50	1	B	28.916	2	11.7	1	94.34
16	24	1	B	28.916	1	11.9	5	50.00
17	45	1	B	28.916	2	12.1	1	94.34
18	60	1	B	28.916	2	12.2	1	94.34
19	40	1	B	28.916	3	12.5	3	89.29
20	25	1	B	28.916	2	12.6	1	94.34
21	65	1	B	28.916	4	13.0	4	71.43
22	51	2	BB	21.452	2	8.4	1	94.34
23	25	2	B	21.452	5	10.4	1	94.34
24	29	2	B	21.452	3	11.5	6	0
25	55	2	CCC	21.452	3	13.5	5	50
26	45	3	B	18.192	2	11.9	1	94.34
27	28	3	CCC	18.192	5	12.4	1	94.34
28	21	3	CCC	18.192	4	12.2	1	94.34
29	21	4	CCC	13.491	5	14.3	4	71.43
30	37	4	CCC	13.491	2	15.4	1	94.34

[1] In billions of Czech korunas.

[9] At the outset, all price indices were deflated by the PPI in order to eliminate the systemic inflationary influence in their evolution.

3. Portfolio value and economic capital according to commercial risk measurement models

3.1 CreditMetrics

In the CreditMetrics model, risk is associated with changes in the portfolio value caused by changes in the credit quality of individual obligors (downgrades or default) over the considered risk horizon (usually one year). We have followed the two standard pillars of the CreditMetrics approach. The analytical pillar requires a derivation of primary risk measures such as means, variances and standard deviations at the asset and portfolio level. The estimation pillar implies generating a simulated portfolio value distribution at the risk horizon. Based on this distribution, estimates of economic capital can be obtained at different confidence levels. Application of the Monte Carlo simulation method to our portfolio in accordance with the CreditMetrics approach is described next.

Monte Carlo simulation

Individual random draws from a multinomial normal distribution ("scenarios" in CreditMetrics terminology) of asset returns contain the same number of components (real numbers) as the number of assets in the portfolio. Each component of each scenario is compared with predefined, rating class-specific threshold values marking the switch from one rating class to another. In this way, within each scenario, a new rating is assigned to each asset (obligor) in the portfolio.

In case of non-default, each asset i is revalued according to the formula:

$$V_i^{g'} = \sum_{t=1}^{T_i-1} \frac{r_i}{\left(1+f_t^g\right)^t} + \frac{r_i + F_i}{\left(1+f_{T_i}^g\right)^{T_i}}, \tag{1}$$

where r_i and F_i are the loan interest payments and the face value of the loan respectively, and f_t^g are the annualised forward zero rates for the years one to T_i, applicable to the rating class g' (here T_i is the maturity of the loan). In this specification it is assumed that the present rating changes from g to g' over the one-year period. In the event of default the present value of the loan is computed as the product of the face value of the loan and a recovery rate.

Note that the way of defining the future loan value as a random variable implies that the distribution of this variable would be taken with respect to the *risk neutral probability* (RNP). CreditMetrics works with the assumption that such a probability is well defined for the studied economy (we take a different view in our own model; see subsections 4.1 and 4.2). Under the risk neutral probability, in contrast to the "physical" one, zero forward rates are unbiased estimates of the future spot interest rates. That is, the capital requirements under CreditMetrics are also derived from the RNP. This might lead to certain discrepancies between the CreditMetrics interpretation by the market (which is based on the RNP) and its interpretation by the regulator (based on the physical probability).

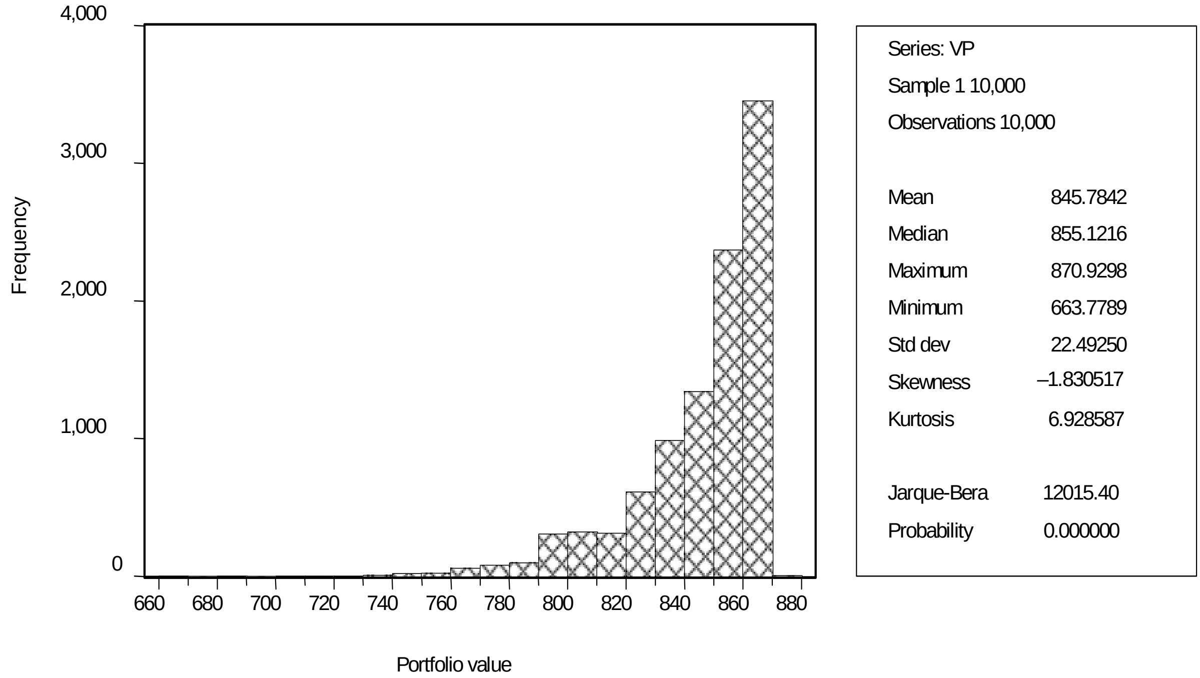
To obtain the portfolio value distribution and derive the economic capital requirement in accordance with the CreditMetrics model, we conducted a Monte Carlo simulation with 10,000 random draws. Each scenario contained 30 correlated random draws from the standard normal distribution. Each element of each scenario represented a standardised return corresponding to one of the 30 assets belonging to the portfolio. Comparing the elements of the scenario with the threshold values characteristic of each rating category, new ratings were assigned to each of the 30 assets. Summing up the values of the 30 assets thus obtained, a new portfolio value resulted for each scenario. Since eight potential new ratings were possible for each of the 30 assets at the end of the year, the total number of potential portfolio values was 8^{30}. In practice, this number was far lower, as some rating class migrations had a zero probability of realisation. Figure 1 shows the distribution of our portfolio value expected at the end of 2000 for the year 2001. On the horizontal axis are the non-overlapping intervals within which the portfolio value falls, while on the vertical axis are the frequencies with which these realisations occurred within each interval in our simulation.

The economic capital is obtained as the difference between the mean of the portfolio value and a p-percentile (p is usually assumed to be 1%, 2% or 5%):

Economic capital = Mean of the portfolio distribution – p-percentile

Figure 1

Monte Carlo simulation for the loan portfolio value



In the case of a discrete portfolio distribution, the p-percentile is obtained by looking at the lowest portfolio value whose cumulative frequency exceeds p%. An interpretation of the p-percentile is that in p% of cases we can expect the portfolio value to take values lower than the p-percentile over the one-year period. For example, in our case we can expect that only 100 times in 10,000 cases could the portfolio value reach a value lower than CZK 767.89 billion (in other words we know with 99% probability that the portfolio value will be higher than CZK 767.89 billion at the year-end). To cover this high loss, however unlikely it is, the bank must keep economic capital of CZK 77.88 billion.

The 1%- and 5%-percentiles take the values CZK 767.89 billion and CZK 796.62 billion respectively. Based on these estimations, the bank's need for economic capital at different confidence levels is:

1%-percentile	**5%-percentile**	**Mean**	**99% economic capital**	**95% economic capital**
767.89	796.62	845.78	77.88	49.16

Macroeconomic fundamentals and the CreditMetrics-based economic capital

The nature of CreditMetrics makes it rather difficult to analyse the impact of shocks to real economic activity on the economic capital allocation. The most natural type of systemic factor analysis that can be conducted in the context of this model concerns the interest rate uncertainty. Subsequently, different business cycle developments can be accommodated in a CreditMetrics environment by assigning them an appropriate change in the yield curve. In Appendix A, we report the results of the corresponding Monte Carlo simulations for our artificial portfolio.

It turns out that the economic capital values are very sensitive to the yield curve movements, and often exhibit a counter-intuitive reaction on certain interest rate developments. For instance, an average downward (clockwise) rotation of the yield curve, which corresponds to the expected loosening of monetary policy, leads to an economic capital increase. This is an inevitable consequence of the fact that CreditMetrics does not work with the duration characteristics of the loans. The model to be described in Section 4 strives to overcome this deficiency.

3.2 CreditRisk+

CreditRisk+ is suitable for assessing credit risk in portfolios containing a large number of obligors with small default probabilities. The model groups bank customers according to their common exposure. The common exposure of an obligor represents the ratio between his bank exposure and a selected unit of exposure (CZK 1 billion in our case). Bank clients can be grouped in homogeneous "bands" that contain obligors with the same common exposure.

For obligor i, we introduce the following notations, taken from the CreditRisk+ technical document: L_i - exposure, P_i - default probability, $\nu'_i = \frac{L_i}{L}$ - common exposure, ν_i - rounded common exposure, $\varepsilon_i = \nu'_i \times P_i$ - expected loss. In addition, at the given Band j-level, ν_j is the common exposure in units of L, $\varepsilon_j = \sum_{i \in Band\ j} \varepsilon_i$ - expected loss in Band j in units of L, $\mu_j = \frac{\varepsilon_j}{\nu_j}$ - expected number of defaults in Band j.

The risk assessment at the asset level consists in estimating the expected loss (ε_i). At the band level, the model estimates the average number of defaults (m_j) as the ratio of the total expected loss in the band (ε_j) to the common exposure characteristic to the obligors from that band (ν_j).

By assumption, the distribution of the number of defaults in each band is of the Poisson type:

$$P_j = P(\textit{number of defaults in Band } j = k) = \frac{m_j^k e^{-m_j}}{k!} \quad k = 0, 1, \ldots$$

Further, using the properties of probability-generating functions, the model estimates recursively the probabilities that the portfolio loss reaches values expressed as multiples of the unit of exposure.

For our test portfolio, the analytical risk assessments at the asset and band level are contained in Tables 6 and 7. Specifically, Table 7 illustrates the partition of the portfolio into bands and the risk characteristics of each band. Four different bands have thus been obtained, with rounded common exposures of 14, 19, 22 and 29.

For each band a probability-generating function is given by:

$$G_j(z) = \sum_{n=0}^{\infty} \Pr\{L_j = n\} z^n = \sum_{k=0}^{\infty} P(k \textit{ defaults}) z^{k\nu_j} = \sum_{k=0}^{\infty} \frac{m_j^k e^{-m_j}}{k!} z^{k\nu_j} = e^{-m_j + m_j z^{\nu_j}}.$$

The probability-generating function for the entire portfolio is the product of the individual probability-generating functions displayed in the last column of Table 7. In this particular example we get:

$$G(z) = e^{-\sum_{j=1}^{m} m_j + \sum_{j=1}^{m} m_j z^{\nu_j}} = e^{-1.877 + 0.381 z^{14} + 0.429 z^{19} + 0.305 z^{22} + 0.762 z^{29}}.$$

To derive the probabilities that loss equals multiples of the unit of exposure, CreditRisk+ constructs a recurrence relationship:

$$P_n = \sum_{\substack{j \\ s.t. \nu_j \le n}} \frac{\nu_j \times m_j}{n} P_{n-\nu_j}$$

that starts with the probability of no loss:

$$P_0 = P(\text{No loss}) = e^{-m} = e^{-\sum_{j=1}^{m} m_j} = e^{-1.877} = 0.153$$

The resulting loss distribution in the case of our portfolio is shown in Figure 2.

The economic capital is given by the difference between the p-percentile and the expected mean of the loss distribution:

Economic capital = p-percentile – Expected loss.

Table 6

Risk assessment at the asset level according to CreditRisk+

Asset *i*	**Exposure (CZK bn)**	**Common exposure (ν_i')**	**Common exposure rounded to multiples of CZK 1 bn (ν_i)**	**Probability of default (P_i)**	**Expected loss ($\varepsilon_i = \nu_i' \times P_i$)**
1	28.92	28.92	29	0	0.00
2	28.92	28.92	29	0.0006	0.02
3	28.92	28.92	29	0.0018	0.05
4	28.92	28.92	29	0.0018	0.05
5	28.92	28.92	29	0.0106	0.31
6	28.92	28.92	29	0.0106	0.31
7	28.92	28.92	29	0.0106	0.31
8	28.92	28.92	29	0.052	1.50
9	28.92	28.92	29	0.052	1.50
10	28.92	28.92	29	0.052	1.50
11	28.92	28.92	29	0.052	1.50
12	28.92	28.92	29	0.052	1.50
13	28.92	28.92	29	0.052	1.50
14	28.92	28.92	29	0.052	1.50
15	28.92	28.92	29	0.052	1.50
16	28.92	28.92	29	0.052	1.50
17	28.92	28.92	29	0.052	1.50
18	28.92	28.92	29	0.052	1.50
19	28.92	28.92	29	0.052	1.50
20	28.92	28.92	29	0.052	1.50
21	28.92	28.92	29	0.052	1.50
22	21.45	21.45	22	0.0106	0.23
23	21.45	21.45	22	0.052	1.12
24	20.38	20.38	22	0.052	1.12
25	21.45	21.45	22	0.1979	4.25
26	18.19	18.19	19	0.052	0.95
27	18.19	18.19	19	0.1979	3.60
28	18.19	18.19	19	0.1979	3.60
29	13.49	13.49	14	0.1979	2.67
30	13.49	13.49	14	0.1979	2.67

Source: Own computation.

Table 7

Band partition and risk assessment at the band level according to CreditRisk+

Band j	**Rounded common exposure in Band j (ν_j)**	**Number of obligors in Band j**	**Expected loss in Band j (ε_j)**	**Expected number of defaults in Band j ($m_j = \varepsilon_j / \nu_j$)**	**Probability-generating function for Band j**
1	14	2	5.340	0.381	$\exp(-0.381 + 0.381z^{14})$
2	19	3	8.147	0.429	$\exp(-0.429 + 0.429z^{19})$
3	22	4	6.704	0.305	$\exp(-0.305 + 0.305z^{22})$
4	29	21	22.092	0.762	$\exp(-0.762 + 0.762z^{29})$

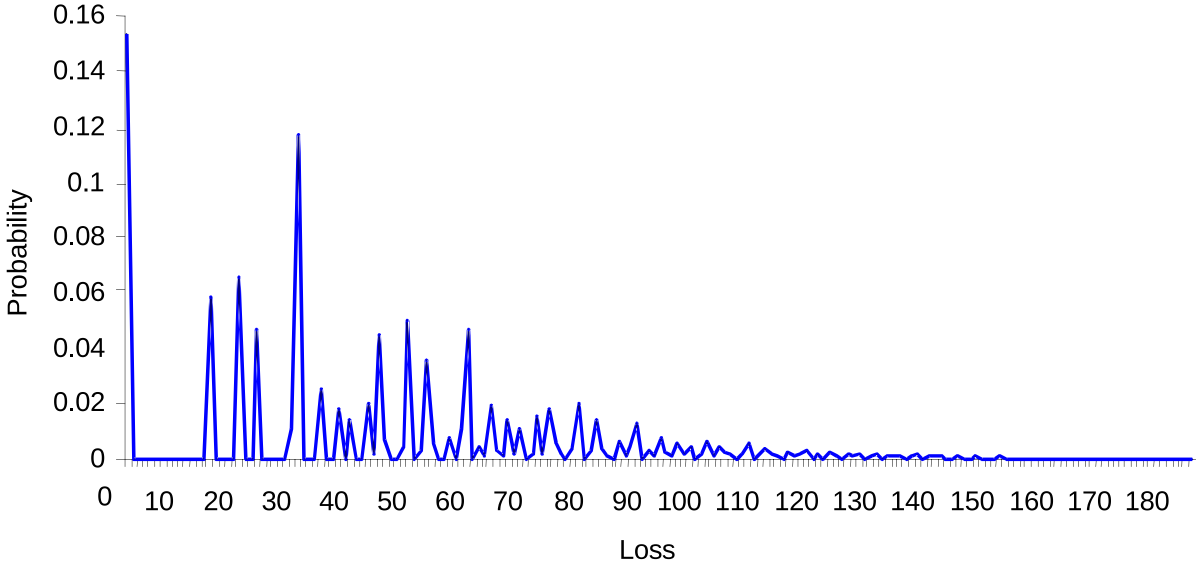
Source: Own computation.

Applying the CreditRisk+ approach to our portfolio, we got estimates of risk capital at different confidence levels as shown below.

1%-percentile	**5%-percentile**	**Expected loss**	**99% capital**	**95% capital**
133	**101**	**42.18**	**90.82**	**58.82**

Figure 2

Loss distribution based on the CreditRisk+ model



4. A structural model of risky debt with a random default arrival

We next give an outline of a model of risky debt, its valuation and the resulting economic capital requirements, going along the "structural" lines of the original KMV model and its ramifications. The term "structural" means that we make the default explicitly dependent on loan and obligor characteristics. However, we borrow an additional element from the so-called "reduced-form" models of default (for a survey of both types of model, see Bohn (1999)), by working with a default process in Poissonian form. This technique was introduced by Jarrow and Turnbull (1995). We follow the variant utilised by Madan and Unal (1998, 1999), in that the default event arrival rate becomes a function of the same obligor fundamentals as the ones that drive the asset prices. However, this principle is developed in a way to establish a link between the default process of the abstract reduced-form models and the empirics inspired by the Expected Default Frequency notion of KMV. The proposed model allows one to deal with a loan portfolio with correlated defaults in a natural way.

4.1 Definitions

Consider a portfolio of n loans issued by n different companies (obligors). Loan i pays a coupon c_t^i at time t and the coupon plus the face value F^i at maturity T^i ($t = 1,\ldots, T^i$). The value of the loan to firm i at time t is denoted by V_t^i. Then, $B_t = \sum_{i=1}^{n} V_t^i$ is the value of the loan portfolio to be found.

There are N assets traded in the market, with prices P_t^j at time t ($j = 1,\ldots, N$). These assets represent all sources of aggregate uncertainty in the economy independent of the actions of obligors defined above. In this sense, the financial markets outside the considered borrower set are complete. These uncertainty factors will be represented by random variables x^j, $j = 1,\ldots, N$. By x, we denote the vector of the unobserved state variables of the model.

The loans are risky. If firm i generates period t-cash flow $Y_t^i = f^i(x_t)$ net of all the other debt service obligations, then the probability of default π_t^i on the loan is an inverse function of the difference $Y_t^i - C_t^i : \pi^i(x) = \pi(Y^i - C^i)$. Here, $C_t^i = c_t^i$ if $t < T^i$ and $C_t^i = c_t^i + F^i$ if $t = T^i$. Hence, the variable driving the default rate in our model is an analogue of the *distance-to-default* measure used in KMV. One shall think of π as approaching unity when the distance to default falls to minus infinity, and approaching zero when it increases to plus infinity.

The space of random events in our model is formed by pairs $\omega = (\chi, b)$, where χ is a realisation of x and $b = [b^1,\ldots, b^n]$, $b^i = S$ if there is no default (survival) and $b^i = D$ if firm i defaults on the loan. The arrival *fact* of the default event itself (which we represent by the Bernoulli process B^i) is assumed independent of x, ie only *the probability value* of the default is x-dependent through the cash flow variable Y^i.

For each loan, there is collateral that is tradable and depends on the same sources of uncertainty as the basic assets. That is, the collateral price for loan j is equal to $Z^i = \zeta^i(\omega)$. If the loan defaults, the bank seizes the collateral, ie receives the value of Z^i.

There are two important cases to be distinguished with regard to the collateral prices. One possibility is to allow ζ^i to depend on both x and b. That is, this collateral is worth different amounts depending on whether the debt has defaulted or not. This would be the case if there were a separate structural factor behind the realisation of b, correlated with the market risk factors x. The same factor should be responsible for the value of the collateral. This situation would allow loan i to be priced in accordance with the risk neutral valuation principles (see an example of such a valuation in Derviz and Kadlčáková (2002, Section 3). It may occur when the collateral is very obligor-specific.

However, an equally legitimate case is that of the collateral being totally unrelated to the operation of the firm (eg securities in its investment portfolio). Then $Z^i = \zeta^i(x)$ and a unique risk neutral valuation of the loan is impossible. In that case, one must resort to pricing techniques based on explicit individual portfolio optimisation. This is done next.

4.2 The individual loan and the portfolio value

Let us consider an optimising investor in discrete time who decides upon allocating his/her wealth between the existing marketable assets, ie the N traded securities, the n collateral assets and the n company loans (all defined above). This is a standard optimisation problem under uncertainty in discrete time.

Let g^j be the stream of coupons/dividends paid out by the basic security j, and h^i the same thing for the collateral security i. These values are unknown at the beginning of each period, when the investor makes the portfolio allocation decisions. Define R^j as the current yield on the basic security and z^i as the current yield on the collateral. In addition, it is convenient to use the notation y^j for the continuously compounded current yield on basic asset j:

$$e^{y^j_{t+1}} = 1 + R^j_{t+1} = \frac{g^j_{t+1} + P^j_{t+1}}{P^j_t},\ 1 + z^i_{t+1} = \frac{h^i_{t+1} + Z^i_{t+1}}{Z^i_t}.$$

r_{t+1} will denote the risk-free short rate between periods t and $t+1$. The period utility function of the investor is a function of the dividend rate withdrawn after the investment strategy gains are realised: $u = u(\tilde{\rho})$. The ρ-dependence is of the standard Inada form. If the time preference rate of the investor is $\beta \in (0, 1)$, the pricing kernel (stochastic discount factor, see Campbell et al (1997), or Cochrane (2001)) is given by:

$$M_t^{t+1} = \frac{\beta u'(\rho_{t+1})}{u'(\rho_t)},\ M_t^\tau = \prod_{k=t}^{\tau} M_k^{k+1},\ \tau > t,\ M_t^t = 1. \tag{2}$$

The information available to the investor at time t consists of the trajectories of g, h, P and Z as well as the default event realisations b, all up to time t. The no-default up to time t subset of the event space for loan i will be denoted by N_t^i. Let the x-dependent statistics of the survival, $S^i_{t,\tau}$ between times t and $\tau \geq t+1$, be defined as:

$$S^j_{t,\tau} = \prod_{k=t+1}^{\tau} \left(1 - \pi_k^i\right). \tag{3}$$

Now we apply the standard asset pricing theory results. The optimal investor behaviour implies the following asset pricing formulae (special cases of the discrete time consumption-based CAPM):

$$E_t\left[M_t^{t+1}(1+R^j_{t+1})\right] = 1,\ E_t\left[M_t^{t+1}(1+z^i_{t+1})\right] = 1,\ E_t\left[M_t^{t+1}\right](1+r_{t+1}) = 1, \tag{4}$$

$$Z_t^i = E_t\left[\sum_{\tau=t}^{\infty} M_t^\tau h_\tau^i\right], \tag{5}$$

$$V_t^i = Z_t^i + E_t\left[\sum_{\tau=t}^{T^i} M_t^\tau B_\tau^i \left(C_\tau^i - h_\tau^i\right)\right]. \tag{6}$$

In view of our assumptions about default arrival independence on x, equation (6) can be rewritten in the form:

$$V_t^i = Z_t^i + E_t{}^*\left[\sum_{\tau=t}^{T^i} S^j_{t,\tau+1} M_t^\tau \left(c_\tau^i - h_\tau^i\right)\right] - E_t{}^*\left[S^i_{t,T^i+1} M_t^{T^i+1} Z^i_{T^i+1}\right], \tag{7}$$

where, now, the conditional expectation $E_t{}^*$ is taken only with respect to the market-wide risk factors. Thus, we have eliminated the firm-specific default event process B^i from the debt pricing formula (6). Also note that the asset pricing equations (4) could be written with expectation $E_t{}^*$ instead of E_t from the outset, since they are default event-independent.

Next, utilising the previously made assumption about market (in)completeness, we note that the value of individual loans and the loan portfolio as a whole can be calculated as soon as one reconstructs a formula for the pricing kernel M_t^τ from the pricing equations (4). Formally, we will be estimating and using *the empirical pricing kernel* in (7) instead of the theoretical pricing kernel (2). The empirical pricing kernel is a projection of the theoretical one on the space of modelled market uncertainty factors x. Due to the law of iterated expectations, the left-hand side of (7) is fully determined by this projection (since the external expectation is taken with respect to x-generated random events). One goes about calculating asset prices in the pricing kernel context by either setting up a parametric model for it or applying orthogonal basic decomposition methods known from numerical mathematics, under a non-parametric approach. Examples in the literature include Ait Sahalia and Lo (2000), Jackwerth (2000), or Rosenberg and Engle (2002).

We proceed by constructing a Gaussian state-space model for the pricing kernel and estimating it on the Czech asset data.

4.3 A hidden factor asset pricing model

The observation process y of our model consists of the above-mentioned traded asset yields $y^1,\ldots,y^N$. The observation equations are:

$$y_{t+1}^j = a_0^j + a_1^j x_t + A^j x_{t+1},\ j = 1,\ldots,N \tag{8}$$

Here, $a_0 = \left[a_0^1,\ldots,a_0^N\right]^T$ is an $N \times 1$-vector of intercepts, a_1 and A are $N \times n$-matrices of coefficients with rows $a_1^j = \left[a_{11}^j,\ldots,a_{1n}^j\right]$ and $A^j = \left[A_1^j,\ldots,A_n^j\right]$ respectively. The n-dimensional vector x of unobserved state residuals follows the VAR(1)-process:

$$x_{t+1} = bx_t + B\varepsilon_{t+1}. \tag{9}$$

Coefficient matrices b and B in (9) are of size $n \times n$. Process ε is an n-dimensional vector of mutually independent standard normal errors. In general, $n \neq N$ and, if there is a reason to assume eg cyclical components in the observations, one will need to take $n > N$.

Another unobserved state variable is the log of the one-period pricing kernel $m_{t+1} = \log M_t^{t+1}$:

$$m_{t+1} = \lambda_0 + \lambda_1 x_t + \Lambda x_{t+1}. \tag{10}$$

Here, λ_0 is a scalar constant, whereas λ_1 and Λ are row vectors of dimension n. The observation equation system (8) together with the state equation system (9), (10) constitutes the state-space representation of the present model. However, this is not the definitive representation to be estimated, since one must incorporate the coefficient restrictions following from the no-arbitrage pricing equations (4) for returns y.

Proposition 1 *The no-arbitrage pricing conditions for the asset return model defined by (8)-(10) above are equivalent to the following constraints on the model coefficients:*

$$a_0^j = -\lambda_0 - \frac{\left|(\Lambda + A^j)B\right|^2}{2},\ a_1^j = -\lambda_1 - (\Lambda + A^j)\mathrm{b},\ j = 1,\ldots,N. \tag{11}$$

The proof is given in Appendix B. Note that, whereas the first equality in (11) is scalar, the second one is for N-dimensional row vectors.

We are able to reduce the number of estimated parameters by simplifying the covariance structure of the state equation through a change of variables. Specifically, assume that B is non-singular and put $x_t = Bu_t$ for all t. Then:

$$u_{t+1} = \Phi u_t + \varepsilon_{t+1},\ \Phi = B^{-1}bB. \tag{12}$$

The log-pricing kernel equation is now given by:

$$m_{t+1} = c_0 + c_1 u_t + Cu_{t+1} \tag{13}$$

instead of (10), with $c_0 = \lambda_0$, $c_1 = \lambda_1 B$, $C = \Lambda B$. Put $\gamma^j = (\Lambda + A^j)B$. It is easily checked that the no-arbitrage pricing conditions (11) of Proposition 1 imply the following equations for the observed yields:

$$\begin{aligned} y_{t+1}^j &= -\lambda_0 - \frac{\left|\gamma^j\right|^2}{2} - \left(c_1 + \gamma^j\Phi\right)u_t + \left(\gamma^j - C\right)u_{t+1} \\ &= -m_{t+1} - \frac{\left|\gamma^j\right|^2}{2} + \gamma^j\varepsilon_{t+1},\ j = 1,\ldots,\mathrm{N}. \end{aligned} \tag{14}$$

Process u will be the state process of the definitive formulation of our model. The state equations for its components are in (12). The observation equations are in (14). The model written in state-space form in (12)-(14) contains restrictions on the fixed coefficients. After having estimated it by means of the Kalman filter method and obtained the (hyper)coefficients Φ, λ_0, c_1, C, γ, we can reconstruct the observation equation coefficients a_0, a_1 by means of (10) and the coefficient matrix A by using the definition of γ:

$$A^j = \gamma^j B^{-1} - \Lambda = (\gamma^j - C)B^{-1}, j = 1, \ldots, N. \qquad (15)$$

This will complete the estimation procedure for the pricing kernel and allow us to price the non-traded debt by means of (5).

One-period interest rate as a basic security

As was already mentioned, the present model is not constructed by directly fitting the yield curve. However, if one works with monthly data, it is convenient to take the one-month risk-free interest rate as one of the basic securities. Let us assign it superscript 1. Then y^1_{t+1} is the continuously compounded one-period rate between t and $t+1$. We incorporate it into the model by imposing the requirement $A^1 = 0$ in (8) for $j = 1$. Equivalently, one must have $\gamma^1 = \Lambda B = C$, then the first observation equation degenerates to:

$$y^1_{t+1} = -\lambda_0 - \frac{|C|^2}{2} - (c_1 + C\Phi)u_t,$$

and the coefficient recovery formulae (15) are applicable for $j = 2, \ldots, N$.

All other points in the yield curve (with longer maturities) generate uncertain one-period yields, so that there is no need for further specialisation in the event that one would want to include any of them in the list of basic securities.

5. Asset pricing data and economic capital estimation

The sample period 1999-mid-2003 is characterised by a negative trend in the Czech interest rate and bond yield data (the disinflation process and the monetary policy rate convergence to the EU level). The model described in the previous section would require a number of messy technical adjustments to accommodate these non-stationary yields along with the stock return data. Since our objective is the modelling of real shock effects on credit risk valuation, we need a pricing kernel projected on the space of most relevant security returns. That is, for our purposes it is sufficient to select the stationary assets with a clear relation to the economic cycle and avoid the problems with fitting the yield curve evolution. Therefore, we have selected a four-factor model based on the following asset returns:

- PX50 stock index return;
- Česká pojišťovna (a major Czech insurance company) stock return;
- DAX stock index return;
- Altana (the pharmaceutical company) stock return.

This choice is motivated by the effort to capture both internal and external risk factors for the small open Czech capital market with a high degree of dependence of the corresponding markets in Germany and the European Union. The stock index returns reflect the direct link to the Czech and euro area business cycles. The additional stocks (Česká pojišťovna on the Czech side and Altana on the EU side) were found to be less than perfectly correlated with the major indices. Therefore, they serve in the model as proxies for the countercyclical risk factors priced in the corresponding markets.

Estimation of the empirical pricing kernel

The unobserved state-space model covered in Section 4, implemented for the four named assets, contains four state variables $u1$, $u2$, $u3$ and $u4$, and their four one-period lags. According to the estimation outcome, the pricing kernel log, m, has the following dependence on these variables:

$$m = 0.023 - 0.00376\, u_1(-1) + 0.0002\, u_2(-1) + 0.00274\, u_3(-1) + 0.00681\, u_4(-1)$$
$$+ 0.00993\, u_1 + 0.00065\, u_2 + 0.0148\, u_3 + 0.0203\, u_4.$$

The estimated autoregression matrix Φ for the states (cf (12) in Section 4) is equal to:

$$\Phi = \begin{bmatrix} 1786 & -0.320 & 0.192 & -0.135 \\ 2.370 & 0.038 & 0.541 & -0.432 \\ -0.0570 & 0.031 & 0.936 & -0.032 \\ -0.023 & -0.010 & 0.004 & 0.997 \end{bmatrix}.$$

Using the estimated state variable series, one can now establish the dependence of the obligor company cash flow (cf (7)) on the hidden risk factors by regressing these flows on the four state series (with first-order lag terms). Considering the way in which the portfolio was constructed, individual asset cash flows are proxied by the price indices corresponding to the relevant industries. We implement a seemingly unrelated regression (SUR) with the price indices acting as the dependent variables and the state vectors as the explanatory variables:

$$Y^i = \beta^i + \sum_{k=1}^{4} \gamma_k^i \cdot u_k + \sum_{k=1}^{4} \delta_k^i \cdot u_k(-1) + \varepsilon_i, \qquad i = 1, \ldots, 30.$$

The coefficients $\beta^i, \gamma_k^i, \delta_k^i$ facilitate the computation of the annual default probabilities for each asset. By assumption, year t default probability of the *i-th* asset is linked to the cash flow Y_t^i according to the formula:

$$\pi_t^i = \frac{e^{-\alpha^i - Y_t^i}}{e^{-\alpha^i - Y_t^i} + e^{\alpha^i + Y_t^i}}, \quad i = 1, \ldots, 30, \ t = 1, \ldots, 5. \tag{16}$$

Intuitively, (16) reflects the fact that high cash flow realisations would yield probabilities of default (PDs) close to zero while highly negative cash flows would yield PDs close to unity. The PD formulae are calibrated to become compatible with the S&P rating structure of the portfolio. This means that the parameters α^i are estimated in such a way as to produce the S&P asset-specific ratings when the cash flows in (16) are zero.

The recovery rates are also modelled as state-dependent variables. The last four of the six types of collateral considered (securities, commercial real estate, other and no collateral) are proxied by the following variables:

- yields on five-year Czech government bonds;
- the real estate price index;
- a linear combination of the PPI and industrial production index (reflecting uncertainty regarding other types of collateral); and
- an insurance sector price index.

As in the case of the assets' cash flows, these variables are regressed on the state variables. By assumption, the recovery rates are exponential functions of the hidden risk factors u:

$$RR_t^i = e^{\left(a^i + b^i + \sum_{k=1}^{4} c_k^i u_k + \sum_{k=1}^{4} d_k^i u_k(-1)\right)}, \quad i = 1, \ldots, 30, \quad t = 1, \ldots, 5 \tag{17}$$

where the coefficients b^i, c_k^i and d_k^i are obtained from the regressions and a^i are parameters that calibrate the recovery rate formulae to the real recovery rates as considered in Table 5.

The annual values of individual loans up to maturity can be calculated as soon as these intermediate valuation elements are available:

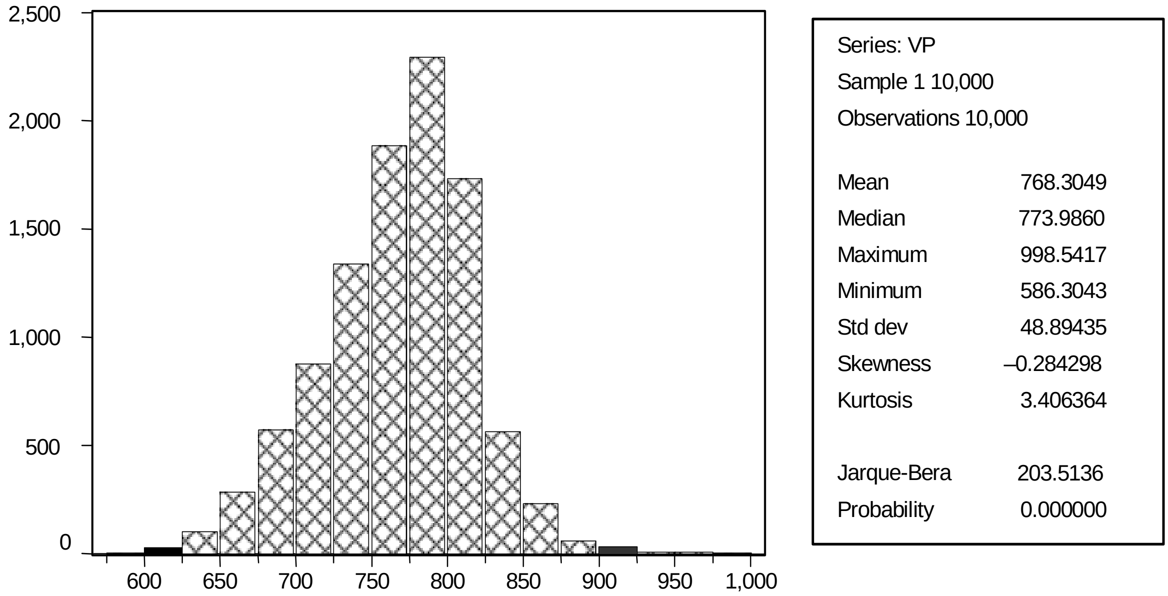
$$V_t^i = U^i \cdot \sum_{\tau=1}^{t} M_\tau \left[c_\tau^i \left(1 - \pi_1^i\right) \ldots \left(1 - \pi_\tau^i\right) + RR_\tau^i \cdot \left(1 - \pi_1^i\right) \ldots \left(1 - \pi_{\tau-1}^i\right) \pi_\tau^i \right], \quad t = 1, \ldots, T^i. \tag{18}$$

Here U^i is the contractual loan volume and c_τ^i is the loan annual payment (the lending rate in the years prior to maturity and one plus the loan lending rate at maturity).

We conduct a Monte Carlo simulation with 10,000 scenarios to generate the distribution of the portfolio value at the risk horizon. Each replica starts with the four-component state vector at the end of 2000. Then, the annual u vectors for the subsequent five-year period are determined using autoregressive state variable formulae as in (12). Additionally, each component of the vector u is shocked each year with an error term randomly drawn from the standard normal distribution. Given the *u*-dependence of the pricing kernel given by (15), default probabilities (see (16)) and recovery rates (see (17)), each loan's valuation is fully determined by (18). The portfolio value is obtained by summing up the individual loan values. This procedure is repeated 10,000 times. The mean of the portfolio distribution thus obtained (see Figure 3) is then used to estimate the economic capital. In our understanding the economic capital represents the difference between the riskless value of the portfolio and the mean value mentioned above. The riskless value of the portfolio is estimated assuming no default events taking place over the risk horizon, thus making the π_t^i equal to zero in (18) for each i and t.

Figure 3

Portfolio distribution according to the pricing kernel model, baseline case



Business cycle events and economic capital in the pricing kernel model

We are now ready to investigate the consequences of the domestic and foreign economic cycle. The latter case will be modelled by constructing the real shocks for Germany since there is no generally accepted industrial production index for the European Union or the euro area as a whole.

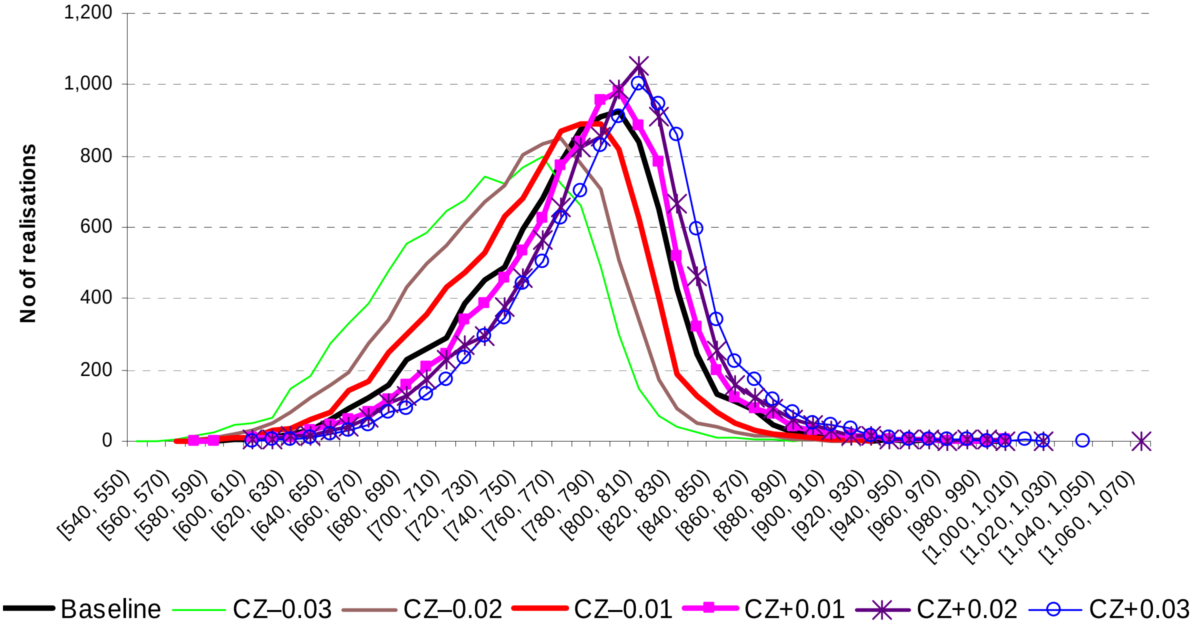
The loan portfolio distributions under different macroeconomic developments in the Czech Republic and Germany are shown in Figure 4. In addition, Figure 5 shows the pricing kernel baseline and the most extreme positive/negative economic activity shock cases in comparison to the CreditMetrics distribution.

When we model the different business cycle developments in the pricing kernel model, we rely on a shock to a corresponding state variable (the first one for the Czech business cycle and the third one for the German). These shocks were selected since the underlying risk factors roughly correspond to the normalised Czech and German industrial production indices. They also produce an almost isolated response in the first (PX50 exchange index) and the third (Dax) of the modelled assets. It turns out that the shocks we model (denoted by CZ $\pm$ 0.01, CZ $\pm$ 0.02, CZ $\pm$ 0.03 and DE $\pm$ 0.01, DE $\pm$ 0.02, DE $\pm$ 0.03 in Figures 4 and 5) correspond to the 1, 2 and 3% rise/decline of the Czech and German Industrial Production Index (IPI), respectively.

Figure 4

Pricing kernel model: loan portfolio distribution under domestic and foreign real economic activity shifts

(a) Shocks to the Czech business cycle



(b) Shocks to the German business cycle

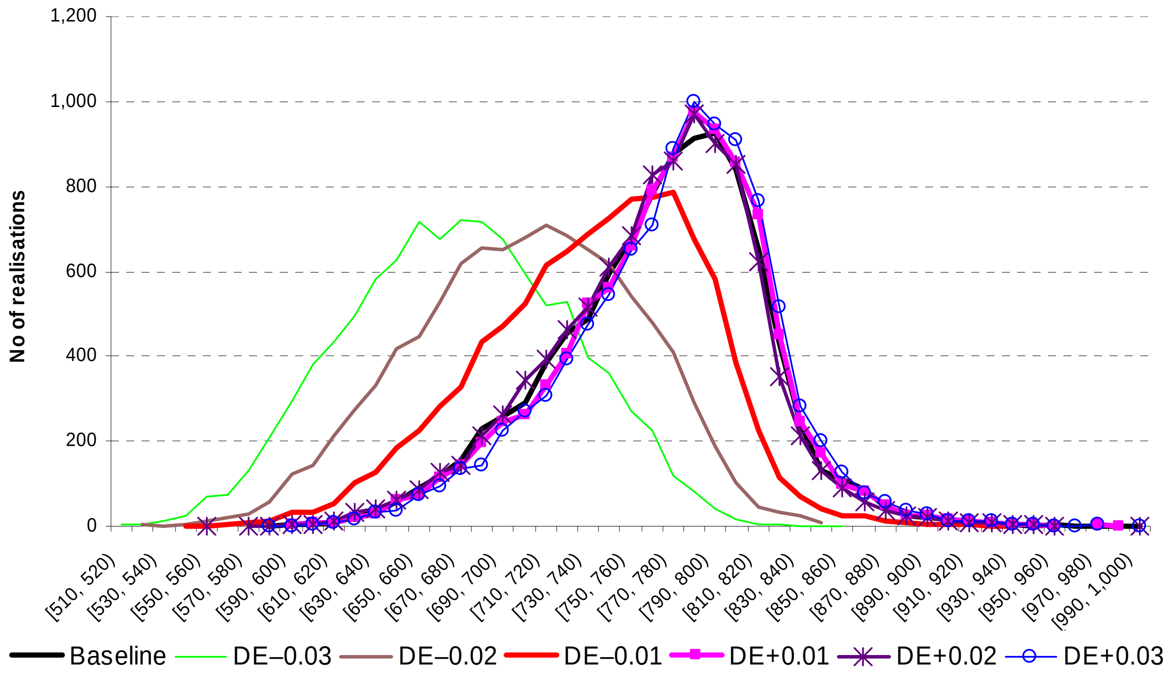


Figure 5

Portfolio value distributions according to the CreditMetrics and pricing kernel models, different growth scenarios

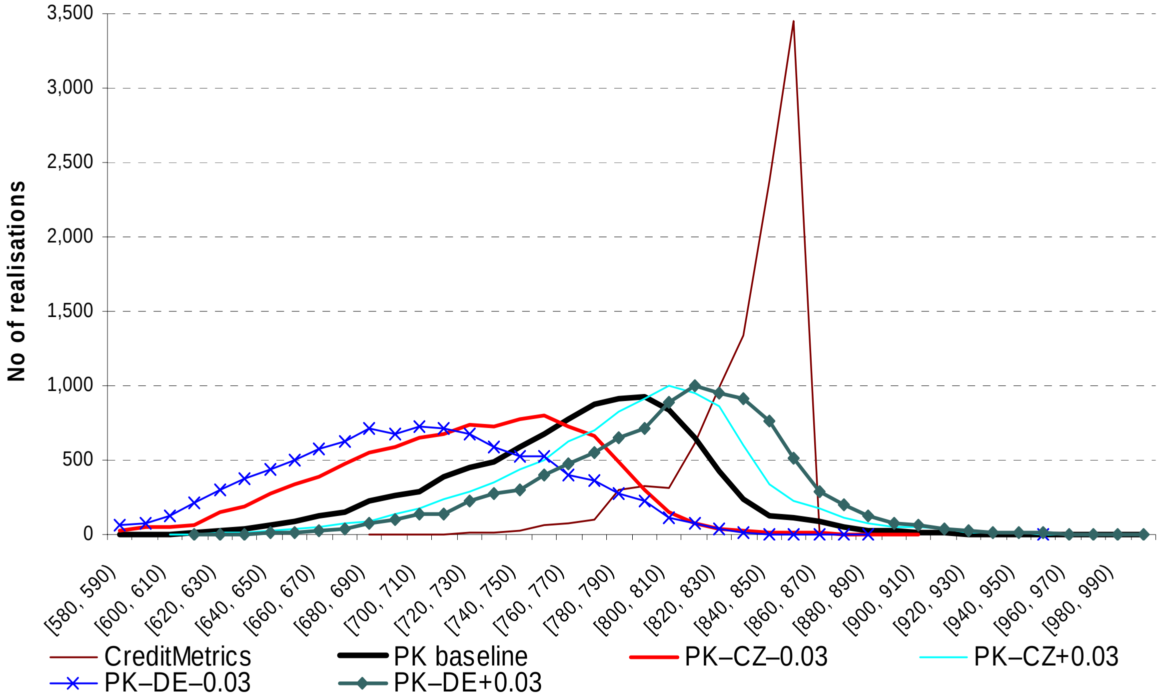


Table 8 sums up the credit risk-related estimations of economic capital in all the cases considered. For comparison, we have also included the regulatory capital measures for the same artificial portfolio (both standardised and IRB approaches) according to the original NBCA guidelines of January 2001 and the 3rd Quantitative Impact Study of October 2002. These were obtained in our earlier paper (Derviz et al (2003)), where a detailed account of the calculation procedure can be found.

6. Conclusion

This paper has an applied objective of analysing the impact of business cycle and monetary policy on credit risk valuation. It does not aspire to create an empirically waterproof econometric model of business cycle effects on asset prices as such. That is why we are not dealing with processes for real economic activity, inflation and monetary policy as (either observable or hidden) explanatory factors for the observed security yields. Instead, we take a shortcut by assuming that at least a subset of the chosen basic asset yields provides a sufficient statistic of either present or future (expected) economic growth and of the monetary policy stance. We are interested in modelling a domestic economic up-/downturn with foreign development being stable, and an expansion/recession abroad with the domestic environment being stable. This has been identified with the corresponding behaviour of the chosen asset returns in both economies considered. Namely, current high/low growth is reflected in the high/low current values of the leading asset return.

The model outlined in the paper is likely to be free of certain deficiencies typical to the two standard ones, whose application to economic capital calculations was described in detail in Section 3. For instance, a counter-intuitive negative dependence of the capital requirement on the market interest rate is an unfortunate feature of the way CreditMetrics works with the relation between the debt value and the economic capital. That model does not have any link from the interest rates to the firm's ability to repay the loan. On the other hand, our model is able to create this link because the obligor's net cash flow will usually be negatively related to the market rates of interest. Therefore, the reduction of the latter (such as the down-translation of the forward zero curve) increases the cash flow and, therewith, reduces the default event rate.

Table 8

Summary of economic capital estimations

	Regulatory capital	1%-percentile	5%-percentile	Mean	99% economic capital	95% economic capital	Non-VaR economic capital
NBCA - Standardised approach (Jan 2001)	51.84						
NBCA - IRB approach (Jan 2001)	165.46						
NBCA - Standardised approach (Oct 2002)	46.90						
NBCA - IRB approach (Oct 2002)	44.79						
CreditMetrics		767.90	796.62	845.78	77.89	49.16	
CreditRisk+ (Loss)		133	101	42.18	90.82	58.82	
Pricing kernel model baseline				**768.30**			**64.56**
CZ–0.03				723.00			109.86
CZ–0.02				737.94			94.93
CZ–0.01				754.27			78.59
CZ+0.01				775.68			57.19
CZ+0.02				783.32			49.54
CZ+0.03				790.24			42.63
DE–0.03				673.73			163.14
DE–0.02				704.80			132.06
DE–0.01				738.20			99.66
DE+0.01				770.62			68.24
DE+0.02				766.15			72.71
DE+0.03				774.14			65.72

Another important advantage of the model is its ability to handle correlated defaults in a natural way. This is because default correlation in the model is not an exogenously given property or an ad hoc assumption, but instead follows by construction from the dependence of default rates on common risk factors.

A certain difficulty lies in the necessity of calculating the pricing kernel recursively for multi-period loan contracts. However, the calculations themselves are routine and are based on well developed numerical techniques, allowing one to apply relatively standard software.

Once the distribution of the loan portfolio market price has been calculated, it can be used to derive the bank-internal measure of economic capital. The latter shall be subsequently compared with the prudential capital values derived directly from regulatory principles. The difference would tell us the degree of discrepancy between internal risk measurement and regulatory mechanisms of risk-based capital allocation by the bank.

The main conclusions of this study can be summarised as follows.

- Bank capital on the basis of our model would react procyclically. However, this reaction differs substantially from loan to loan, so that certain “countercyclical” loans may even be assigned lower capital values under a downturn. Also, stochastic properties of the collateral (their non-zero covariances with the underlying systemic uncertainties) mitigate the procyclical economic capital allocation in our model.

- When floating interest rates and changes in yield curves were modelled, the estimate of economic capital was generally higher. This result is intuitive, as increased uncertainty should generally impose higher required levels of capital on banks.
- The particular changes in the forward zero curves analysed in this paper in the CreditMetrics context did not impose significantly different levels of economic capital than did the case with stable forward zero curves. The scenario where forward zero rates fell (both translation and clockwise rotation) required slightly more capital, and this situation persisted at all considered confidence levels.
- Risky debt valuation by the traditional asset pricing methods currently in use by the banking industry tends to generate higher loan values and reduce economic capital requirements, compared to other possible regulatory and model-based risk measurement methods. Therefore, the regulator may see an effort on the part of the banks to treat different parts of loans on their balances differently in terms of economic capital. The difference will go in the direction of reducing capital allocation (and specialising collateral requirements) in those segments of the loan portfolio that exhibit strong correlation with traded risks.
- Asset pricing methods of risk measurement may lead to a better recognition of the role of the business cycle and other systemic macroeconomic factors in economic capital determination. Therefore, the feedback from the business cycle-related events in the security markets to economic capital, as captured in our model but also present in many existing risk management procedures, may make natural (and desirable) countercyclical economic capital adjustments possible.
- Those methods of credit risk measurement which explicitly deal with market incompleteness (ie the lack of market valuation of both the loan itself and the assets of the obligor) lead to a better recognition of the role of the business cycle and other systemic macroeconomic factors in economic capital determination. Therefore, the regulator should encourage the use of methods that allow for countercyclical adjustments by banks of procyclically biased risk management procedures.

Appendix A:
Yield curve fluctuations, CreditMetrics and economic capital

A.1 CreditMetrics - allocation of economic capital under floating lending rates and fixed forward zero curves

We assume that the mechanism of future changes in lending interest rates is as follows:

$$1+\frac{r_t^{g'}}{100}=\left(1+\frac{r_t}{100}\right)*e^{\sigma_t}$$

Here σ is a normally distributed random variable, so the exponential follows a log-normal distribution.

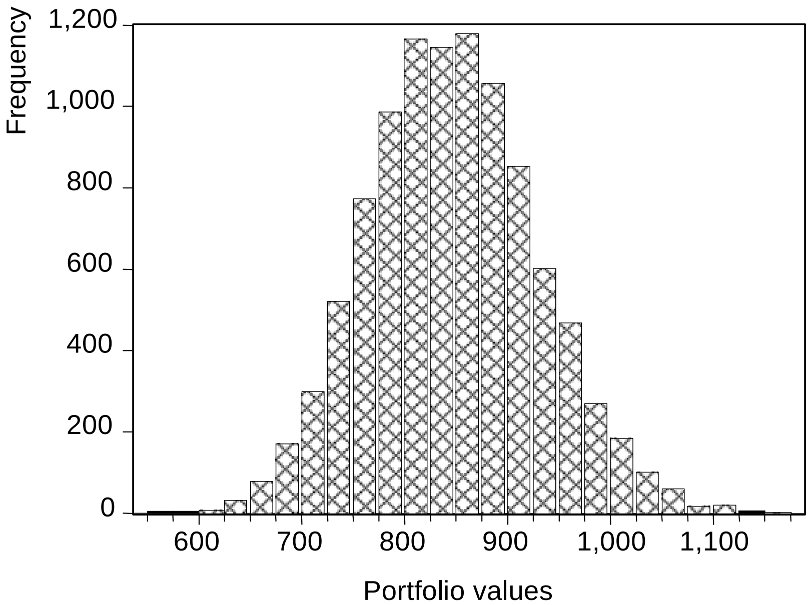
In this formulation, interest rates on loans preserve the markups that capture obligor-specific risk or liquidity premium (already incorporated in the old interest rates), but also contain a random component reflecting uncertainty related to the future course of the money market rate (Pribor). The random component does not vary across obligors. In our simulations it was obtained as a random draw from a log-normal distribution. The parameters (mean and standard deviation) describing the log-normal distribution were estimated on actual Czech data (1Y Pribor over the year 2000, a period of relative rate stability and no monetary policy changes).

A.2 Monte Carlo simulation with floating interest rates

A total number of 10,000 random draws from the standard log-normal distribution were performed to determine the random component of Pribor and thus the loans' interest rates. Due to the floating nature of the interest rates, the valuations of each asset and of the portfolio varied in line with the particular values of the random draws. A total number of 10,000 portfolio values at the end of the risk horizon were thus obtained. Figure A1 shows the relative frequencies of these values at the year-end.

Figure A1

The empirical distribution of the portfolio value with floating interest rates



Series: VP	
Sample 1 10,000	
Observations 10,000	
Mean	848.3348
Median	845.5493
Maximum	1167.615
Minimum	554.5349
Std dev	81.97142
Skewness	0.195979
Kurtosis	3.005247
Jarque-Bera	64.02450
Probability	0.000000

Note: The horizontal axis shows non-overlapping intervals that cover the entire range of the estimated portfolio values, and the vertical axis shows the frequencies with which the portfolio values fell into those intervals.

The estimation of economic capital in this case is given in Table A1.

Table A1

Capital requirements assuming different changes in interest rates and forward zero curves (FZCs)

(10,000 random draws)

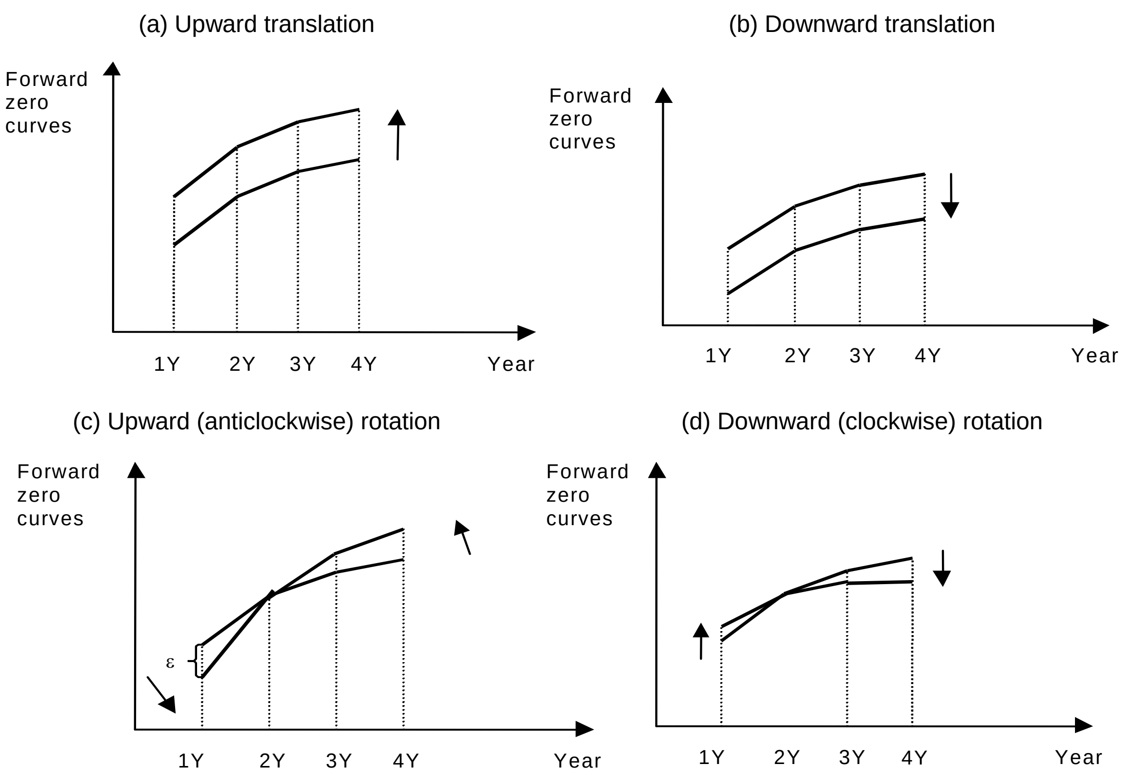
	1%-percentile	5%-percentile	Mean	99% ec capital	95% ec capital
Fixed interest and fixed FZC	767.90	796.62	845.78	77.89	49.16
Floating interest and fixed FZC	669.65	718.48	848.33	178.69	129.85
Upward translation of FZC	655.15	705.70	834.25	179.11	128.56
Downward translation of FZC	672.44	724.32	856.94	184.50	132.62
Anticlockwise rotation of FZC	668.83	722.92	839.14	170.31	116.22
Clockwise rotation of FZC	668.83	722.92	853.01	184.18	130.09

A.3 Monte Carlo simulation with floating interest rates and stochastic forward zero curves

The next four cases retain the assumption that floating interest rates were charged by the bank on its loans. Additional uncertainty is added with regard to changes in forward zero curves (the discount factors entering the loan valuations) over the one-year period. We analysed the impact on the bank's need for economic capital under the following changes in the forward zero curves: upward translation, downward translation, clockwise rotation and anticlockwise rotation. Rotations of the forward zero curves are around the point determined by the two-year forward rate. These changes are illustrated in Figure A2.

Figure A2

Simulated changes in the forward zero curves



We assumed that these changes reflected subjective expectations concerning the evolution of the Czech forward zero curves over a one-year period. They became rating class-specific by adding the US forward spread (the difference between the US rating class-specific and the US forward zero curves). We incorporated these changes into the forward zero curves in the model according to the formula:

$$1+\frac{f_t^{g'}}{100}=\left(1+\frac{f_t}{100}\right)*\left(1+\frac{s_t^g}{100}\right)*e^{\varphi_t} \quad t=1, 2, 3, 4 \tag{A1}$$

Here f_t is the original Czech forward zero rate at year t, s_t^g is the spread in forward zero rates characteristic of the g-th rating class at time t, and φ_t is a random draw from the normal distribution (thus $e^{\varphi}{}_t$ is log-normally distributed).

The proposed changes in the forward zero curves were captured in the model by considering particular random variables ϕ_t in (A1):

$\phi_t = \mu + \varepsilon$, $t=1, 2, 3, 4$, for an upward translation;

$\phi = -\mu + \varepsilon$, $t=1, 2, 3, 4$, for a downward translation;

$\phi_1 = -\mu + \varepsilon$, $\phi_3 = \mu + \varepsilon$, $\phi_4 = 2\mu + \varepsilon$, for an upward rotation;

$\phi_1 = \mu + \varepsilon$, $\phi_3 = -\mu + \varepsilon$, $\phi_4 = -2\mu + \varepsilon$, for a downward rotation.

The overall effect of an upward shift in the forward zero curve is a decrease in the present value of all loans in the portfolio. If the bank heavily discounts the future, the opportunity cost of granting loans increases, since alternative assets may provide higher returns in the future. Accordingly, the present value of the cash flows accrued from the loans is lower compared with the case where the forward zero curves remain unchanged. The effect of a downward shift of the forward zero curve is the opposite of the one mentioned above. Rotations of the forward zero curve affect assets' valuations depending on maturity. For example, the anticlockwise (upward) rotation discounts assets with a short maturity less and assets with a long maturity more. Therefore, the valuation of the portfolio is very sensitive to the portfolio composition. If more assets fall into the long-maturity category the present value of the portfolio tends to decrease, while if they fall into the short-maturity category the present value of the portfolio tends to increase.

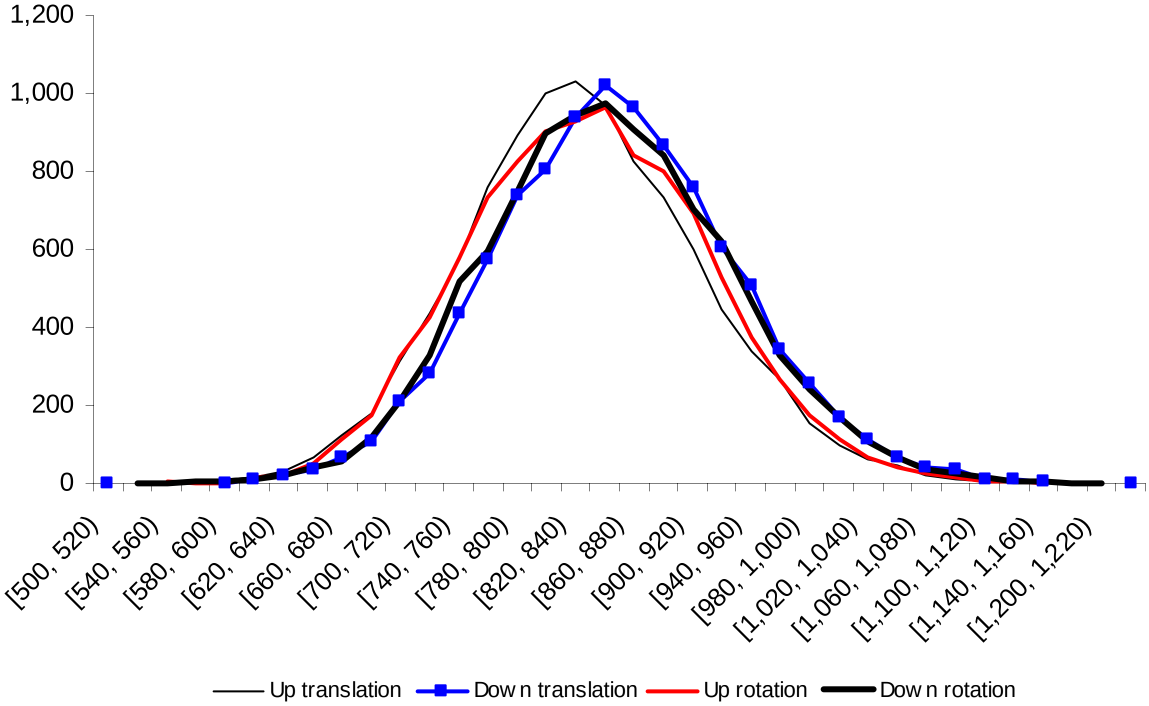
In all previous formulations of the φ_t distribution, the parameter μ determined the magnitude of the deterministic change in forward zero curves and ε added random deviations. We wanted changes in the deterministic part of the discount factors not exceeding 1% (thus, if for a given maturity the forward zero rate was 3.4%, we wanted it to deviate upward to 4.4% only). This assumption implied a value for μ of 0.01. In each case, ε was assumed to follow a normal distribution with mean $-\frac{\sigma^2}{2}$ and standard deviation σ, so that the mean of the log-normally distributed factor e^{ε} is equal to 1. Under these conditions φ became normally distributed with mean $\mu + m$. The standard deviation σ of ε was estimated by computing the standard deviation of the log(1+1Y Pribor/100) variable using daily observations over the year 2000 after removing the trend. The same estimated values of the parameters μ and σ were used in all cases of random changes in forward zero curves.

We performed Monte Carlo simulations containing 10,000 scenarios that simultaneously accounted for random changes in interest rates and forward zero curves. Shown next are the portfolio value distributions and the estimates of economic capital based on simulations that incorporated the proposed changes into the forward zero curves.

Figure A3 displays the portfolio value distributions when the four forward zero curve change cases discussed above are compared.

Figure A3

Portfolio distributions under different forward zero curve movements



Economic capital estimations at different confidence levels are displayed in Table A1. The downward translation and the clockwise rotation of the forward zero curves impose the highest requirements of economic capital in this particular example. However, economic capital seems to converge towards the fixed forward zero curves (with floating lending rates) case when the confidence level is reduced.

Appendix B:
Proof of Proposition 1

The pricing relation (4) for asset j can be written as $E_t\left[e^{m_{t+1}+y^j_{t+1}}\right]=1$. Since both m and y^j are normally distributed with known conditional expectations and variances for each date t, the last equation can be rewritten as:

$$E_t\left[m_{t+1}+y^j_{t+1}\right]+\frac{1}{2}\,Var_t\left[m_{t+1}+y^j_{t+1}\right]=0.$$

In accordance with (8)-(10), this is equivalent to:

$$\lambda_0+\lambda_1 x_t+\Lambda b x_t+a^j_0+a^j_1 x_t+A^j b x_t+\frac{1}{2}\left|(\Lambda+A^j)B\right|^2=0 \tag{B1}$$

for all t for each j. Equations (B1) can be considered as identities involving a non-trivial vector autoregressive state process x. They are satisfied if and only if all coefficients on the left-hand side of (B1) are identically zero. This means:

$$a^j_0+\lambda_0+\frac{\left|(\Lambda+A^j)B\right|^2}{2}=0,\; a^j_1+\lambda_1+(\Lambda+A^j)b=0$$

for all j, which is equivalent to (11).

References

Ait Sahalia, Y and A Lo (2000): “Nonparametric risk management and implied risk aversion”, *Journal of Econometrics*, 94, pp 9-51.

Ang, A and M Piazzesi (2003): “A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables”, *Journal of Monetary Economics*, 50, pp 745-87.

Barnhill, T and W Maxwell (2002): “Modeling correlated market and credit risk in fixed income portfolios”, *Journal of Banking and Finance*, 26, no 3, pp 347-74.

Bohn, J (1999): “A survey of contingent-claims approaches to risky debt valuation”, *working paper*, Haas School of Business, U C Berkeley, June.

Campbell, J, A Lo and C MacKinlay (1997): *The econometrics of financial markets*, Princeton Univ Press, Princeton, NJ.

Cochrane, J H (2001): *Asset pricing*, Princeton, NJ, Princeton Univ Press.

Derviz, A and N Kadlčáková (2002): “Strategic outcomes of credit negotiation and prudential capital regulations”, Charles University, Prague, *CERGE-EI Discussion Paper* no 82, March.

Derviz, A, N Kadlčáková and L Kobzová (2003): “Credit risk, systemic uncertainties and economic capital requirements for an artificial bank loan portfolio”, Czech National Bank, Research Dept, *Working Paper* no 9, November.

Duffie, D and K Singleton (1998): “Simulating correlated defaults”, *working paper*, Graduate School of Business, Stanford University.

Flavin, T and M Wickens (2003): “Macroeconomic influences on optimal asset allocation”, *Review of Financial Economics*, 12, pp 207-31.

Hou, Y (2002): *Integrating market risk and credit risk: a dynamic asset allocation perspective*, mimeo, Yale Univ, Dept of Economics, November.

Jackwerth, J (2000): “Recovering risk aversion from option prices and realized returns”, *Review of Financial Studies*, 13, pp 433-51.

Jarrow, R, D Lando and S Turnbull (1997): “A Markov model for the term structure of credit risk spreads”, *Review of Financial Studies*, vol 10, no 2, pp 481-523.

Jarrow, R A and S M Turnbull (1995): “Pricing derivatives on financial securities subject to credit risk”, *Journal of Finance*, 50(1), pp 53-85.

Madan, D and H Unal (1998): “Pricing the risks of default”, *Review of Derivatives Research*, 2, pp 121-60.

——— (1999): “A two-factor hazard-rate model for pricing risky debt and the term structure of credit spreads”, *working paper*, Smith School of Business, Univ of Maryland, November.

Rosenberg, J and R Engle (2002): “Empirical pricing kernels”, *Journal of Financial Economics*, 64, pp 341-72.

Wickens, M and P Smith (2002): “Macroeconomic sources of FOREX risk”, *CEPR DP* no 3148, January.

A simplified credit risk model for supervisory purposes in emerging markets[1]

Javier Márquez Diez-Canedo[2]

1. Introduction[3]

Currently, the mainstream methodologies that are most widely used to measure credit risk can be divided into two broad categories: mark to market models and default models. The differences between these paradigms rest first on the scope of the losses considered. Whereas in default models an obligor can be in only one of two states, default and non-default, so that losses are exclusively those resulting from debtor defaults, mark to market models also consider losses resulting from a change of value of the loans due to credit quality migration. Further differences arise from the functional forms assumed for the underlying probability distributions, and the way in which these are related to obtain the loan portfolio's loss distribution. For example,[4] in CreditMetrics™, which is a mark to market methodology, the key component is the transition matrix related to a rating system, which provides the probabilistic mechanism that models the quality migration of loans. This determines the losses due to obligor defaults, and the changes in the market value of the loans in the portfolio due to quality migration through a Monte Carlo simulation process, to finally obtain the loss distribution for the portfolio. Whereas the transition matrix, the changes of value, the loss-given-default of the loans, and the migration covariances are theoretically estimated from statistical data and market information, the simulation process relies heavily on a normality assumption around the transition probabilities and Merton's[5] asset value model to establish a relation between credit quality and asset value of the debtor firms, and to determine the joint migration behaviour of the loans in the portfolio.

KMV's[6] methodology is also based on Merton's model[7] and defines a *distance to default*, which is the difference between the value of a company's assets and a certain liability threshold, such that if this quantity is negative, the company is bankrupt and will therefore default on its obligations. For standardisation purposes, this distance to default is measured as a multiple of the standard deviation of the value of the firm's assets. KMV has accumulated a large database, which it uses to estimate default probabilities and correlations, as well as the loss distributions due to debtor default and quality migration. For a specific company, this probability is approximated by the *expected default frequencies*, ie the ratio of the number of companies with the same distance to default that actually defaulted to the total number of companies with the same distance to default in the database. Being a mark to market methodology, it differs significantly from CreditMetrics™ in that it relies on EDFs for each debtor rather than average transition rates as estimated from the historical data produced by the rating agencies. There are also considerable differences in the assumptions and the functional forms utilised.

CreditRisk$^+$ is a default model[8] in which the cornerstone of the methodology is the set of individual default probabilities of the loans in the portfolio. A basic assumption is that the default probabilities are

---

1 An earlier version of this model was published in English in Economia, Societa' e Istituzioni. See Márquez (2002). The model presented here is an updated version with significant differences compared with the original and several new results.

2 Bank of Mexico. The views expressed are those of the author and do not necessarily reflect those of the Bank of Mexico.

3 A good detailed review of the different approaches is presented by Crouhy et al (2000).

4 CreditMetrics™ is a spin-off from the JP Morgan Risk Management systems development group.

5 The reader unfamiliar with the methodology is referred to Section 8 of the CreditMetrics™ technical document and Merton (1974).

6 This is the proprietary methodology of KMV corporation.

7 See Kealhofer (1998, 1999).

8 CreditRisk$^+$ is marketed by Credit Suisse Financial Products.

always small, so that the number of defaults in the portfolio can be approximated according to a Poisson probability distribution. In its more general version, where default probabilities can change over time, it is further assumed that these probabilities are entirely driven by a weighted sum of *K risk factors*, each distributed according to an independent Gamma distribution. The weights of the risk factors differ depending on the individual rating of the obligor and, conditional on these risk factors, individual obligor defaults are assumed to be independent Bernoulli trials. In the general case, default correlation is *implicit* in the covariation behaviour of the risk factors, and *the Poisson assumption* leads to a negative binomial for the distribution of the number of defaults. Having obtained the distribution of the number of defaults in the portfolio, proceeding in the typical actuarial fashion by selecting a unit of loss and given the recovery rates for the individual loans, these are then grouped into *buckets of equal loss-given-default*, and the *probability generating function* of the loss distribution is obtained. From here it is necessary to resort to a numerical recursion procedure to obtain the loss distribution.

Another popular default methodology is Credit Portfolio View,[9] which is a discrete multiperiod model. Apart from the fact that it is conceived from the beginning as a dynamic model, the highlight of the methodology is the determination of default probabilities, which are *logit functions* of indices of macroeconomic variables. The portfolio is segmented according to geographical location and economic activity of the debtors, and the indices for each segment are linear functions of the associated macroeconomic variables for the segment. In turn, each macroeconomic variable is assumed to obey a second-order univariate, autoregressive process, and due to cross-correlations in the error terms of the linear models for the indices and the autoregressive expressions of the underlying macroeconomic variables, the parameters of both are estimated simultaneously from a system of equations. Credit Portfolio View also resorts to simulation on transition matrices to obtain the loss distribution.

All of the above methodologies have contributed greatly to the understanding of the key issues in credit risk modelling and it is now accepted that all models are converging to produce comparable results. Research by Finger (1998), Crouhy et al (2000) and Gordy (2000) discusses how under certain parametric equivalents the mainstream methodologies such as CreditMetrics™ and CreditRisk+ can be mapped into each other. It is important to note that the emphasis in all of these methodologies is on producing a distribution of losses which is as realistic as possible. Although one can hardly argue against this principle, the computational effort required can be impractical for certain users, such as regulators, who have to oversee the whole financial system and not just one individual bank. Furthermore, the development of management tools such as simple rules for establishing capital adequacy, identifying segments of excessive credit risk concentration and setting single obligor limits to loans that are explicitly related to the risk profile of the portfolio is not directly addressed.

The model presented here assumes that the default probabilities of the loans and their covariances are given. From here, a default model is developed which obtains an explicit functional form for the loss distribution, assuming that it can be characterised by two parameters: the mean and the variance. Given a specific mean-variance distribution of losses, not necessarily normal, it is possible to obtain the value-at-risk (VaR) for the portfolio as the expected loss plus a certain multiple of the standard deviation of losses. This leads to a lower bound on the bank's capitalisation ratio and the resulting inequality establishes capital adequacy. The model is developed in a way which explicitly measures the concentration of the loan portfolio. We can see that the *Herfindahl-Hirschman index* emerges naturally as a measure of concentration, providing a precise quantification of how concentration contributes to the overall credit risk of the portfolio. Two new properties of the index are obtained that relate single obligor limits to concentration along different segments of the portfolio so as to ensure capital adequacy. Furthermore, the research shows how correlation affects concentration and this leads to the definition of a *risk concentration measure*. Finally, it is shown that the model can be implemented with limited information on the actual composition of bank loan portfolios, which is a crucial factor for regulators inasmuch as their capacity to obtain up to date and timely information from banks is limited.

Examples of numerical exercises performed to date on real loan portfolios are shown, and are seen to provide results comparable to those obtained using other methodologies, at a considerable reduction in computational effort. Finally, since all the relevant elements for measuring default credit risk are

---

9 This product is offered by McKinsey, the consulting firm. The classic reference is Wilson (1997a,b).

explicitly parameterised, the shortcomings of available information can be compensated by a judicious use of assumptions on the values of the relevant parameters. The computational efficiency of the model results in rapid feedback on the implications and sensitivity of the risk profile of a loan portfolio to changes in the parameters.[10] Since the measurement of concentration is at the heart of the model, we begin with a discussion of this topic.

2. The concentration issue

Loan concentration has long been identified as an important source of risk for banks and loan portfolios. Judging from current technical literature on credit risk, as far as concentration goes the establishment of a generally accepted paradigm has remained elusive in spite of the importance of the problem.[11] The more formal approaches, which look to portfolio theory,[12] have been mainly concerned with optimal diversification of portfolios of traded fixed income assets where information compatible with traditional Markowitz (1959) type models can be obtained in a cost-effective manner. It must be pointed out however, that traditional portfolio theory approaches deal with the concentration issue indirectly, since the preoccupation is the allocation of assets through the well known mean-variance trade-off, but a clear measure of concentration and its relation to risk has never been made explicit. Kealhofer (1998) has an interesting discussion of the issue from the point of view of diversification. First he states that "there has been no method for actually measuring the amount of diversification in a debt portfolio", and that "ex ante, no method has existed which could quantify concentrations"; concentrations have only been detected ex post. He then argues that "measuring the diversification of a portfolio means specifying the range and likelihood of possible losses associated with the portfolio". He goes on to provide a definition that allows the comparison of diversification of two portfolios as:

> "*Portfolio A is better diversified than portfolio B if the probability of loss exceeding a given percent is smaller for A than for B, and both portfolios have the same expected loss*".

Thus, when dealing with portfolios of traditional bank loans, no formal methodology for measuring concentration seems to have emerged. As pointed out by Altman and Saunders (1998), the concentration measurement issue has mainly been dealt with through subjective analysis. Typically, banks and other agents apply a scoring technique based on the opinion of a group of experts about the degree of concentration observed along and across different segments of a portfolio, as regards some classification criterion, in order to obtain an indicator of loan concentration. Generally, the number obtained is of more value in cardinal or hierarchical terms than it is as a direct measure of risk that can quickly be translated into potential losses or value-at-risk.[13]

The approach adopted in the following analysis does not solve all the aforementioned problems, but it does provide a theoretical framework that might allow, ex ante, the detection of risk concentration. The proposed risk concentration measure is consistent with Kealhofer's notion as previously stated. Example 6.2 illustrates how the risk concentration measure can be used to detect the more risky segments of a loan portfolio.

3. Value-at-risk, concentration and the "single obligor limit": the simplest case

Traditionally, banks deal with concentration risk by placing a limit on the maximum amount that can be loaned to a single debtor, along the different dimensions where concentration can occur, ie industry, geographical region, product, country, etc. Normally, the "single obligor limit" is expressed as a

---

10 Due to the closed form expression for value-at-risk, it is also possible to perform analytical exercises.

11 See Caouette et al (1998), Chapters 17 and 18. See also Kealhofer (1998).

12 See, for example, Bennet (1984).

13 See, for example, Moody's Investor Services (1991) and the Coopers and Lybrand (1993) report.

proportion δ of the capital K of the bank. However, when discussing loan concentration, one normally addresses the issue of *how much of the total loans outstanding is concentrated in an individual or group*. Thus, whatever the virtues of setting limits as a percentage of capital, this does not give much information as to the actual concentration of loans in the portfolio. To see this, note that, at least theoretically, a bank could have only one loan that respects the limit but have a totally concentrated portfolio. On the other hand, the bank can have a million uncorrelated loans of exactly the same size, in which case the portfolio would be completely diversified, regardless of whether each loan respects the limit or not. Thus, one can have highly concentrated portfolios as well as highly diversified portfolios that respect the constraint in terms of capital.[14] We will therefore part with tradition, since for the purpose at hand it is better to think of concentration in terms of proportions of the total value of the loan portfolio, and fix limits accordingly. Throughout this paper, individual limits on loans will be expressed as proportions θ of the total value of the loan portfolio V. Furthermore, no generality is lost since δ *and* θ *are linearly related*, so the results are not altered. To see this, let f_k denote the value of the kth of N loans, and analyse the *single obligor limit* as represented by the following constraint:

$$f_k \leq \delta K = \delta \frac{K}{V} \cdot V = \delta\psi V = \theta V; \quad k = 0,1,2,3,\ldots,N \tag{3.1}$$

where $\psi = \frac{K}{V}$ is *the capitalisation ratio*. Thus, $\theta = \delta\psi$,[15] and the single obligor limit will be expressed as:

$f_k \leq \theta V \quad k = 1,2,\ldots,N$

If all loans have the same default probability p, and assuming independence, one can define N binary random loss variables x_i as:

$$x_i = \begin{cases} f_i & \text{with probability } p \\ 0 & \text{with probability } 1-p \end{cases}$$

Clearly $E(x_i) = pf_i$ and $Variance(x_i) = p(1-p)f_i^2$. Since the variables are independent:

a. $$\mu = E\left(\sum_{i=1}^{N} x_i\right) = \sum_{i=1}^{N} pf_i = pV; \quad \text{where } V = \sum_{i=1}^{N} f_i$$

b. $$\sigma^2 = Variance\left(\sum_{i=1}^{N} x_i\right) = \sum_{i=1}^{N} Variance\,(x_i) = p\,(1-p)\sum_{i=1}^{N} f_i^2$$

Since the distribution of loans (f_i) is totally arbitrary, it is difficult to know the exact distribution of $\sum_{i=1}^{N} x_i$. For the moment, assume that it can be approximated by the normal distribution,[16] so that:

$$VAR_\alpha = \mu + z_\alpha \sigma = pV + z_\alpha \sqrt{p(1-p)\sum_{i=1}^{N} f_i^2} \tag{3.2}$$

If $VAR_\alpha \leq K$, after a little algebra one arrives at the following expression:

---

14 For example, if loans are constrained not to exceed 12% of capital, this can be done with only one loan in the portfolio, in which case concentration is maximum. On the other hand, if the portfolio has a thousand loans all representing 12% of capital, it would be a highly diversified portfolio.

15 Note that if there is only one loan in the portfolio, then it is necessarily true that f_i = V so that $\psi\delta = \theta = 1$, which in turn implies that the portfolio is totally concentrated in one loan.

16 See, for example, DeGroot (1988, p 263).

$$H(F)=\frac{\sum_{i=1}^{N} f_i^2}{\left(\sum_{i=1}^{N} f_i\right)^2} \le \frac{(\psi-p)^2}{z_\alpha^2 p(1-p)} = \Theta(p,\psi,\alpha) \qquad (3.3)$$

In this expression, portfolio concentration is measured by:

$$Concentration = H(F) = \frac{\sum_{i=1}^{N} f_i^2}{\left(\sum_{i=1}^{N} f_i\right)^2}$$

Readers familiar with the literature of industrial organisation will have recognised that the above measure is the *Herfindahl-Hirschman* concentration index.[17]

4. Analysis of the capital adequacy inequality

The first observation is that, with the obvious limitations, it seems that portfolio concentration risk can be managed using a very general measure of concentration other than the single obligor limit. Next, it is interesting to note that capital adequacy as represented by the capitalisation ratio ψ requires that

$$\psi \ge p + z_\alpha \sqrt{p(1-p)H(F)} \qquad (4.1)$$

This inequality relates capital adequacy to the probability of default, the confidence level used for value-at-risk, and the concentration index. It also shows that there is a direct relation between the Herfindahl index and the variance of losses. Since the index takes on values between the reciprocal of the number of loans N and one, where high concentration is present the variance of losses will vary between $\sqrt{p(1-p)/N}$ and $\sqrt{p(1-p)}$, depending on $H(F)$. Furthermore, note that the role played by $H(F)$ in the above is totally consistent with Kealhofer's definition of concentration since it is obvious from (4.1) that the lower the value of $H(F)$, the lower the probability of loss exceeding a specified level, for the same expected loss.

In what follows, we can see that everything behaves as it should. The following theorem summarises the main implications for risk managers of the previous analysis. These results are introduced early because they remain basically unchanged throughout all future generalisations.

Theorem 4.1

The bound $\Theta(p,\psi,\alpha)$ on the concentration measure has the following properties:

$\Theta(p,\psi,\alpha)$ *varies in direct proportion to the capitalisation ratio* ψ *and inversely to the default probability* p *and the value-at-risk confidence level* z_α.

If the concentration measure exceeds the bound (ie $H(F) > \Theta(p,\psi,\alpha)$), then the capital of the bank is at risk for the given confidence level.

If the default probability p exceeds the capitalisation ratio ψ, then the capital of the bank is at risk for any confidence level, regardless of the concentration of the loan portfolio.

If $\Theta(p,\psi,\alpha) > 1$, no degree of concentration of the loan portfolio places the capital of the bank at risk.

[17] See, for example, Shy (1995) or Tirole (1995).

Proof

Point one is obvious from the form of $\Theta(p,\psi,\alpha$. The second point is easily verified, ie: if $H(F) > \Theta(p,\psi,\alpha)$ then,

$$VAR_\alpha = \left(p + z_\alpha\sqrt{\theta(F)pq}\right)V > \left(p + z_\alpha\sqrt{\Theta pq}\right)V = \left(p + \frac{z_\alpha\sqrt{pq}(\psi - p)}{z_\alpha\sqrt{pq}}\right)V = K$$

Point three follows directly from 4.1:

$$VAR_\alpha \le K \Leftrightarrow \psi \ge p + z_\alpha\sqrt{p(1-p)H(F)}$$

Point three is also verified easily. If $p > \psi$, then 4.1 is violated:

$$VAR_\alpha = \left(p + z_\alpha\sqrt{H(F)pq}\right)V > \left(\psi + z_\alpha\sqrt{H(F)pq}\right)V = K + z_\alpha V\sqrt{H(F)pq} > K$$

As for point four, it is well known that $H(F) \le 1$ for any arbitrary F.[18]

Capital adequacy Theorem 4.1 provides some useful rules for the risk manager and for the regulator. First, one can determine capital adequacy because one obtains precise measures of the adjustments in the capitalisation ratio required by variations in the default rates and/or the concentration of the loan portfolio. Furthermore, depending on the amount of control that banks have on the default ratio and loan concentration, adjustments in the default probability and the concentration of the loan portfolio necessary to maintain capital adequacy can also be calculated. Thus, if the concentration of the loan portfolio exceeds the bound at the desired confidence level, inequality (3.2) provides a convenient means of fine-tuning the adjustments required in ψ, p and $H(F)$ so that credit risk does not place the capital of the bank in jeopardy. Also interesting is that if the default rate of the portfolio exceeds the capitalisation ratio, the risk manager and the financial authorities are alerted that the banks' capital is at risk regardless of the concentration of the loan portfolio and the confidence level adopted.

5. A closer look at the Herfindahl index

One of the main features of the approach taken is that a measure of loan concentration as it relates to risk arises naturally. The Herfindahl-Hirschman index (HHI) has been extensively studied in relation to industrial concentration, and it is known to have several important properties. Thus, it is known that the index takes values between the reciprocal of N and one,[19] and that it behaves well in terms of "*the five properties of inequality measures*".[20] We now investigate how the HHI relates to the intuitive notion that concentration is related to the minimum number of obligors where credit is more concentrated. A better understanding of the relation between the single obligor limit and the concentration index has important risk management and regulatory implications.

In order to examine how concentration relates to the notion that more credit in fewer hands means more concentration, it must be consistent with the notion that maximum concentration occurs when all credit is held by a single obligor and the minimum is when all debtors owe the same amount. Formally:

a. The maximum concentration occurs when, for some i, one has that:

$$f_j = \begin{cases} V & \text{for } j = i \\ 0 & \text{for } j \ne i; \quad j = 1,2,\ldots,N \end{cases}$$

[18] See Encaoua and Jacquemin (1980).

[19] A simple normalisation is possible, from which we can easily see that $\phi(F)$ as defined below satisfies $0 \le \phi \le 1$.

$$\phi(F) = \frac{N - \frac{1}{H(F)}}{N - 1}$$

[20] See Cowell (1995) and Encaoua and Jacquemin (1980).

ie $F^{\max} = Ve^i$, where $e^i \in E^N$ is the ith unit vector.

b. The minimum concentration occurs when $f_i = \dfrac{V}{N}$ for $i = 1,2,\ldots,N$

Concentration has to do with numbers, and the HHI has several interesting *numbers-related properties*. The best known is *Adelman's* "*numbers-equivalent*",[21] which for loan concentration states that its inverse can be interpreted as "*the minimum number of loans of equal size that would result in a specific value of the index*". It is now shown that *the value of the index is maximised under the single obligor limit, when all credit is concentrated in the minimum number of obligors, and each obligor holds credit up to the limit.* The theorem establishes the relation between the single obligor limit and the Herfindahl-Hirschman measure of concentration, and in so doing, it shows that Adelman's numbers-equivalent is in fact the maximum concentration possible, when loans are constrained by a certain limit.[22] In what follows, we let F denote the vector of loans $f_k \geq 0$ for $k = 1,2,\ldots,N$. Without loss of generality, we also assume that the elements of this vector have been sorted in decreasing order:

$f_1 \geq f_2 \geq \ldots \geq f_N$

We can also assume that $V = \sum_{k=1}^{N} f_k = 1$. The following proposition is an important basic property of the index.

Proposition 5.1

Assume $F = (f_k)$ is such that $f_i \geq f_{i+1} \geq 0$ for $i = 1,2,3,\ldots,N-1$ and $\sum_{i=1}^{N} f_i = 1$. Then:

a. For f_i, f_j such that $1 \leq i \leq j$; $f_j > 0$ and $\varepsilon > 0$ such that $f_j - \varepsilon > 0$ define the vector $F' = (f_k')$ to be

$$f_k' = \begin{cases} f_k; & k = 1,2,\ldots,N; \quad k \neq i,j \\ f_i + \varepsilon; & k = i \\ f_j - \varepsilon; & k = j \end{cases}$$

then $H(F') > H(F)$.

b. If $f_i > f_j$ and $0 < \varepsilon \leq f_i - f_j$, then the vector $F'' = (f_k'')$ defined as:

$$f_k'' = \begin{cases} f_k; & k = 1,2,\ldots,N; \quad k \neq i,j \\ f_i - \varepsilon; & k = i \\ f_j + \varepsilon; & k = j \end{cases}$$

has the property $H(F'') < H(F)$.

Proof

To prove (a), simply note that:

$$H(F') - H(F) = \sum_{k=1}^{N} \left[(f_k')^2 - f_k^2 \right] = 2\varepsilon \left[(f_i - f_j) + \varepsilon \right] > 0$$

The Proof of (b) is similar: $H(F'') - H(F) = 2\varepsilon \lfloor (f_i - f_j) - \varepsilon \rfloor \leq 0$ since $\varepsilon \leq (f_i - f_j)$

(note that $\varepsilon > f_i - f_j$ implies case (a)).

21 See Adelman (1969) and Kelly Jr (1981). For other interesting numbers-related properties, see Weinstock (1984).

22 Although the result conforms to intuition, no formal proof has been detected by the authors in the more frequent references, such as Sleuwaegen et al (1989), Weinstock or Encaoua and Jaquemin op cit.

The proposition states that if some element f_k is increased at the expense of decreasing a smaller element f_j, the concentration index will increase. If on the other hand, an element is increased at the expense of a larger element, then the concentration index will decrease. To continue with the analysis, it is now shown that if all credit is concentrated in the minimum number of debtors, while subject to the constraint $f_k \le \theta V$, then $H(F) \le \theta$.

Proposition 5.2

Let $\theta \in (0, 1)$ and $n = \left\lfloor \frac{1}{\theta} \right\rfloor$ be the integer part of $\frac{1}{\theta}$. Let $\varepsilon \in [0, 1)$ be such that $\theta = \frac{1-\varepsilon}{n}$. Then, for the distribution,

$$f_k = \begin{cases} \theta; & k = 1, 2, \ldots, n \\ \varepsilon; & k = n+1 \\ 0 & else \end{cases}$$

we have that $H(F) \le \theta$.

Proof

Note that $\sum f_k = n\theta + \varepsilon = 1$ and therefore:

$H(F) = n\theta^2 + \varepsilon^2 = n\theta^2 + (1 - n\theta)^2 = n(n + 1)\theta^2 - 2n\theta + 1$

For $H(F) = \theta$ one must solve the quadratic equation,

$$(n + 1)n\theta^2 - 2n\theta + 1 = \theta \quad \text{ie} \quad n(n + 1)\theta^2 - (1 + 2n)\theta + 1 = 0 \tag{5.1}$$

It is simple to verify that (5.1) has the following two solutions:

$$\theta_1 = \frac{1}{n} \text{ and } \theta_2 = \frac{1}{n+1}$$

This means that if θ^{-1} is an integer, then $H(F) = \theta$. Thus, examine what happens in the interval $\left(\frac{1}{n+1}, \frac{1}{n}\right)$. To do this, let

$$\theta(\lambda) = \lambda \left(\frac{1}{n}\right) + (1-\lambda) \frac{1}{n+1} = \frac{n+\lambda}{n(n+1)} \quad \text{with } \lambda \in (0,1)$$

Substituting $\theta(\lambda)$ in the left-hand side of (5.1), one obtains:

$$\begin{aligned} n(n+1)\left(\frac{n+\lambda}{n(n+1)}\right)^2 - (1+2n)\frac{n+\lambda}{n(n+1)} + 1 &= \frac{1}{n(n+1)}\left\{n^2 + 2\lambda n + \lambda^2 - n(1+2n) - \lambda(1+2n) + n^2 + n\right\} \\ &= \frac{\lambda(\lambda-1)}{n(n+1)} < 0 \quad \forall \lambda \in (0,1) \end{aligned}$$

It is now shown that if all loans respect the single obligor limit $f_k \le \theta V$, then $H(F) \le \theta$ and the distribution of loans of the previous proposition maximises the value of the index under the single obligor constraint.

Theorem 5.3

Let $F = (f_k)$ be such that:

$$f_k = \begin{cases} \theta; & k = 1,2,\ldots,n \\ \varepsilon; & k = n+1 \\ 0; & k = n+2,\ldots,N \end{cases}$$

with $\theta, \varepsilon \geq 0$; $\varepsilon < \theta$ and $\sum f_k = 1$. Then F maximises $H(F)$ for all F such that $f_k \leq \theta \ \forall\ k$ and $H(F) \leq \theta$.

Proof [23]

Proposition 5.2 states that $H(F) \leq \theta$ for this distribution. Necessarily, $n = \left[\frac{1}{\theta}\right]$ and $\varepsilon \geq 0$ are such that $\theta = \frac{1-\varepsilon}{n}$ in order to have $\sum f_k = 1$. Furthermore, any vector with $f_k' = \theta + \delta;\ \delta > 0$ would violate the constraint $f_k \leq \theta \ \forall\ k$. Therefore, the only possibility of altering the distribution of loans would be to decrease some element $f_k = \theta$ or $f_{n+1} = \varepsilon$ by some quantity $\delta > 0$. But then proposition 5.1(b) states that $H(F') < H(F) \leq \theta$.

This result has important implications for risk management and regulation since de facto it states that by placing a limit on individual loans as a propotion of the value of the portfolio, one is also placing a limit on concentration as measured by the HHI by the same amount θ. Therefore, it is simple to check for capital adequacy by

$$\theta \leq \frac{(\psi - p)^2}{z_\alpha^2 p(1-p)} = \Theta(p, \psi, \alpha) \tag{5.2}$$

Alternatively, from (4.1), one can obtain the capital adequacy relation in terms of the single obligor limit (2.6), that is:

$$\psi \geq p + z_\alpha \sqrt{p(1-p)\theta} \tag{5.3}$$

Thus, (5.2) provides a very simple means to check for capital adequacy, without performing complicated calculations. Although crude, simply take θ to be the ratio of the largest loan to the total value of the loan portfolio and the observed default rate as an ex post proxy of default probability and substitute these values in the right-hand side of (5.1). Since Theorem (5.1) guarantees $H(F) \leq \theta$, if the inequality holds it is a good sign that the bank is adequately capitalised.

It should be realised however, that this condition is sufficient but not necessary. As will be shown in the following theorem, if one chooses to explicitly constrain the portfolio to satisfy $H(F) \leq \theta$, it is possible to have specific loans that as a proportion of the total value of the portfolio represent a quantity larger than θ. Intuitively, granting a very large loan while satisfying the constraint on the index is only possible at the expense of the other loans in the portfolio so that in the optimum, the portfolio is composed only of one large loan and all others are small and of equal size.

Theorem 5.4

If $H(F) \leq \theta$ then:

$$f_i \leq \frac{1}{N}\left(1 + \sqrt{(N\theta - 1)(N-1)}\right) < \sqrt{\theta} \quad \text{for } i = 1,2,3,\ldots,N$$

Proof

The idea behind the proof is that under the constraint $H(F) \leq \theta$, a very large loan is only possible at the expense of all the other loans, which must become progressively smaller and of equal size. So, given the constraint $H(F) \leq \theta$, let us maximise the largest element f_1. Suppose $f_1 = a$ is the largest loan possible, then necessarily $f_2 = f_3 = \ldots = f_N = b$; for some $b > 0$; $b < a$. To see this, consider any other distribution with $f_i > f_j$ and $1 < i < j$. Then there exists $\varepsilon > 0$ such that $f_i' = f_i - \varepsilon > f_j' = f_j + \varepsilon > 0$. Proposition 5.1 then states that $H(F') < H(F) \leq \theta$. Now, by continuity of the index on each f_i and

[23] This proof and the one for the next theorem are different from the original proofs in Márquez (2002). They are due to Fausto Membrillo and are more intuitive and elegant than the original.

because of Theorem 5.3, there exists $\varepsilon' > 0$ such that any loan distribution F'' with $f_1'' = f_1 + \varepsilon'$ and $f_j'' = f_j' - \varepsilon' \geq 0$ satisfies $H(F') < H(F'') \leq \theta$, which contradicts the assumption that F is a distribution where f_1 is a maximum. Therefore, if $f_1 = a$, for some $a > 0$, the loan distribution which maximises f_1, subject to the constraint $H(F) \leq \theta$, can be represented as

$$f_k = \begin{cases} a; & k = 1 \\ b; & k = 2, 3, \ldots, N \end{cases}$$

and $a > b$; therefore:

$$H(F) = a^2 + (N-1)b^2 \leq \theta \tag{5.4}$$

Furthermore $a + (N-1)b = V$. Solving for b:

$$b = \frac{1-a}{N-1}$$

Substituting b in (5.4) one obtains:

$$a^2 + (N-1)\left(\frac{a-1}{N-1}\right)^2 \leq \theta$$

This leads to the following quadratic equation:

$$Na^2 - 2a + \lfloor 1 - \theta(N-1) \rfloor \leq 0 \tag{5.5}$$

Equating to zero, the solution of (5.5), yields:

$$a = \frac{1}{N}\left[1 + \sqrt{(N\theta - 1)(N-1)}\right]$$

Note that $a \to \sqrt{\theta}$ when $N \to \infty$, and it is simple to obtain the last inequality:

$$\begin{aligned}(\theta-1)^2 > 0 &\Leftrightarrow \theta^2 - 2\theta + 1 > 0 \Leftrightarrow \theta^2 + 2\theta + 1 > 4\theta \Leftrightarrow (\theta+1)^2 > 4\theta \Leftrightarrow \theta + 1 > 2\sqrt{\theta} \\ &\Leftrightarrow -N(\theta+1) < -2\sqrt{\theta}N \Leftrightarrow N^2\theta - N(\theta+1) + 1 < N^2\theta - 2\sqrt{\theta}N + 1 \\ &\Leftrightarrow N^2\theta - N\theta - N + 1 < \left(\sqrt{\theta}N - 1\right)^2 \Leftrightarrow (N\theta-1)(N-1) < \left(\sqrt{\theta}N - 1\right)^2 \\ &\Leftrightarrow \sqrt{(N\theta-1)(N-1)} < \sqrt{\theta}N - 1 \Leftrightarrow \frac{1}{N}\left(1 + \sqrt{(N\theta-1)(N-1)}\right) < \sqrt{\theta}\end{aligned}$$

Having a good concentration index is desirable from the regulatory point of view, since it facilitates comparisons of loan concentration between different institutions, and leads to an assessment of concentration risk for the financial system as a whole. For the risk manager of an individual bank, besides measuring his own risk, it provides benchmarks for setting business strategy and goals, and allows comparisons with the competition. The HHI seems particularly well suited for the task, since besides measuring concentration it is directly related to risk, and provides a quick means to check capital adequacy. In the following section it will be seen that the concept is robust under much more general conditions.

5.1 A numerical example

In order to illustrate the results obtained so far, consider the following example taken from the CreditRisk$^+$ manual:

Table 5.1

No of loans	Rating							Total
	A	**B**	**C**	**D**	**E**	**F**	**G**	
1	$4,728	$5,528	$3,138	$5,320	$1,800	$1,933	$358	$22,805
2	$7,728	$5,848	$3,204	$5,765	$5,042	$2,317	$1,090	$30,994
3			$4,831	$20,239	$15,411	$2,411	$2,652	$45,544
4			$4,912			$2,598	$4,929	$12,439
5			$5,435				$6,467	$11,902
6							$6,480	$6,480
Total	$12,456	$11,376	$21,520	$31,324	$22,253	$9,259	$21,976	$130,164

Default probabilities for the loans are taken from the following table:

Table 5.2

Rating	**Default prob**
A	1.65
B	3.00
C	5.00
D	7.50
E	10.00
F	15.00
G	30.00

For this first example let the default probability for the loans be the weighted average of the probabilities of Table 5.2; that is 10.89%. The HHI for the portfolio is 6.61%. Assuming normality and choosing a 5% confidence level, $z_\alpha = 1.96$ and one obtains:

$$\psi \geq p + z_\alpha \sqrt{p(1-p)H(F)} = 0.1089 + 1.96\sqrt{0.1089 \times 0.8911 \times 0.0661} = 0.2658$$

Then the bank's economic capital must be at least:

$$VaR_{0.05} = 0.2658 \times V = 0.2658 \times \$130{,}164.00 = \$34{,}602.79$$

Suppose economic capital is $35,000, then the capitalisation ratio is:

$$\psi = \frac{K}{V} = \frac{35{,}000}{130{,}164} = 0.2689$$

Since $0.2689 > 0.2658$, the bank exhibits capital adequacy. Now, under 3.2, the maximum concentration that the portfolio can assume is:

$$\frac{(\psi - p)^2}{z_\alpha^2 p(1-p)} = \frac{(0.2689 - 0.1089)^2}{1.96^2 \times 0.1089 \times 0.8911} = 0.0687$$

Since $H(F) = 6.61\%$, the portfolio is not excessively concentrated.

Since the maximum value of the index is 6.87%, no loan in the portfolio should exceed:

$f^* = \sqrt{0.0687} \times V = 0.2621 \times 130{,}164 = 34{,}107.88$

Table 5.1 shows that the largest loan is the $20,239 D-loan, which is smaller than the aforementioned amount. It is interesting to note that the single obligor limit would be violated. According to Theorem 5.2, loans should not exceed:

$f_i \leq 0.0687 \times 130{,}164 = \$8{,}942.27$

There are two loans in the portfolio that are greater than this amount: the $20,239 D-loan and the $15,411 E-loan, confirming that the condition is sufficient but not necessary. Finally, we can see that the largest loan in the portfolio is within the bounds provided by Theorem 5.2, ie $\$8{,}942.27 \leq \$20{,}239 \leq \$34{,}107.88$.

6. Accounting for default correlation and different default probabilities

The results obtained so far rely on the following assumptions:

a. The loss distribution can be characterised by its mean and variance.

b. Default probabilities are homogeneous and independent from each other, for all loans along the dimension where loan concentration can occur.

c. There is only one dimension of possible loan concentration.

d. Nothing is recovered from defaulting loans.

In this section the model is generalised by relaxing the second and third assumptions. We first examine the case where default probabilities can be different and are correlated.

6.1 A general model

Assume that the portfolio loss distribution can be characterised by its mean and its variance and that the vector of default probabilities π and the covariance matrix M are given exogenously. Proceeding along the same lines of the previous analysis, the VaR to capital inequality is now:

$$VAR_\alpha = \pi^T F + z_\alpha \sqrt{F^T M F} \leq K \qquad (6.1)$$

Since M is positive definite, it is well known that there exists a matrix Q such that,

$$M = Q \Lambda Q^T \qquad (6.2)$$

where Λ is the diagonal matrix of characteristic values of M, and Q is an *orthogonal matrix of the eigenvectors of* M, with the property that $Q^{-1} = Q^T$.[24] Let $S = Q\sqrt{\Lambda}\,Q^T$, where $\sqrt{\Lambda}$ is the diagonal matrix of the square roots of the eigenvalues of M, so that $M = S^T S$. Now change the variable to $G = SF$ so that $F^T M F = G^T G$. This change of variable effectively *rescales* F in terms of the matrix S which in turn is representative of the "*square root*" of the covariance matrix M. It is well known that this is equivalent to rescaling the loans in the portfolio according to the covariances of the default probabilities between the loans, so that loans with higher loss covariances will increase in size, while the opposite will happen to loans with smaller loss covariances. Although much credit in few hands is potentially dangerous, it is even more dangerous when *too much risk is concentrated in a particular group of debtors*, as suggested by the rescaling of the loan portfolio in terms of S. Thus, at a given moment a numerically highly diversified portfolio of small loans that exhibit large variances and are highly correlated may be riskier than a numerically small portfolio of large loans that are uncorrelated and have low default probabilities. In the next section, the discussion is taken a step further.

---

[24] Any intermediate text on matrix theory can be consulted. See, for example, Strang (1988), or Mirsky (1990).

To continue with the development of the model, multiplying and dividing F^TMF by F^TF, and dividing by $V = 1^TF$, the following capital adequacy relation, relative to the value of the loan portfolio, is obtained:

$$\psi \geq \bar{p} + z_\alpha \sqrt{\frac{F^TMF}{F^TF} H(F)} = \bar{p} + z_\alpha \sigma \sqrt{H(F)} \qquad (6.3)$$

where

$$\sigma^2 = \frac{F^TMF}{F^TF} = R(F,M) = \textit{Rayleigh's quotient} \qquad (6.4)$$

is a measure of the standard deviation of losses and

$$\bar{p} = \frac{\pi^T F}{V} \qquad (6.5)$$

is the expected loss of the portfolio relative to its value which is nothing more than the weighted average of default probabilities. Proceeding in the usual way, and applying Theorem 5.1, one obtains a limit on concentration and single obligor limits as:

$$H(F) \leq \theta \leq \left(\frac{\psi - \bar{p}}{z_\alpha \sigma} \right)^2 \qquad (6.6)$$

Note that relations (6.3) and (6.6) have the same structure as those obtained for the simple cases of equal default probabilities and independent loans. In this general case, Rayleigh's quotient measures the variance of losses. One can verify that this reduces to the case of equal default probabilities for all loans and uncorrelated defaults, and that all the results of Theorem 4.1 are still true under this generalisation.

Note that the total variance of losses $\sigma\sqrt{H(F)}$ is decomposed into the variation-covariation effect, represented by σ, and concentration $H(F)$. This emphasises the fact that resizing the loan vector through the covariance matrix M implies that concentration in the number of loans is not necessarily a good measure of risk concentration.

6.2 A measure of risk concentration

In order to investigate how correlation affects concentration and increases risk, consider the special case when all loans have the same default probability p and each pair of loans is similarly correlated through ρ. Then, the covariance of defaults between any two loans (i, j) is:

$$\sigma_{ij} = \sigma_i \sigma_j \rho_{ij} = \sqrt{p_i(1-p_i)}\sqrt{p_j(1-p_j)}\,\rho_{ij} = p(1-p)\rho \quad \forall\ i, j \qquad (6.7)$$

In this case the covariance matrix has the following structure:

$$M = p \cdot (1-p) \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho \\ \rho & \cdots & \rho & 1 \end{pmatrix} \qquad (6.8)$$

It is convenient to represent this as:

$$M = p(1-p)\{\rho \mathbf{1}\mathbf{1}^T + (1-\rho)I\}, \qquad (6.9)$$

"**1**" is the "sum vector" ie $\mathbf{1}^T = (1,1,1,...,1)$ and "I" is the identity matrix.

Thus, the variance of losses of the portfolio is:

$$F^TMF = p(1-p)\{\rho(\mathbf{1}^TF)^2 + (1-\rho)F^TF\}$$

Proceeding in the usual way, and noting that $V = \mathbf{1}^TF$, this leads to a VaR of:

$$VaR = V\left\{p + z_\alpha \sqrt{p \cdot (1-p)}\sqrt{\rho + (1-\rho)H(F)}\right\} \qquad (6.10)$$

In this expression, loss variance is decomposed into two distinct elements. The first is the Bernoulli variance $p(1 - p)$, while concentration is captured by:

$$H' = \rho + (1 - \rho)H(F) \tag{6.11}$$

Note that under positive correlation, H' can be interpreted as a convex combination between the HHI of a totally concentrated portfolio ($H(.) = 1$) and the HHI of the portfolio $H(F)$. Clearly, H' increases with ρ and for $\rho = 0$ we have $H' = H(F)$; whereas $H' = 1$ if $\rho = 1$. In other words, if all the loans of a portfolio are perfectly and positively correlated, in terms of risk they behave as a single loan. In general, one can say that the correlated portfolio behaves exactly the same as an uncorrelated portfolio, whose concentration index is H', instead of $H(F)$. Thus, H' could be considered a *correlation-adjusted concentration index*.

Furthermore, (6.11) can be used to compute such an index for any given portfolio by computing p and ρ such that:

$$p(1 - p)\cdot H' = p\cdot(1 - p)\cdot[\rho + (1 - \rho)H(F)] = R(M,F)\cdot H(F) \tag{6.12}$$

Letting $p = \dfrac{\pi^T F}{V}$, solving for ρ gives:

$$\rho = \frac{\left(\dfrac{R(M,F)}{p\cdot(1-p)} - 1\right)}{\left(\dfrac{1}{H(F)} - 1\right)} = \frac{\left[R(M,F) - p\cdot(1-p)\right]H(F)}{p\cdot(1-p)\left[1 - H(F)\right]} \tag{6.13}$$

The expression provides an equivalent correlation measure which summarises how loan defaults are pairwise correlated within the portfolio.

Example 6.1

Consider the loan portfolio of the previous examples. The correlation matrix used in this exercise is as shown in Appendix A, and is segmented into three groups:

$$M = \begin{bmatrix} M_1 & C_{1,2} & C_{1,3} \\ C_{2,1} & M_2 & C_{2,3} \\ C_{3,1} & C_{3,2} & M_3 \end{bmatrix}$$

Assuming normality and a 5% confidence level, VaR is:

$$VaR_{0.05} = \pi^T F + z_{0.05}\sqrt{F^T MF} = 14{,}179 + 1.96(21{,}176) = \$55{,}683$$

From previous examples we know that $p = 0.1089$, $H(F) = 0.0661$, and computation yields:

$$\sigma = \sqrt{\frac{F^T MF}{F^T F}} = \sqrt{0.4006} = 0.6329$$

Thus, capital adequacy requires:

$$\psi > \overline{p} + z_\alpha \sigma\sqrt{H(F)} = 0.4278$$

Assume $K = \$60{,}000$, so that $\psi = \dfrac{60{,}000}{130{,}164} = 0.4610.$ Relation (1.5) provides single obligor limits:

$$\theta \le \left(\frac{\psi - \overline{p}}{z_\alpha \sigma}\right)^2 = \left(\frac{0.4610 - 0.1089}{1.96(0.6329)}\right)^2 = 0.0805$$

That is:

$f_i \le 0.0805 \times \$130{,}164 = \$10{,}482$

From Table 5.1 we can see that there are only two loans that exceed the limit.

Let us now examine the impact of correlation on concentration. From (6.13):

$$\rho = \frac{[0.4006 - 0.0978] \times 0.0661}{0.0978 \times [1 - 0.0661]} = 0.2191$$

From (6.11), the risk concentration index is:

$H' = 0.2191 + (1 - 0.2191) \times 0.0661 = 0.2707$

Beside the fact that the portfolio of this example is a pretty bad one, if one adds 22% correlation to the high default probability of 10.89% one obtains unexpected losses of $\sigma\sqrt{H(F)} = 0.1627$, as opposed to $\sqrt{p(1-p)H(F)} = 0.0801$ if the loans were independent. Thus, the 22% equivalent correlation doubles the standard deviation of losses over the uncorrelated case. It is also interesting to compare the risk concentration index of $H' = 27.07\%$, which is four times greater than $H(F) = 6.61\%$. In terms of capital adequacy, the correlated portfolio requires a capitalisation ratio $\psi \geq 43\%$, which is substantially greater than the 27% required if the loans were independent.

6.3 Dealing with different dimensions of concentration

Generally, banks partition loan portfolios into subportfolios or "buckets" according to some practical criterion which is somehow related to the way in which they do business. For the purpose of credit risk in general and concentration in particular, it may be desirable to adopt a different criterion. As mentioned initially, one of the most difficult problems is to *determine ex ante* potentially dangerous *dimensions of concentration*, and these may have nothing to do with the organisational structure of the bank. The model permits a totally arbitrary segmentation of the portfolio, in order to determine the segments where concentration is potentially riskier. This permits the differentiation of limits for each segment, as well as differentiation in the allocation of capital.

6.3.1 The analysis of individual segments

Suppose that F is arbitrarily partitioned into h segments, $F^T = (F_1,\ldots,F_h)$, where F_i is a vector whose elements are the amounts outstanding of the loans in group i. Now partition the default probability vector and the associated covariance matrix accordingly:

a. $\pi = (\pi_i)$; where π'' is the vector of default probabilities of segment i; $i = 1,2,3,\ldots,h$

b. The covariance matrix is partitioned as:

$$M = \begin{bmatrix} M_1 & C_{12} & \cdots & C_{1h} \\ C_{21} & M_2 & \cdots & C_{2nh} \\ \vdots & \vdots & \ddots & \vdots \\ C_{h1} & C_{h2} & \cdots & M_h \end{bmatrix}$$

Each diagonal block M_i is the covariance matrix of defaults for the loans in segment i and has dimension $(N_i \times N_i)$, where N_i is the number of loans in the segment. Matrices C_{ij} contain the covariances of the defaults between the loans of segments i and j. Let $V_i = \sum_{j \in F_i} f_j$ be *the value of the portfolio of segment* i, and $\sum_{i=1}^{h} V_i = V$. Let $K_i = \gamma_i K$, where γ_i is the proportion of capital allocated to segment i; $\gamma_i \in [0,1]$ $\forall\, i$; $\sum_{i=1}^{h} \gamma_i = 1$. Note that when analysing individual segments, only correlations between defaults of the loans in segment i with loans of the other groups should be considered, while correlations of other groups between themselves are irrelevant. Thus, from M construct matrices S_i with the following structure:

$$S_i = \frac{1}{2}\begin{bmatrix} 0 & \cdots & C_{1i} & \cdots & 0 \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ C_{i1} & \cdots & 2M_i & \cdots & C_{ih} \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ 0 & \cdots & C_{hi} & \cdots & 0 \end{bmatrix} \tag{6.14}$$

Note that $\sum_i S_i = M$. When integrating the analysis of individual segments into the overall portfolio, it is important that the relative weights of each segment in the overall portfolio do not distort the results for the portfolio as a whole. An additivity property is necessary so that addition of over individual segments is consistent for the portfolio. Let

$$\phi = \frac{\sqrt{F^T MF}}{\sum_{i=1}^{h}\sqrt{F^T S_i F}} \tag{6.15}$$

In what follows, we will see that this constant permits the summation of the individual VaR_i. Proceeding in the usual way, the value-at-risk inequality for each segment is:

$$\nu_i = \pi_i^T F_i + z_\alpha \phi \sqrt{F^T S_i F} \le K_i = \gamma_i K \quad \text{for } i = 1, 2, \ldots, h \tag{6.16}$$

where $\gamma_i \ge 0$ and $\sum_i \gamma_i = 1$. It is easily verified that $\sum_i \nu_i = VaR_\alpha = \pi^T F + z_\alpha \sqrt{F^T MF}$.

Dividing by V_i, leads to capital adequacy for each individual segment:

$$\psi_i \ge \frac{\nu_i}{V_i} = \bar{p}_i + z_\alpha \phi \sqrt{R_i(F_i, M_i)\, H(F_i) + \frac{1}{(1^T F_i)^2}\sum_{\{j\,/\,j \ne i\}} F_i^T F_{ij} F_j} \tag{6.17}$$

Solving for $H(F_i)$ one obtains

$$H(F_i) \le \left(\frac{\psi_i - \bar{p}_i}{z_\alpha \phi \sigma_i}\right)^2 - \frac{1}{(\sigma_i V_i)^2}\sum_{\{j | j \ne i\}} F_i^T\ C_{ij}\ F_j \tag{6.18}$$

where

$$\sigma_i = \sqrt{\frac{F_i^T M_i F_i}{F_i^T F_i}} = \sqrt{R_i(F_i, M_i)} \tag{6.19}$$

Single obligor limits per segment are obtained by applying Theorem 5.3:

$$\theta_i \le \left(\frac{\psi_i - \bar{p}_i}{z_\alpha \phi \sigma_i}\right)^2 - \frac{1}{(\sigma_i V_i)^2}\sum_{\{j | j \ne i\}} F_i^T\ C_{ij}\ F_j \tag{6.20}$$

It is interesting to note that the bound on concentration now includes a correction for default correlation with the loans in other groups: namely, the second term on the right-hand side of the inequality. This conforms to intuition, since higher correlation of defaults with the loans in the other groups means that less concentration can be tolerated in group i, namely:

$$\frac{1}{(\sigma_i V_i)^2}\sum_{\{j | j \ne i\}} F_i^T\ C_{ij}\ F_j \tag{6.21}$$

6.3.2 Overall capital adequacy in a segmented portfolio

Note that all of the above expressions are obtained from ν_i / V_i, so that the weight of the segments within the portfolio is not accounted for. Therefore, a simple summation of terms can be misleading as to the overall capital adequacy of the segmented portfolio. Letting $\gamma_i = \frac{V_i}{V}$, then if 6.17 is satisfied for all the segments, $\psi = \sum_{i=1}^{h} \gamma_i \psi_i$ ensures capital adequacy for the portfolio.

Example 6.2

Refer to the portfolio of the previous examples. The partition is shown in Table 6.1.[25] The loans vector is partitioned as: $F^T = (F_1 \ \ F_2 \ \ F_3)$,

Table 6.1

Rating	F_1	Rating	F_2	Rating	F_3
A1	$4,728	B1	$5,528	A2	$7,728
C2	$3,204	C1	$3,138	B2	$5,848
C4	$4,912	C3	$4,831	C5	$5,435
D1	$5,320	E2	$5,042	D2	$5,765
D3	$20,239	E3	$15,411	E1	$1,800
F1	$1,933	F3	$2,411	F2	$2,317
F4	$2,598	G1	$358	G3	$2,652
G2	$1,090	G5	$6,467	G4	$4,929
				G6	$6,480
Total	$44,024	Total	$43,186	Total	$42,954

Next, the default probabilities vector and the covariance matrix are partitioned to be consistent with the partition of the loans vector as:

$\pi^T = (\pi_1 \quad \pi_2 \quad \pi_3)$ and $M = \begin{bmatrix} M_1 & C_{1,2} & C_{1,3} \\ C_{2,1} & M_2 & C_{2,3} \\ C_{3,1} & C_{3,2} & M_3 \end{bmatrix}$ where:

- M_1, M_2, and M_3 are the idiosyncratic covariance matrices for the three groups respectively.
- $C_{12}^T = C_{21}$ is the covariance matrix between the loans of groups one and two. Likewise, $C_{13}^T = C_{31}$ is the covariance matrix between the loans of the first and third groups and $C_{23}^T = C_{32}$ is the covariance matrix between the loans of the second and third (see Appendix A).

Table 6.2 shows the value of the loans of each segment, the corresponding HHI, and the associated capital allocation γ_i.

Table 6.2

Segment *i*	V_i	$H(F_i)$	γ_i	K_i
1	$44,024	0.2613	0.3382	$20,293
2	$43,186	0.2008	0.3318	$19,907
3	$42,954	0.1293	0.33	$19,800

[25] A1 is the first A-rated loan, C2 is the second C-rated loan and so on.

Refer to Appendix A for the variance covariance matrix used for this example. The S_i matrices for each segment have the form:

$$S_1 = \frac{1}{2}\begin{bmatrix} 2M_1 & C_{12} & C_{13} \\ C_{21} & 0 & 0 \\ C_{31} & 0 & 0 \end{bmatrix}, S_2 = \frac{1}{2}\begin{bmatrix} 0 & C_{12} & 0 \\ C_{21} & 2M_2 & C_{23} \\ 0 & C_{32} & 0 \end{bmatrix} \text{ and } S_3 = \frac{1}{2}\begin{bmatrix} 0 & 0 & C_{13} \\ 0 & 0 & C_{23} \\ C_{31} & C_{32} & 2M_3 \end{bmatrix}.$$

Note that $\psi_i = \frac{K_i}{V_i} = \frac{\gamma_i \times K}{V_i} = \frac{K}{V} = \frac{60{,}000}{130{,}164} = 0.4610$ for all segments, since $\gamma_i = V_i/V$.

From 6.15, parameter ϕ, which allows summation of individual VaRs, is:

$$\phi = \frac{\sqrt{F^T MF}}{\sum_{i=1}^{3}\sqrt{F^T S_i F}} = 0.5783$$

Calculation of ν_i with 6.16, using a 5% confidence limit and assuming normality, yields:

$\nu_1 = \$16{,}255 < K_1 = \$20{,}293$,

$\nu_2 = \$19{,}368 < K_2 = \$19{,}907$,

$\nu_3 = \$20{,}060 > K_3 = \$19{,}800$.

First note that:

$$VaR_\alpha = \sum_{i=1}^{3}\nu_i = \$55{,}683$$

Moreover,

$$\psi = 0.4610 \geq \sum_{i=1}^{3}\frac{\nu_i}{V_i} = \frac{VaR}{V} = \frac{55{,}684}{130{,}164} = 0.4278$$

Thus, the portfolio as a whole exhibits capital adequacy, in spite of the fact that the third segment does not comply with its individual capital requirement. This means that the segment will not satisfy any of the other conditions. Using the data in Tables 6.1 and 6.2, the equivalent correlation for each segment is calculated from equation (6.13) and the risk concentration measure from (6.11). The results are summarised in Table 6.3:

Table 6.3

p	**ρ**	***H*(*F*)**	***H′***	***H′*/*H*(*F*)**	**Loss std dev**
0.0774	0.1404	0.2613	0.365	1.3969	0.1614
0.1162	0.1746	0.2008	0.3403	1.6947	0.1869
0.1339	0.2792	0.1293	0.3724	2.8801	0.2078

With these values, one can verify all the capital adequacy relations. As was to be expected, the third segment does not comply with the limit on concentration.

$$H(F_3) = 0.1293 > \left(\frac{\psi_i - \bar{p}_i}{z_\alpha \phi \sigma_i}\right)^2 - \frac{1}{(\sigma_i V_i)^2}\sum_{\{j|j\neq i\}} F_i^T C_{ij} F_j = 0.1115$$

Now single obligor limits can be obtained:

$\theta_1 \leq 1.1478 - 0.3895 = 0.7583$; $f_1 \leq 0.7583 \times \$44{,}024 = \$33{,}384$

$\theta_2 \leq 0.5314 - 0.2860 = 0.2454$; $f_2 \leq 0.2454 \times \$43{,}186 = \$10{,}596$

$\theta_3 \leq 0.2492 - 0.1377 = 0.1115$; $f_3 \leq 0.1115 \times \$42{,}954 = \$4{,}790$

In summary, no loan in the first group exceeds its limit, while the $15,411 loan exceeds its limit in the second group. As was to be expected, the third group is the most problematic, since only the three smallest loans in the segment comply with the limit.

Note that although the third segment is the least numerically concentrated as measured by $H(F)$, it has the highest level of risk concentration H'. Although the first segment also exhibits high risk concentration, since it has the lowest average default probability it is the least risky of the three. Note also that the first is the numerically more concentrated segment, but since its equivalent correlation is relatively low, its risk concentration relative to its HHI is the smallest of the three. These numbers also illustrate the interplay between default probabilities and concentration in the loss variance of each segment, pointing to the third segment as the riskiest, because its equivalent correlation, risk concentration and average default probability are the largest of the three, providing the highest standard deviation of losses.

The example evidences the analytical power of the model. If one had restricted the exercise to using the general model without analysing individual segments, the risky third segment would have passed undetected. It is also clear that the results depend on the segmentation criterion used, since one can classify the loans in such a way that all segments comply with the relevant relations, and risky groups of loans will remain undetected. However, the example also indicates how one can obtain insight into the ex ante concentration issue, in the worst case by trial and error.

7. Accounting for recovery rates

It is simple to extend all the relations so far obtained to include loan recovery rates. Doing so leads to less restrictive limits in terms of tolerable concentration along the different dimensions where concentration can occur. Basically, there are two ways to account for recovery rates. The first is to define F directly as the vector of "*loss given default*" *(LGD), as opposed to the outstanding balance*, where it is assumed that nothing is recovered if loans default. This would be very much in line with current practice.[26] Thus if an estimation of the LGD vector is at hand, one can simply use this in the relations derived without any changes, although they should be reinterpreted accordingly.

Alternatively, assuming that the portfolio is segmented such that recovery rates are the same for all loans in the group, let r_i be the recovery rate for defaulted loans in segment i, so that the loss-given-default vector is simply $L_i = (1 - r_i)F_i$. Proceeding in the usual manner for each segment leads to:

$$H(F_i)R_i(M_i,F_i) + \frac{2}{V_i^2(1-r_i)}\sum_{\{j/j\neq i\}}(1-r_j)\,F_i^T\;C_{ij}\,F_j \leq \left(\frac{\psi_i-(1-r_i)\bar{p}_i}{z_\alpha\phi\,(1-r_i)}\right)^2 \qquad (7.1)$$

and adding over all segments:

$$\sum_i H\,(F_i)R_i(M_i,F_i) + 2\sum_i \frac{1}{V_i^2(1-r_i)}\sum_{\{j/j\neq i\}}(1-r_j)\,F_i^T\;C_{ij}\,F_j \leq \sum_i\left(\frac{\psi_i-(1-r_i)\bar{p}_i}{z_\alpha\phi\,(1-r_i)}\right)^2 \qquad (7.2)$$

The expression shows that any change in recovery rates has a double impact. On the one hand, the importance of each segment's correlation with loans of other segments is increased or decreased, depending on the ratio of loss rates between the loans in the segment with respect to that of the others. Additionally, its contribution to the expected loss also decreases (increases) in the numerator of the right-hand side, increasing (decreasing) the established bound on concentration. It is not difficult to show that the denominator of the right-hand side behaves accordingly, decreasing as the recovery rate increases and vice-versa. So, if recovery rate data is inadequate or non-existent, one can perform exercises using different recovery rates, or using some kind of reference.

26 See Basel Committee on Banking Supervision (1999).

8. The normality assumption

Up to this point, it has been assumed that the loss distribution is normal. In this section we discuss the approximation of the loss distribution using a gamma distribution, which can also be characterised by its mean and variance and captures the asymmetry typically observed in credit loss distributions. The gamma density function can be written as:[27]

$$f(x|\alpha,\beta) = \frac{x^{\alpha-1}}{\beta^{\alpha}\Gamma(\alpha)} e^{-\frac{x}{\beta}}$$

The mean and the variance are $E(x) = \alpha\beta$ and $VAR(x) = \alpha\beta^2$ respectively, and there is only one solution for any given pair of parameters (α, β).

Several exercises have been performed to date, to compare the results of the model presented here and CreditRisk$^+$, on random portfolios from the SENICREB database of the central bank.[28] Without claiming to have conducted a rigorous and exhaustive study, we can say that the results obtained are encouraging. In the next example the results for the best and worst fits are shown.

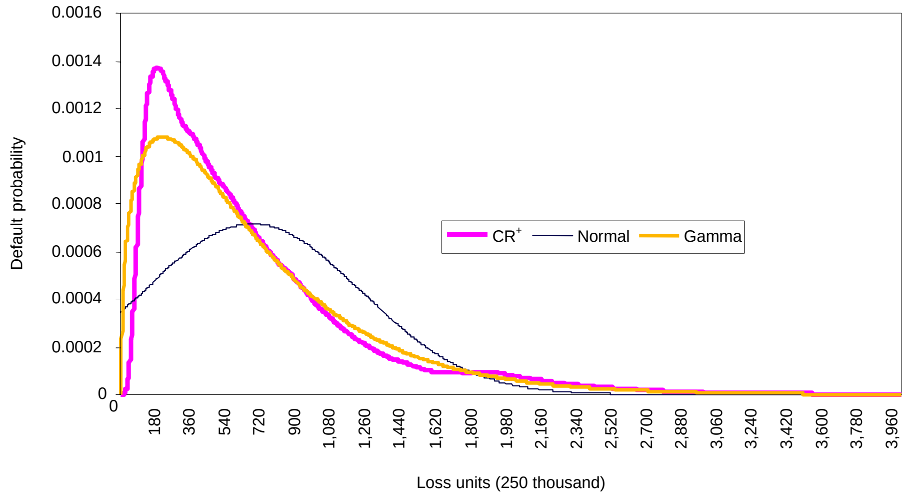
Example 8.1

The results for the first exercise, Figure 8.1, compare the loss distributions obtained for a random portfolio of 3,000 loans in the SENICREB database. Whereas the normal approximation can differ with CreditRisk$^+$ as much as 37.7% in VaR at the 99% confidence level, the difference using the gamma approximation is only 0.45%.

Figure 8.2 shows the results on a random portfolio of 1,320 loans from the same source. The loss distribution obtained using CreditRisk$^+$ has two “humps”. This is because this sample contained a very large loan in comparison with the other loans in the portfolio, which, due to the bucketing procedure required by CreditRisk$^+$, creates discontinuities and gaps in the possible losses. As shown in the table, the largest difference of VaR between the two methodologies using the gamma approximation is 12.34% at the 99% confidence level. The figures using the normal approximation are worse.

Figure 8.1

Comparison of loss distributions on a random sample of 3,000 loans



[27] There are many ways in which the gamma distribution can be written. The one adopted here follows the convention used in CreditRisk$^+$.

[28] SENICREB (Servicio Nacional de Información de Créditos Bancarios), is a loan database of the Mexican banking system, and is managed by the Bank of Mexico.

Table 8.1

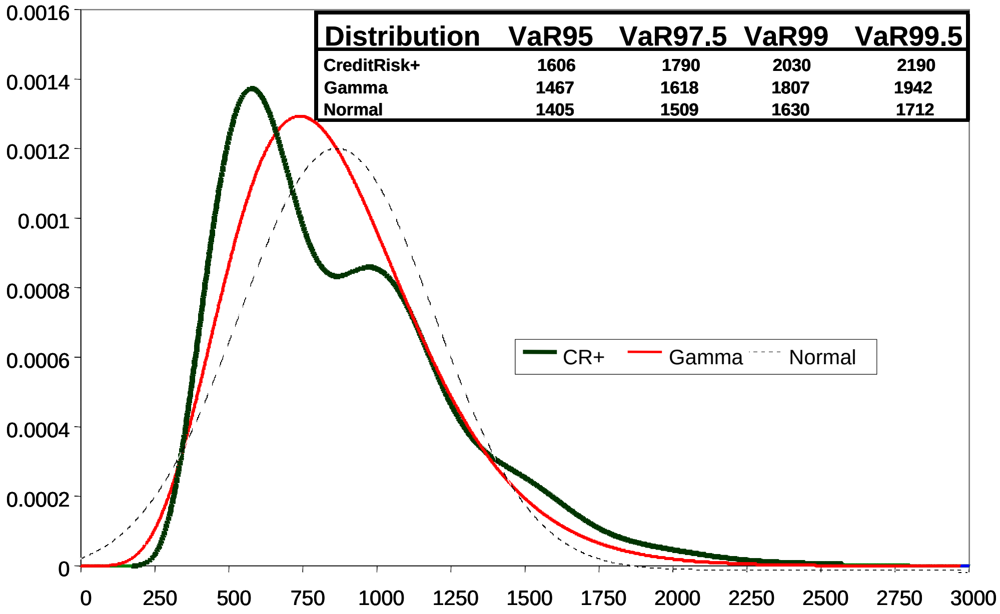
VaR comparative statistics for the sample

VaR confidence level	0.95	0.975	0.99	0.995
CR+	1,878	2,212	2,623	2,932
Normal	1,590	1,765	1,969	2,108
Gamma	1,770	2,120	2,577	2,919

Loss distribution statistics	Mean	Variance	Std dev	alfa	beta
CR+	673	312,277	559	–	–
Normal	674	310,116	557	–	–
Gamma	674	310,116	557	1.46	460.26

Figure 8.2

Comparison of loss distributions on a random sample of 1,320 loans



It should be pointed out that not all of the exercises produced VaR differences where the model underestimated the results of CreditRisk+. Some of the random portfolios provided results where the opposite occurred using the gamma approximation. In all of these cases the differences were small.

It is not always the case that the VaR obtained by CyRCE is less than the corresponding VaR obtained using CreditRisk+. Although the preceding examples are interesting and serve to illustrate the kind of results obtained by both methods, they are far from being a rigorous comparative study. In particular, it is interesting to examine how the two methodologies behave as the number of loans in the portfolio increases. In order to explore this behaviour, a simulation experiment was carried out, taking random samples of portfolios of increasing numbers of loans, and their VaR was calculated by both methods, for different confidence levels.[29] The results of the exercise are summarised in Figure 8.3. The number of loans in the portfolios is shown on the x-axis, while the y-axis shows the average of the following statistic:

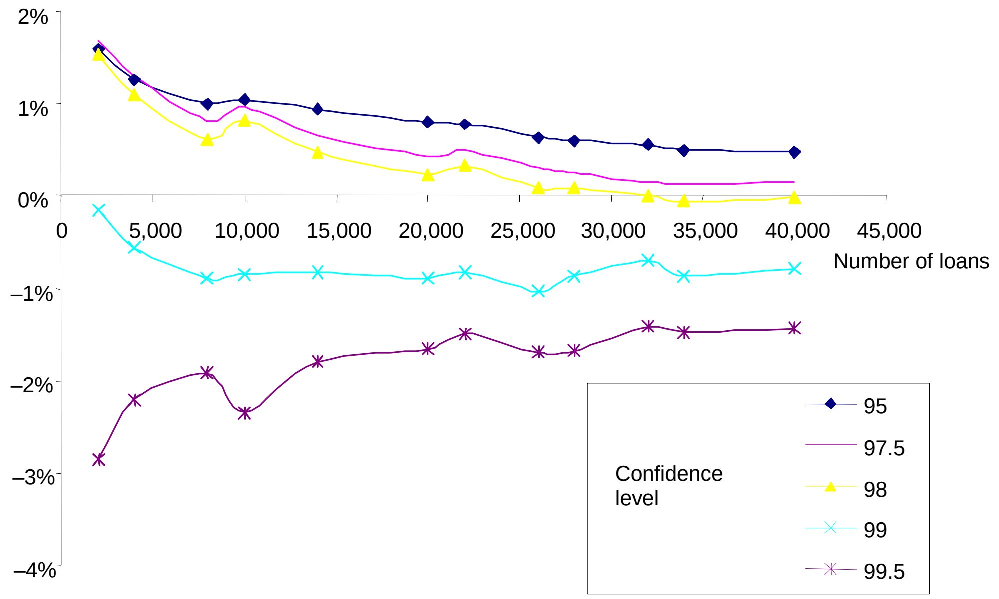
$$\Delta \frac{VaR_{CyRCE} - CaR_{CreditRisk^+}}{\text{Total value of loan portfolio}}$$

The curves in the graph represent the average differences in VaR relative to the value of the portfolio, for different confidence levels. The gamma distribution was used for approximating the loss distribution obtained by CyRCE.

Figure 8.3

Comparison between CreditRisk$^+$ and CyRCE as the number of loans increases in the portfolio, for different confidence levels

$$\frac{\text{VaR CyRCE – VaR CreditRisk}^+}{\text{Total loan}}$$



First, it is interesting to note that the average difference of VaRs relative to the size of the portfolio decreases as the number of loans increases. This provides some empirical evidence that there is some sort of large numbers effect. Next, the graph shows that, on average, the VaR obtained by CyRCE overestimates that obtained by CreditRisk$^+$ for confidence levels below 98%, and underestimates them for higher confidence levels. Undoubtedly, this is due to the heavier tails of the loss distribution generated by CreditRisk$^+$.

9. Application of the model with limited portfolio information

Any credit risk model requires two types of information: a description of the loans in the portfolio, and the default behaviour of the loans it contains (ie default probabilities and correlations). The model

[29] Thus, for each of the 23 sizes of portfolio, between 2,000 and 64,000 loans, 500 simulation runs were performed. Due to the characteristics of the SENICREB database, sampling was done with replacement for the larger sizes.

presented here allows several options for performing calculations with limited information. Regardless of the quality of information available on default rates of loans in a portfolio, it is the author's experience that in the worse case, bankers have some idea of what these are, even if this information is not available in some sort of systematised database. The estimation of default probabilities and correlations from default rates is a topic in itself and will not be dealt with here. On the other hand, the difficulty in obtaining portfolio information is of particular relevance to regulators, and probably constitutes the largest stumbling block for effective credit risk supervision. Banks are reluctant to provide regulators with this information on an ongoing basis simply because of the huge quantities of data involved. Even if the data could be obtained in an appropriate and systematic way, it would be difficult to handle. Private banks with large portfolios would also benefit from reducing information requirements to run their models. We now address this issue.

As seen in the derivation of the model, it is not strictly necessary to know the credit portfolios in detail. Given an adequate segmentation of the portfolio, the only information required by the model is:

a. The total value of the loans in each segment V_i.

b. Enough information about the loan distribution within each segment, which allows an estimate of its HHI.

c. Estimates of p_i, ρ_i and ρ_{ij}.

In what follows, we will discuss how estimates of HHI can be obtained from some very basic statistics. Thus, suppose that the portfolio has been segmented into h segments. If for each segment one knows the value of the segment, V_i, and the value of the largest loan in each segment, f_j^*, then Theorem 5.3 states that:

$$H(F_i) \leq \theta_i = \frac{f_i^*}{V_i} \tag{9.1}$$

Therefore, $H(F_i) = \theta_i$, is an estimate of HHI for each segment, although perhaps somewhat crude. In fact, Theorem 5.3 can be used to obtain a slightly tighter bound. To see this, remember that the largest concentration occurs when the portfolio has the following distribution as a proportion of its value V:

$$f_k = \begin{cases} \theta; & k = 1, 2, \ldots, n \\ \varepsilon; & k = n+1 \\ 0; & k = n+2, \ldots, N \end{cases}$$

$$\sum f_k = n\theta + \varepsilon$$

For this distribution,

$$H(F) = n\theta^2 + \varepsilon^2 = n\theta \cdot \theta + \varepsilon^2 = (1 - \varepsilon)\theta + \varepsilon^2.$$

This expression is minimum when $\varepsilon = 0.5\theta$. Since it is virtually impossible to have such a distribution in practice, if only the largest loan in each segment is known, one could argue that a good bound on HHI is:

$$H(F) < \theta(1 - 0.5\theta) \tag{9.2}$$

If the number of loans per segment N_i is known, as well as the average size loan $\overline{f_i}$ and the variance $\overline{\sigma}_i^2$, then HHI can be obtained. To see this, first note that $V_i = N_i \overline{f_i}$ is the value of each segment and $V = \sum_{i=1}^{h} V_i$ is the value of the portfolio. Then, by the definition of variance:

$$\overline{\sigma}_i^2 = \frac{\sum_k \left(f_k^i - \overline{f_i}\right)^2}{N_i - 1} = \frac{\left(N_i \overline{f_i}\right)^2}{(N_i - 1)\left(N_i \overline{f_i}\right)^2} \left\{ \sum_k \left(f_k^i\right)^2 - N_i \left(\overline{f_i}\right)^2 \right\}$$

$$= \frac{(V_i)^2}{(N_i - 1)} \left\{ H(F_i) - \frac{1}{N_i} \right\}$$

Solving for HHI one obtains:

$$H(F_i) = \frac{(N_i - 1)}{N_i^2}\left(\frac{\sigma_i}{\bar{f}_i}\right)^2 + \frac{1}{N_i} \tag{9.3}$$

Now, having estimates of HHI for each segment, one can obtain the HHI of the whole portfolio as follows:

$$H(F) = \sum_{i=1}^{h}\left(\frac{V_i}{V}\right)^2 H(F_i) \tag{9.4}$$

Note that (9.3) and (9.4) are *exact* values for the concentration indices, and that they can be obtained with very limited information.

10. Systemic credit risk analysis of the Mexican banking system

Using the SENICREB database, the model is currently being used to analyse the credit risk profile and capital adequacy of the 20 banks of the Mexican financial system. The results are presented to the board of governors of the central bank on a monthly basis. In the exercise presented here for illustrative purposes, we assume that the underlying loss distribution is normal and VaR computations are for a monthly horizon at the 2.5% confidence level. Due to the centralisation that characterises the Mexican economy, it is difficult to segment by geographical region. As a starting point, the loan portfolio of the system has been segmented by bank and economic activity. Loans rated by one or more of the major rating agencies form a separate segment, because there are relatively few rated loans so the observed history of their default behaviour is insufficient for default probability and correlation estimation. Fortunately, the rating agencies themselves provide good estimates of these parameters.

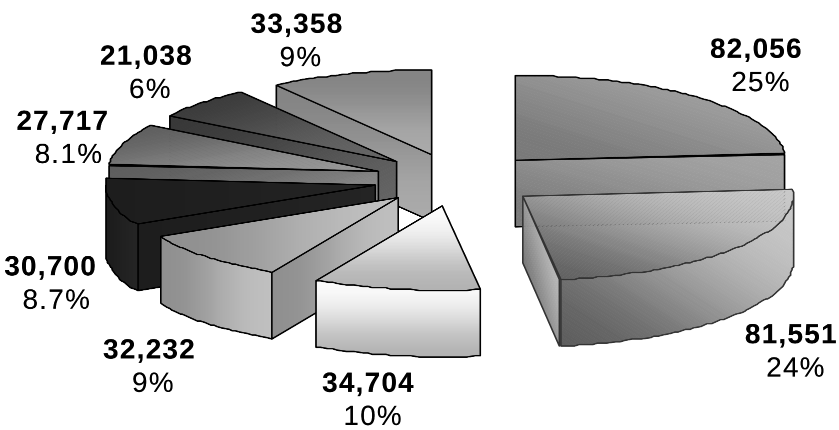
Figure 10.1 shows how the loan market was distributed among the major banks in Mexico at the end of March 2002. Two banks have 48% of the market, and 91% of all loans are held by seven banks.

Figure 10.1

Distribution of the loan portfolio by major banks

March 2002

Loan portfolio of the system = 343,357[1]



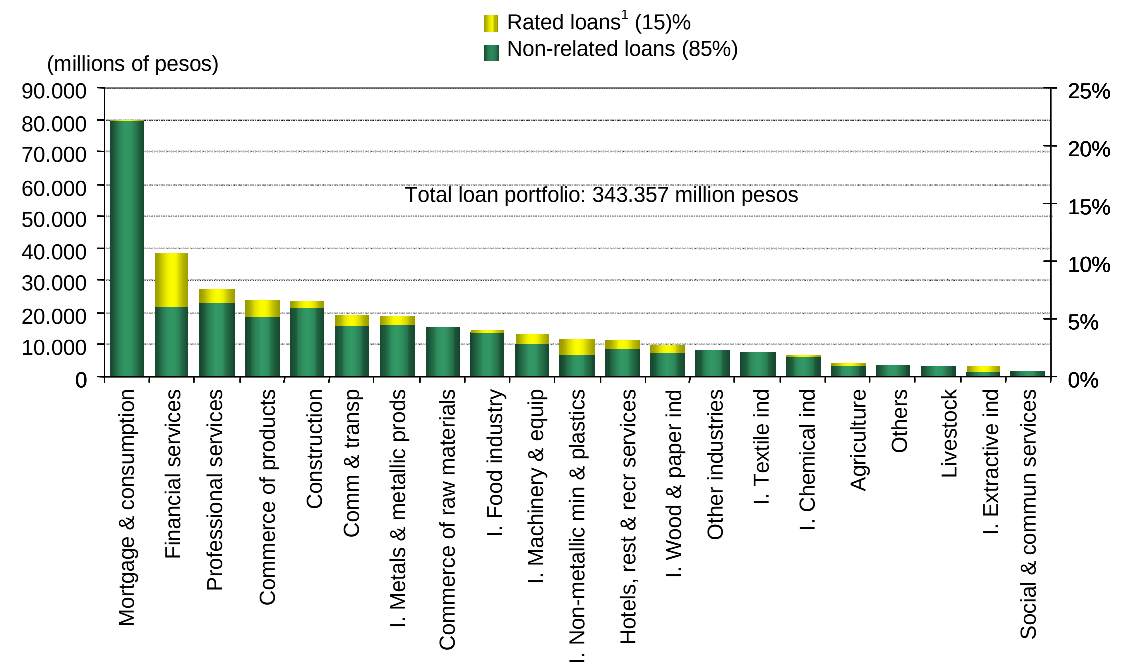
[1] Millions of pesos.

Figure 10.2 shows the distribution of the loan portfolio of the banking system segmented by economic activity. From the bar chart, we see that the largest segment is represented by mortgage and consumer loans, followed by financial services and so on.

Figure 10.2

Distribution of the loan portfolio by economic activity

March 2002

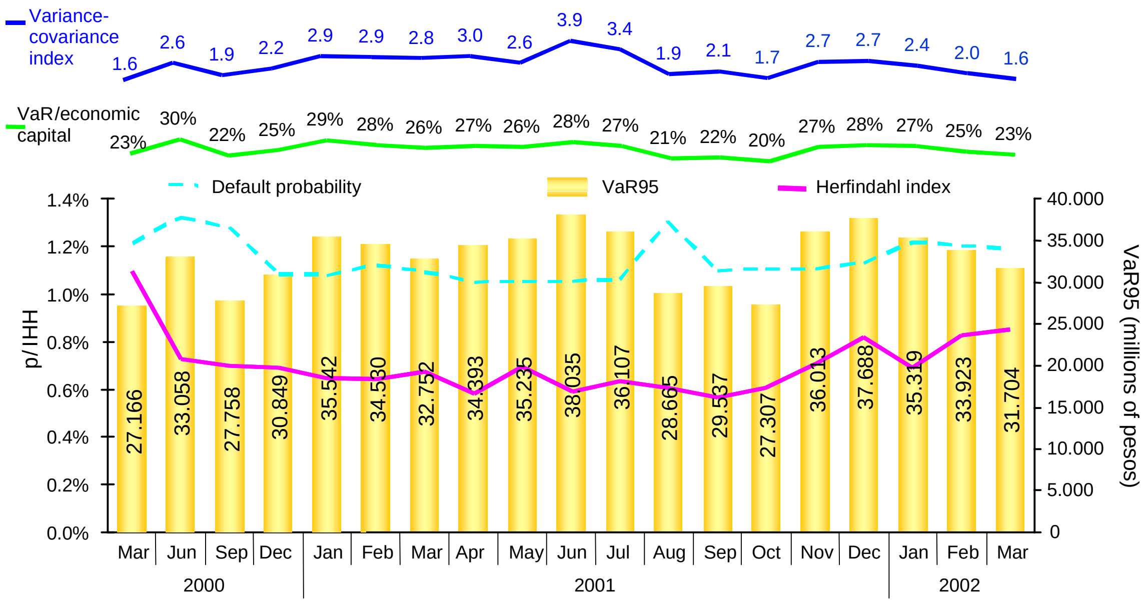


[1] Rated by Standard & Poor's, Moody's and Fitch.

Figure 10.3, shows the variation over time of VaR, default probabilities, HHI concentration index, covariation index and VaR relative to economic capital for the banking system.

Figure 10.3

Evolution of the risk profile of the Mexican banking system



p	1.2%	1.3%	1.3%	1.1%	1.1%	1.1%	1.1%	1.1%	1.1%	1.1%	1.1%	1.3%	1.1%	1.1%	1.1%	1.1%	1.2%	1.2%	1.2%
IHH	1.1%	0.7%	0.7%	0.7%	0.6%	0.6%	0.7%	0.6%	0.7%	0.6%	0.6%	0.6%	0.6%	0.6%	0.7%	0.8%	0.7%	0.8%	0.9%

This graph summarises how systemic VaR responds to the main risk drivers, and provides a first indication of how well the banking system is capitalised relative to the level of risk taken.

Figure 10.4 shows the concentration of risk in the system. Banks are sorted by their contribution to the overall VaR of the system, and arranged from largest to smallest. A Lorenz curve is constructed so that the amount of risk concentrated in a specific number of banks can be seen. The bars represent the ratio of each bank's VaR relative to its net capital. The horizontal line is the average VaR/net capital ratio for the system, so that one can see the relative position of each bank with respect to the system. Thus, 80% of the risk is concentrated in five banks and the third of these have a VaR/net capital ratio of 64%, which is relatively high when compared to the 23% average of the system.

Figure 10.4

Individual banks' contribution to systemic risk

March 2002

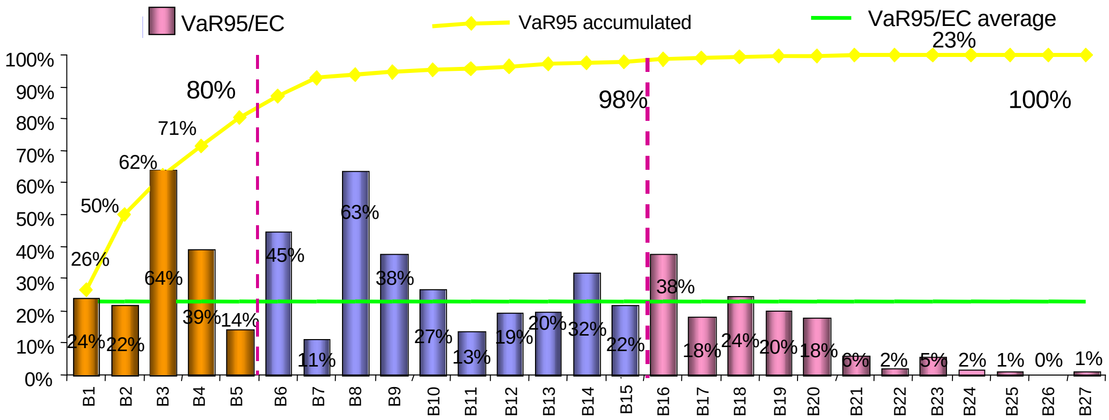


Figure 10.5

Contribution to systemic risk by economic activity

March 2002

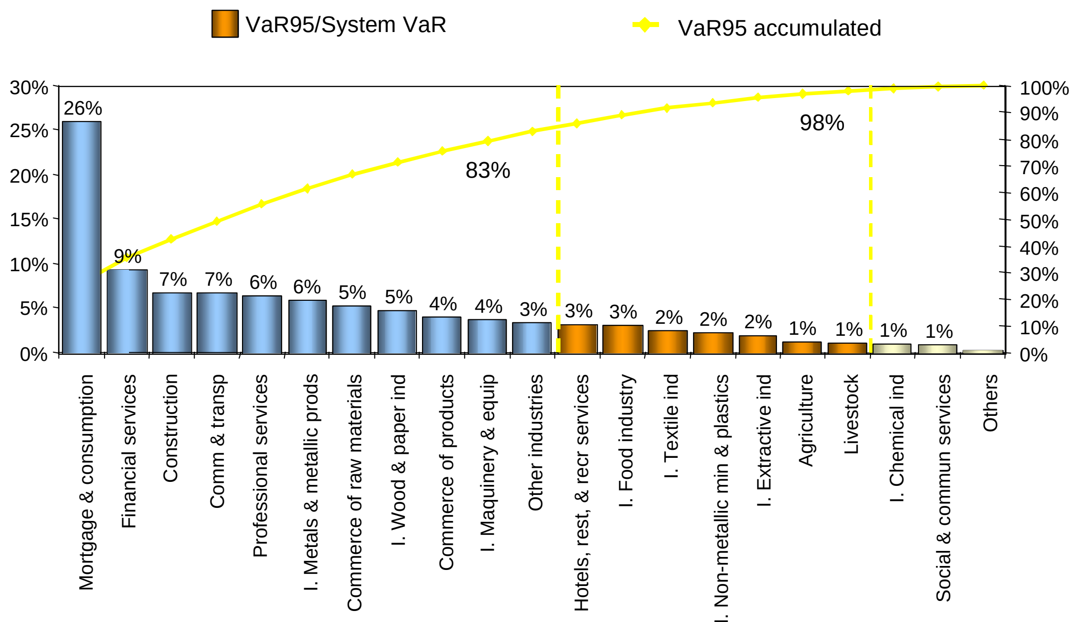


Figure 10.5, shows the contribution of the individual segments of economic activity. Most of the risk is in consumer and mortgage loans, followed by financial services. Note that construction and communications and transportation have moved up to third and fourth place respectively, from the fifth and sixth positions they occupy in terms of loan value. This is due to relatively high concentration and default probabilities within these sectors. Note also how the food industry, which is in the eighth position in terms of loan value, occupies the 13th slot in terms of risk.

Figures 10.6 to 10.8 provide some statistical results on the behaviour of the credit risk driver, ie default probabilities, concentration and loss variation-covariation indices, for all banks and segments considered.

Figure 10.6

Default probability histogram

April 2001-March 2002

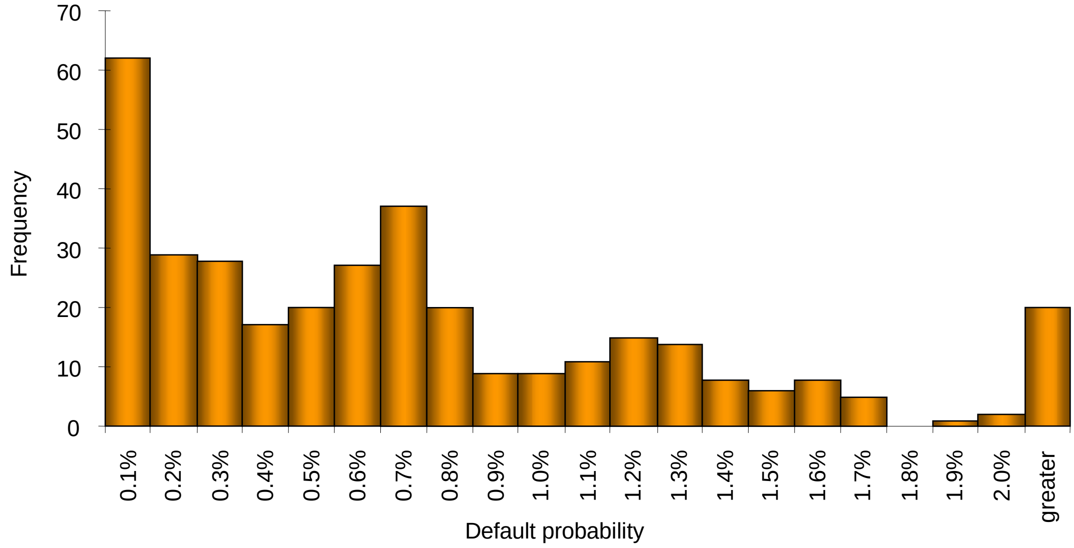


Figure 10.7

Herfindahl-Hirschman index histogram

April 2001-March 2002

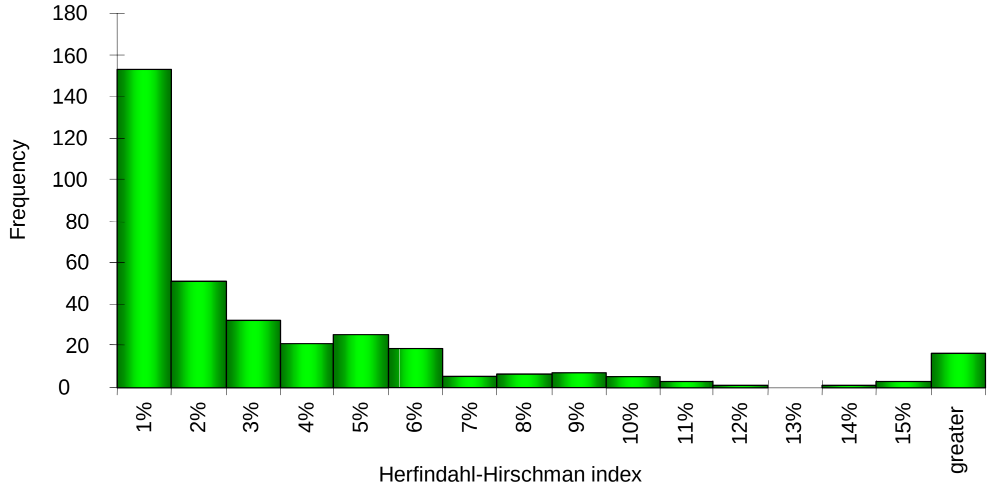


Figure 10.8

Loss variance-covariance index histogram

April 2001-March 2002

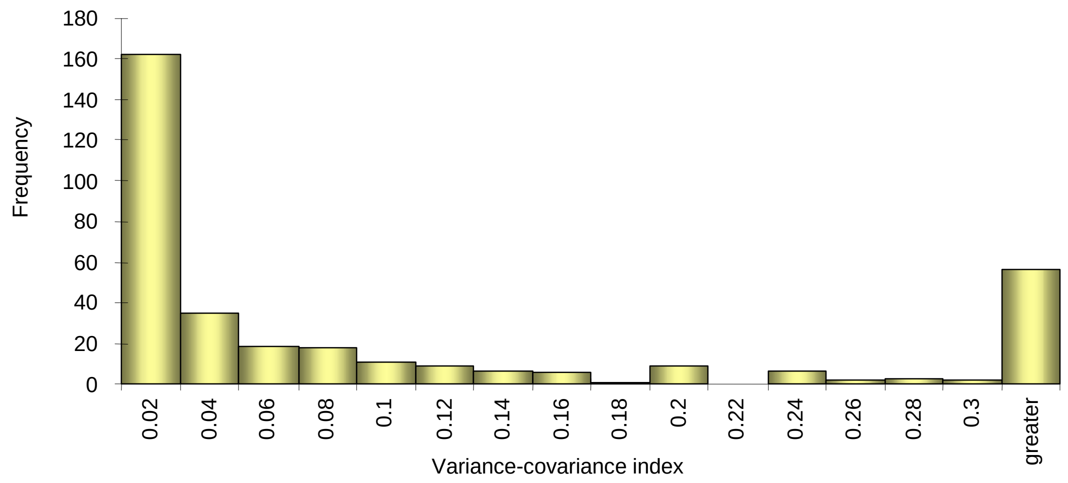


Figure 10.9 is a histogram of how well banks are capitalised as measured by the capital adequacy relationship. On average, the excess of net capital over VaR as a proportion of the value of the portfolios is around 25%, and in the past two years no negative quantities have been observed.

Figure 10.9

Capital adequacy histogram (EC – VaR95)/loan portfolio value > 0

April 2001-March 2002

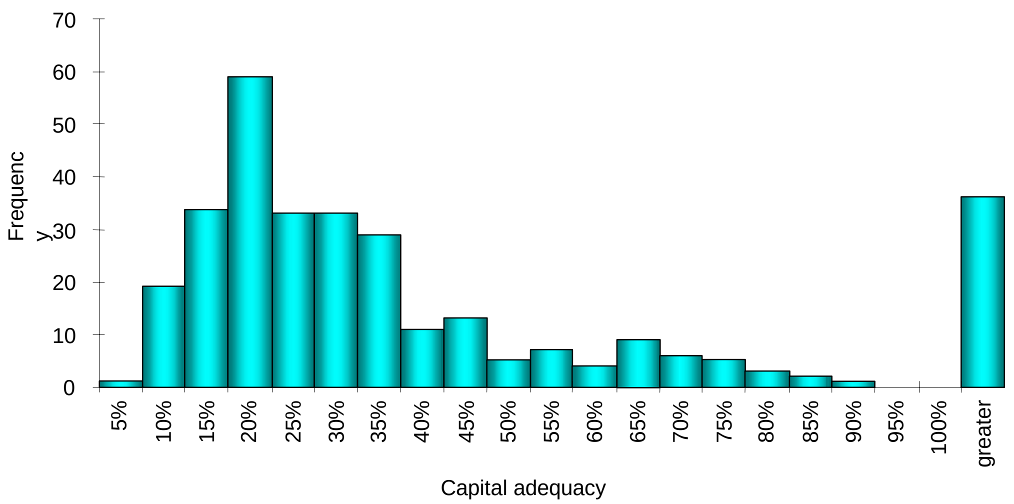
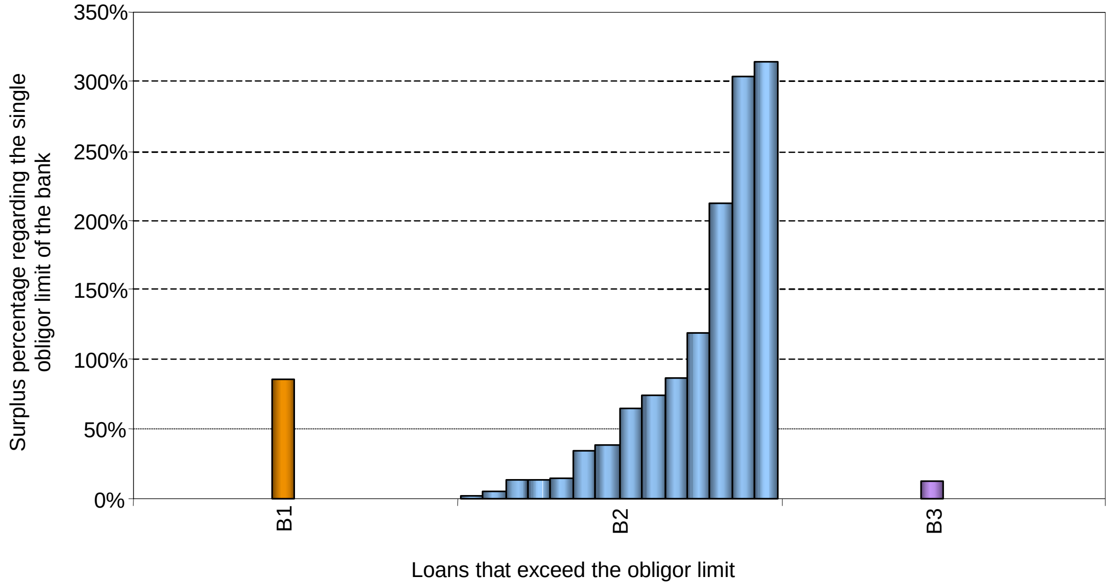


Figure 10.10 shows the analysis of compliance with the theoretical single obligor limits. Among the five banks that account for 80% of the risk of the system, only the three shown in the graph have loans that exceed the limit. Bank number two of this graph, which is the same as bank number three of Figure 10.5, has many such loans. Analysis of the bank's risk revealed that the large VaR/EC ratio is due to some extent to concentration. However, it should be emphasised that the limit is only a sufficient condition for capital adequacy but not a necessary one. So even if it helps to know how many and which loans are in violation of the condition, further analysis is needed in order to assess the gravity of the situation.

Figure 10.10

Loans that exceed single obligor limits in five major banks

March 2002



These graphs give a good idea of the type of analysis that is possible using the model. The same amount of detail is obtained for every bank in the system, permitting a more in-depth analysis of their situation.

11. Concluding remarks

The results obtained are very appealing for managing credit risk, since they provide explicit formulae to measure risk and permit a precise quantification of the policy actions that should be adopted in order to maintain capital adequacy. If default or recovery rates change, or concentration along a particular segment is excessive, the relations can be used to determine the adjustments to the capitalisation ratio and/or concentration composition of the portfolio that would re-establish capital adequacy. If banks have control over default or recovery rates to some degree, these can be part of the management instruments that can be used to maintain capital adequacy.

Since single obligor limits and the HHI are related to concentration, and since the measures are subject to the same bound, either one can be used as a policy instrument. In fact, both measures can be used in conjunction. Whereas the single obligor limit is easy to implement and supervise, it may lead to overly constrained loan distributions. For example, if a greedy bank manager decides to grant a loan exceeding his limit, the gravity of the transgression may be assessed using the HHI. It may be that, apart from misbehaviour, the infraction is not serious in terms of risk.

It is clear that default probability distributions as well as recovery rates exhibit random behaviours through time, depending on economic and financial factors. In contrast to market risk, where risk factors can be modelled using continuous processes, because loan defaults are discrete events in time, default behaviour in a certain group can also change in pronounced discrete jumps. This is one reason why it is difficult to establish ex ante concentration. Under different economic conditions, default probabilities and correlations can increase for a certain group of debtors, which otherwise appeared to be unrelated (eg mortgage loans to employees of a large company that goes bankrupt). Since the model can handle arbitrary segmentations of the portfolio, and it is relatively simple to stress particular segments and analyse the consequences, it provides a means for detecting ex ante concentration.

There is no reason why default probabilities cannot be made to depend on risk factors and credit drivers through logit models as in Credit Portfolio View, or as linear combinations of these factors.

Since these are determined exogenously, the results can be embedded in simulation models which generate scenarios of trajectories of these variables through time that exhibit the discontinuities typical of default-related events. From these, one can obtain stressed loss distributions and all the related statistics for each segment and the portfolio. This type of experiment may be what is needed to set an appropriate policy on the capitalisation ratio and single obligor limits. Note that since the simulation process in this case is for only a few variables, it can be done very efficiently.

Whatever the dynamics, it is always possible to make the necessary adjustments through time by monitoring only a few variables.

Appendix A:
Variance-covariance matrix

Example 6.1

Table A.1

M_1

	1	2	3	4	5	6	7	8
1	0.0162	0.0050	0.0050	0.0060	0.0060	0.0082	0.0082	0.0105
2	0.0050	0.0475	0.0086	0.0103	0.0103	0.0140	0.0140	0.0180
3	0.0050	0.0086	0.0475	0.0103	0.0103	0.0140	0.0140	0.0180
4	0.0060	0.0103	0.0103	0.0694	0.0125	0.0169	0.0169	0.0217
5	0.0060	0.0103	0.0103	0.0125	0.0694	0.0169	0.0169	0.0217
6	0.0082	0.0140	0.0140	0.0169	0.0169	0.1275	0.0230	0.0295
7	0.0082	0.0140	0.0140	0.0169	0.0169	0.0230	0.1275	0.0295
8	0.0105	0.0180	0.0180	0.0217	0.0217	0.0295	0.0295	0.2100

Table A.2

$C_{12}^T = C_{21}$

	1	2	3	4	5	6	7	8
9	0.0063	0.0108	0.0108	0.0130	0.0130	0.0177	0.0177	0.0227
10	0.0081	0.0138	0.0138	0.0166	0.0166	0.0226	0.0226	0.0290
11	0.0081	0.0138	0.0138	0.0166	0.0166	0.0226	0.0226	0.0290
12	0.0111	0.0190	0.0190	0.0229	0.0229	0.0311	0.0311	0.0399
13	0.0111	0.0190	0.0190	0.0229	0.0229	0.0311	0.0311	0.0399
14	0.0132	0.0226	0.0226	0.0273	0.0273	0.0370	0.0370	0.0475
15	0.0169	0.0290	0.0290	0.0350	0.0350	0.0475	0.0475	0.0609
16	0.0169	0.0290	0.0290	0.0350	0.0350	0.0475	0.0475	0.0609

Table A.3

$C_{13}^T = C_{31}$

	1	2	3	4	5	6	7	8
17	0.0039	0.0067	0.0067	0.0081	0.0081	0.0109	0.0109	0.0140
18	0.0052	0.0089	0.0089	0.0108	0.0108	0.0146	0.0146	0.0188
19	0.0067	0.0114	0.0114	0.0138	0.0138	0.0187	0.0187	0.0240
20	0.0081	0.0138	0.0138	0.0167	0.0167	0.0226	0.0226	0.0290
21	0.0092	0.0157	0.0157	0.0190	0.0190	0.0257	0.0257	0.0330
22	0.0109	0.0187	0.0187	0.0226	0.0226	0.0306	0.0306	0.0393
23	0.0140	0.0240	0.0240	0.0290	0.0290	0.0393	0.0393	0.0504
24	0.0140	0.0240	0.0240	0.0290	0.0290	0.0393	0.0393	0.0504
25	0.0140	0.0240	0.0240	0.0290	0.0290	0.0393	0.0393	0.0504

Table A.4

***M*$_2$**

	9	10	11	12	13	14	15	16
9	0.0291	0.0086	0.0086	0.0118	0.0118	0.0140	0.0180	0.0180
10	0.0086	0.0475	0.0109	0.0150	0.0150	0.0179	0.0230	0.0230
11	0.0086	0.0109	0.0475	0.0150	0.0150	0.0179	0.0230	0.0230
12	0.0118	0.0150	0.0150	0.0900	0.0207	0.0246	0.0316	0.0316
13	0.0118	0.0150	0.0150	0.0207	0.0900	0.0246	0.0316	0.0316
14	0.0140	0.0179	0.0179	0.0246	0.0246	0.1275	0.0376	0.0376
15	0.0180	0.0230	0.0230	0.0316	0.0316	0.0376	0.2100	0.0483
16	0.0180	0.0230	0.0230	0.0316	0.0316	0.0376	0.0483	0.2100

Table A.5

$\boldsymbol{C}_{32}^{T} = \boldsymbol{C}_{23}$

	17	18	19	20	21	22	23	24	25
9	0.0070	0.0093	0.0119	0.0144	0.0164	0.0195	0.0250	0.0250	0.0250
10	0.0089	0.0119	0.0152	0.0184	0.0209	0.0249	0.0320	0.0320	0.0320
11	0.0089	0.0119	0.0152	0.0184	0.0209	0.0249	0.0320	0.0320	0.0320
12	0.0122	0.0164	0.0209	0.0253	0.0288	0.0343	0.0440	0.0440	0.0440
13	0.0122	0.0164	0.0209	0.0253	0.0288	0.0343	0.0440	0.0440	0.0440
14	0.0146	0.0195	0.0249	0.0301	0.0343	0.0408	0.0524	0.0524	0.0524
15	0.0187	0.0250	0.0320	0.0386	0.0440	0.0524	0.0672	0.0672	0.0672
16	0.0187	0.0250	0.0320	0.0386	0.0440	0.0524	0.0672	0.0672	0.0672

Table A.6

***M*$_3$**

	17	18	19	20	21	22	23	24	25
17	0.0162	0.0093	0.0119	0.0144	0.0164	0.0196	0.0251	0.0251	0.0251
18	0.0093	0.0291	0.0160	0.0193	0.0220	0.0262	0.0336	0.0336	0.0336
19	0.0119	0.0160	0.0475	0.0247	0.0281	0.0335	0.0429	0.0429	0.0429
20	0.0144	0.0193	0.0247	0.0694	0.0340	0.0404	0.0519	0.0519	0.0519
21	0.0164	0.0220	0.0281	0.0340	0.0900	0.0461	0.0591	0.0591	0.0591
22	0.0196	0.0262	0.0335	0.0404	0.0461	0.1275	0.0704	0.0704	0.0704
23	0.0251	0.0336	0.0429	0.0519	0.0591	0.0704	0.2100	0.0903	0.0903
24	0.0251	0.0336	0.0429	0.0519	0.0591	0.0704	0.0903	0.2100	0.0903
25	0.0251	0.0336	0.0429	0.0519	0.0591	0.0704	0.0903	0.0903	0.2100

References

Adelman, M A (1969): "Comment on the 'H' concentration measure as a numbers-equivalent", *Review of Economics and Statistics*, February, pp 99-101.

Altman, E I and A Saunders (1998): "Credit risk measurement: development over the last 20 years", *Journal of Banking and Finance*, vol 21, pp 1721-42.

Basel Committee on Banking Supervision (1999): *Credit risk modelling: current practices and applications*.

Bennet, P (1984): "Applying portfolio theory to global bank lending", *Journal of Banking and Finance*.

Caouette, J B, E I Altman and P Narayanan (1998): *Managing credit risk: the next great challenge*, John Wiley & Sons, Inc.

Coopers and Lybrand (1993): "Growth and diversification: are the benefits always there?", in *Mortgage banking*, J Lederman (ed), Probus Publishing Company, Chicago.

Cowell, F A (1995): "Measuring inequality", *LSE Handbooks in Economics*, 2nd edition, Prentice Hall/Harvester Wheatsheaf.

Credit Risķ (1996): Credit Suisse Financial Products, New York, NY.

CreditMetricsTM (1997): technical document, 2 April.

Crouhy, M, D Galai and R Mark (2000): "A comparative analysis of current credit risk models", *Journal of Banking and Finance*, vol 24, pp 59-117.

DeGroot, M H (1988): *Probabilidad y estadística*, Addison-Wesley Iberoamericana, Wilmington Delaware.

Encaoua, D and A Jacquemin (1980): "Degree of monopoly, indices of concentration and threat of entry", *International Economic Review*, 21.

Finger, C C (1998): *Sticks and stones*, The RiskMetrics Group.

Gordy, M B (2000): "A comparative anatomy of credit risk models", *Journal of Banking and Finance*, vol 24, pp 119-49.

Kealhofer, S (1998): *Portfolio management of default risk*, KMV Corporation, revised.

——— (1999): *Modeling default risk*, KMV Corporation, revised.

Kelly Jr, W A (1981): "A generalized interpretation of the Herfindahl index", *Southern Economic Journal*, 48(1), pp 50-7.

Márquez Diez-Canedo, J (2002): "Capital adequacy and credit risk in bank loan portfolios: a default model that relates capital adequacy to credit risk, measures concentration, and obtains single obligor limits", *Economia, Societa' e Istituzioni*, Anno XIV/n° 1 Gennaio, Aprile.

Markowitz, H (1959): *Portfolio selection: efficient diversification of investments*, John Wiley & Sons, New York.

Merton, R C (1974): "On the pricing of corporate debt: the risk structure of interest rates", *Journal of Finance*, 29, pp 449-70.

Mirsky, L (1990): *An introduction to linear algebra*, Dover Publications, Inc.

Moody's Investor Services (1991): *Rating cash flow transactions backed by corporate debt*. Moody's Special Report, April.

Shy, O (1995): *Industrial organization*: *theory and applications*, MIT Press.

Sleuwaegen, L E, R R De Bondt and W V Dehandschuter (1989): "The Herfindahl index and concentration ratios revisited", *The Antitrust Bulletin*, Fall.

Strang, G (1988): *Linear algebra and its applications*, Academic Press, New York.

Tirole, J (1995): *The theory of industrial organization*, MIT Press.

Weinstock, D S (1984): "Some little-known properties of the Herfindahl-Hirschman index: problems of translation and specification", *The Antitrust Bulletin*, Winter.

Wilson, T C (1997a): "Portfolio credit risk (I)", *Risk*, 10(9), September.

——— (1997b): "Portfolio credit risk (II)", *Risk*, 10(10), October.

Lending decisions, procyclicality and the New Basel Capital Accord

Fabrizio Fabi, Sebastiano Laviola and Paolo Marullo Reedtz[1]
Bank of Italy

1. Introduction

At the end of April 2003 the Basel Committee on Banking Supervision released a third consultative paper (CP3) containing a proposal for a new accord on bank capital (Basel II). The proposal contains important changes with respect to an earlier paper, published in January 2001 (CP2). The reform process has been undertaken in response to the increase of financial innovations in banking products and enhancements in the measurement of banking risks, which have highlighted some inadequacies in the simplified framework underlying the 1988 Accord (the "current" Accord). Indeed, the current Accord does not fully reflect changes in risk. As a consequence, it may understate the risks and hence overstate the capital adequacy of banks. It may also create incentives for banks to make high-risk investments.

A more differentiated assessment of banks' risk exposures and the provision of incentives to banks to improve their risk measurement and management capabilities are the key objectives of the new proposal. With regard to the level of overall capital, the Basel Committee has explicitly declared that in the standardised approach minimum capital requirements have to bring about a level of capital that is on average equal to the current requirement (8%), while banks applying the more advanced approaches should receive on average a small capital incentive.

As is well known, the proposal is based on three pillars - minimum capital requirements, supervisory review of banks' capital adequacy and, market discipline - and foresees a plurality of methods to calculate capital requirements, according to the degree of development of banks' risk management systems.

Through the consultation with the banking industry and three impact studies performed by a large number of intermediaries, the Basel Committee has aimed at aligning prudential regulation with the best practices of risk management developed in the marketplace.

Some important changes have been introduced in CP3. The most significant improvements are in the field of defining the capital requirements connected with the corporate and retail portfolios. The rise in capital requirements with the increase in the borrowers' probability of default was deemed to be too sharp in the proposal issued in January 2001. This would have implied a serious impact on the financing of small and medium-sized enterprises (SMEs), which tend to have relatively higher probabilities of default than large corporates. In order to comply with higher capital requirements, banks would have been induced to increase the interest rates charged to high-risk borrowers or to cut the amount of lending.

Moreover, such a conservative calibration of overall capital was likely to lead to a potential increase in the procyclicality of the supply of credit: in times of recession, when the quality of borrowers tended to deteriorate, banks would reduce lending (and therefore risk-weighted assets) in order to comply with the increase in capital requirements.

Capital requirements that change according to the riskiness of bank borrowers are a built-in effect of any risk-sensitive prudential regulation. What is really relevant is that, even under the current Accord, in which essentially all corporate and retail loans are subject to the same capital charge, lending to borrowers with a different financial situation is priced at different interest rates and risk premia are

---

[1] Banking and Financial Supervision departments. The views expressed in this paper are those of the authors and do not necessarily reflect those of the Bank of Italy. We thank M Kwast (US Federal Reserve Board) for helpful comments and suggestions. Please address correspondence to: Banca d'Italia, Servizio Concorrenza, normativa e affari generali, Via Milano 53, 00184 ROMA, tel: +39 06 4792 4458-4601-4455; fax: +39 0647924460; e-mail: Fabi.Fabrizio@insedia.interbusiness.it; Laviola.Sebastiano@insedia.interbusiness.it; Marulloreedtz.Paolo@insedia.interbusiness.it.

usually negatively correlated with the rate of growth of GDP. Such circumstances can also easily be recorded for the period before 1988, when no capital regulation was in force at international level. This implies that the new regulatory proposal could be blamed for altering the lending policies of banks only in the event that the assessment of credit risk implicit in the risk-weight functions substantially differed from banks' perception of risk as reflected in the interest rates they charge to the borrowers.

The aim of this paper is to provide an empirical evaluation of the impact of the CP3 proposals on the lending policies of Italian banks, ie on interest rates on bank loans. We address this issue through two separate steps: first, we compare the interest rates charged to a large set of Italian firms with the cost brought about by the change in the calculation of capital requirements, so as to have an assessment of the impact of the new regulatory scheme on banks' lending policies; and second, we measure the change in interest rates which would be consistent with a sudden deterioration in the cyclical conditions of the corporate sector under the new regulatory scheme, in order to have an indication of the procyclicality effect embodied in the New Accord.

The rest of the paper is organised as follows. In Section 2 we briefly review the main aspects of credit risk measurement under the new capital adequacy framework which are relevant for the empirical exercises conducted later on; in Section 3 we look at the impact of the proposed capital requirements on banks' loan rates to a large sample of non-financial firms. In Section 4 we conduct a "stress testing" exercise, in order to assess the procyclicality of capital requirements on lending conditions in a negative economic scenario. Section 5 draws some conclusions.

2. CP3: credit risk measurement and the IRB approach

As regards Pillar 1 of Basel II, the purpose of creating a more risk-sensitive framework is pursued through a range of options for addressing credit risk, including: (a) a standardised approach, under which risk weights are based on the evaluation of credit quality by external credit assessment institutions (rating agencies and other institutions authorised according to a set of specified criteria); (b) a "foundation" internal ratings-based (IRB) approach, based on both banks' internal assessments of risk components and supervisory parameters; and (c) an "advanced" IRB approach, in which all risk components are estimated internally by banks.

Both IRB approaches to computing risk-weighted assets rely on four quantitative risk factors: (1) the probability of default (PD), which measures the likelihood that the borrower will default over a given time horizon; (2) the loss-given-default (LGD), which measures the proportion of the exposure that will be lost if default occurs; (3) the exposure at default (EAD), which includes the on-balance sheet exposure and an estimate of the off-balance sheet one (as an example, for loan commitments the purpose is to measure the amount of the facility that is likely to be drawn if a default occurs); and (4) the maturity (M) of the exposure, which measures the remaining economic maturity of the asset. For corporate, sovereign and interbank exposures, under the "foundation" IRB approach banks satisfying minimum supervisory requirements will be allowed to input their own assessment of the probability of default associated with the borrower. The value of the other risk factors, such as EAD, LGD and maturity, will be determined by supervisors. Under the advanced IRB approach banks will provide internal estimates of LGD, EAD and M as well as PD.

For each of the relevant portfolios, a risk-weight function translates the risk components into specific capital requirements.[2] In the CP2 document only one risk-weight function was established for bank exposures to the corporate sector. The formula proposed delivered an 8% capital requirement for a benchmark unsecured loan having a 0.7% PD, a 50% LGD and a three-year maturity.

The comments from financial institutions, other market participants and national authorities, as well as the results of an in-depth Quantitative Impact Study on a sample of international banks, and pointed out that minimum capital charges tended to exceed current ones under the revised standardised approach; in turn, the standardised approach requirements were lower than those computed

[2] For further details see BCBS (2001a, 2003a).

according to the foundation IRB approach.[3] This was not consistent with the declared objectives of the Committee.

The steepness of the risk-weight curve in the IRB approach for the corporate portfolio was mentioned among the factors responsible for such a result. Many comments focused on the impact of the proposed regulatory framework on the potential procyclicality effects of the new regulatory scheme and on the financing of SMEs, as well as on the treatment of expected losses.[4]

With reference to the expected loss (EL) treatment, it was argued that it did not recognise the specific provisions made on loans to offset the capital requirements, or the general provisions not included in supplementary capital. This would not encourage adequate provisioning policies and could create competitive disadvantages for banks subject to more rigorous prudential standards.

As regards the procyclicality issue, the influence of capital regulation on the potential propensity of the banking system to increase macroeconomic fluctuations is a theme often addressed in the economic literature, but it has rarely been possible to come to clear-cut conclusions. While it is widely accepted that the banking system is inherently procyclical, it has not been possible to establish a clear link between binding capital requirements and macroeconomic outcomes.[5] However, with the new regulation a potential fluctuation of capital requirements over the business cycle is to a certain extent an inevitable result of the higher risk sensitivity.[6] Since the publication of CP2 the issue of procyclicality has therefore stimulated a great debate in the literature; many papers have recently addressed the link between credit risk measurement and procyclicality of the financial system, from both a theoretical and an empirical point of view.[7]

The main cyclical element in credit risk measurement comes from rating migration; both internal and external credit ratings improve during phases of economic expansion and deteriorate during contraction, so that measured risk falls in good times and increases in bad times. Therefore, the level of capital required by the new proposal will probably rise in economic downturns and fall in expansionary phases. The changes can be more pronounced to the extent that rating systems rely on market-based information (for example KMV) as opposed to relying on the methods employed by credit rating agencies (through-the-cycle ratings). Banks use a variety of rating systems; some are similar to the approach followed by KMV or to that of rating agencies. Many banks, however, use systems that are in-between, whereby the PD is derived from internal models or from expert judgment systems relying heavily on the experience of credit officers. In the latter case, it is not clear how much the raters take into account the future evolution of the state of the economy.

On the other hand, the use of more accurate rating systems is likely to bring about improvements in risk management practices; therefore, deteriorations in credit quality should be detected earlier than before, and banks could take the appropriate measures. Moreover, even though the regulation does not require the rating of borrowers through the cycle, it encourages banks to take greater account of uncertainty in economic conditions. In the longer term, banks could choose to run their internal rating processes in a way that incorporates greater provision for unexpected events.

In sum, even with the existing capital standards there is a definite cyclical element in the banking system. To the extent that Basel II encourages banks to be more forward-looking, this could reduce procyclicality because such behaviour would cause banks to react more quickly to changing conditions and expectations.

As far as loans to SMEs are concerned, it was argued that small firms usually have a higher PD but are relatively less sensitive to the evolution of the macroeconomic framework, while large enterprises

---

3 See BCBS (2001b).

4 See, among others, ABI (2001), IIF (2001) and ISDA (2001).

5 See, for example, Jackson et al (1999).

6 In the current regulation there can be a lower contribution of earnings to capital as a consequence of the greater losses during a downturn; with the new proposal there would also be a fluctuation in the risk-weighted assets, given the migration of borrowers to higher risk classes.

7 See, among others, BCBS (2001c), Borio et al (2001), Danielsson et al (2001), ECB (2001), Ervin and Wilde (2001), Estrella (2001), Allen and Saunders (2002), Ayuso et al (2002), Jordan et al (2002), Lowe (2002), Resti (2002), Segoviano and Lowe (2002) and Catarineu-Rabell et al (2003).

tend to behave in the opposite way. In other words, small firms' loans tend to be riskier because of the firms' own specific characteristics; this means the effect of systematic risk on their financial conditions is proportionately lower. In the simplified (with respect to fully fledged state-of-the-art credit risk models) framework for the determination of IRB capital charges, the effect of systematic risk is basically taken into account by the value of the average asset correlation across obligors, which is established by the regulators. Therefore, for a given PD of individual borrowers, a portfolio of loans to SMEs is less risky than a single loan to a large firm, because the asset correlation is lower, all else equal.[8]

On the basis of the comments received and of further empirical evidence, the treatment of exposures to corporates, and to SMEs in particular, has been considerably improved.

In the first place, the steepness of the risk-weight curve has been lowered for all corporate exposures by shifting the threshold for neutrality vis-à-vis the 1988 Accord to a 1.0% PD.[9] The reduction in risk weights is much stronger for higher PD levels. Therefore, this modification has made it possible to significantly reduce the potential degree of procyclicality of the framework and to indirectly take into account, at least partially, the SME issue.

Further, in the IRB approach included in CP3 banks are permitted, separately for any asset class (corporate, retail, interbank, etc), to recognise provisions to offset the EL component of the capital charge.[10] This modification, in addition to providing incentives to banks to make adequate provisions, also makes it possible to reduce the procyclicality of the regulation; specifically, in a downturn the possibility to offset the EL charge with provisions reduces the increase in the requirement connected with the migration of loans towards lower-quality risk buckets.

Finally, the Basel Committee has established that in the IRB approach banks would have to conduct reasonably conservative stress tests of their own design, with the aim of estimating the extent to which their IRB requirements could increase during a stress scenario. The results of these stress tests would be used by supervisors in order to ensure that banks were holding a sufficient capital buffer under Pillar 2 of the New Accord.

With reference to the treatment of SME loans, the smaller size of firms has been recognised as a factor potentially allowing banks to reduce capital requirements on loans to non-financial firms, other things being equal. Specifically, banks will be permitted to distinguish between exposures to large firms and those to SMEs, defined as corporate exposures where the reported sales for the consolidated group of which the firm is a part are lower than €50 million. Loans to SMEs will attract a capital requirement, for a given PD, LGD and maturity, lower than that attracted by larger firms. The capital reduction increases linearly from 0% to 20% with sales going from €50 to €5 million, and remains at 20% for firms with sales figures lower than the latter threshold.

Moreover, loans extended to small businesses can be treated according to the risk-weight formula established for the retail portfolio provided that: (a) each of them represents a small portion of a large pool of loans with similar risk characteristics which are managed by the bank on a pooled basis; (b) the total exposure of the banking group to an individual small business is less than €1 million. In this case, the capital requirements are lower than those for SMEs classified in the corporate portfolio, all else equal.

A third Quantitative Impact Study (QIS3) was performed between October and December 2002 with the cooperation of 365 banks from 43 countries. A total of 188 G10 banks were divided into two groups: Group 1 is representative of the large and internationally active banks; Group 2 includes smaller and, in many cases, more specialised banks.

---

8 For empirical studies on the asset correlations of portfolios of large and small corporates see Cannata et al (2001), Lopez (2002), Dietsch and Petey (2003) and Dullmann and Scheule (2003).

9 In the context of the calibration of the capital charges contained in CP3, in the IRB foundation approach the LGD of a senior unsecured loan has been reduced from 50% to 45% of the nominal exposure, and the residual maturity of the asset, originally set at three years, has been lowered to 2.5 years.

10 Provisions exceeding those already included in supplementary capital (exceeding 1.25% of risk-weighted assets) can continue to be used as one-for-one offsets to capital requirements on performing loans, but only to the extent that the EL portion of the IRB capital requirement also exceeds the maximum amount of provisions eligible for inclusion in Tier 2.

Although banks did not succeed in completely simulating the provisions of the new regulation, as regards, for example, the range of collateral instruments allowed or a more clear-cut separation across portfolios because of shortcomings in their information systems, on the whole the results are consistent with the Committee's objectives: (1) minimum capital requirements would be broadly unchanged for Group 1 banks, which are likely to use IRB approaches; (2) for Group 2 banks capital requirements under the IRB approach could be substantially lower than under the current Accord, due to the relatively larger size of retail portfolios.

Within the IRB approach, capital requirements for the financing of the corporate sector are lower than under the current Accord for both Group 1 and Group 2 banks, in connection with the higher quality of the borrowers. As expected, capital requirements both on loans to SMEs treated as corporate and on loans to small businesses included in the retail portfolio also turn out lower than currently. Overall, capital requirements for credit risk show a sharp reduction in comparison with the current Accord especially for Group 2 banks, thanks above all to the better treatment of retail portfolios. However, the overall result is substantially affected by the operational risk requirement.[11]

3. The impact of the new capital requirements on the pricing of bank loans

The results of QIS3 have confirmed the improvements that have been made in the proposed regulation.

However, this is not enough to be able to argue that banks' lending policies will not be distorted by the new regulatory framework. In the current situation, given the dispersion of interest rates by economic sectors and geographical areas (Graph 1), the pricing of individual bank loans is unlikely to be affected by the flat capital requirement.

In general terms, the pricing of bank loans reflects both financial and operating costs, the market power of the bank, and a risk premium computed by the bank according to its internal procedures, in some cases through VaR methodologies. In the context of the IRB approach, internal ratings and default and loss estimates must play an essential role in the credit approval, risk management and internal capital allocation functions of banks using this approach. This implies that a potential change in lending policies could be introduced if the regulatory treatment of credit risk were inconsistent with the internal assessments of banks, as reflected in the pricing of their lending operations.

We define the overall risk component, ORC, (or the "total regulatory cost of risk") of any lending operation, measured as a percentage of the nominal exposure, as: $ORC = EL/EAD + k(REQ\text{-}EL)/EAD$, where EL is the expected loss, REQ is the capital requirement as measured in CP3, k is the rate of return on bank capital and EAD is the nominal exposure. Since the CP3 capital requirement includes both EL and UL (unexpected loss), the formula is equivalent to: $ORC = EL/EAD + k{*}UL/EAD$.

In order to measure the capital requirement connected with each lending operation we need estimates of all relevant risk parameters. In the context of the IRB foundation approach, we derived an estimate of the probability of default of each borrower, while we used the supervisory parameters for the loss-given-default (ie, we considered all the exposures as uncollateralised) and maturity. As to EAD, we considered only the on-balance sheet nominal amount.

We refer to the Italian framework, for which a large amount of data on both lending relationships and balance sheets of industrial and commercial firms is available. Quantitative information can be drawn from CERVED's Company Accounts Register and from the Credit Register run by the Bank of Italy. In the Company Accounts Register both the balance sheets and the profit and loss accounts of a large set of Italian firms have been collected since 1993 according to a simplified reclassification scheme including 70 elementary items. The Credit Register records individual credit positions above approximately €75,000; non-performing loans are recorded no matter what their amount. The interest rates charged to individual borrowers by individual banks are also available, with reference to a sample of 68 banks accounting for 80% of total loans.

---

[11] For further details see BCBS (2003b,c).

3.1 Measuring probabilities of default and capital requirements

The probabilities of default of a large sample of corporate borrowers are estimated on the basis of a scoring model developed for research purposes at the Bank of Italy. A logit model is used in order to distinguish sound from insolvent firms.[12] In particular, balance sheet data at time t and Credit Register information at time $t+1$ are used to assess the probability for each firm of being recorded as defaulted at time $t+2$. A firm is regarded as defaulted if it is reported in the Credit Register's bad debt category for the first time in the year $t+2$ by at least one lending bank.

In the Credit Register bad debts are defined as all exposures to insolvent borrowers, regardless of any collateral received. Debtors are considered insolvent if they are globally unable to cover their financial obligations and are not expected to recover, even if this does not result in a legal bankruptcy procedure.[13]

The estimation procedure was applied to a set of 180,000 firms divided into four sectors of economic activity (manufacturing, trade, construction and services); a separate regression model was estimated for each sector. Through a stepwise procedure, 11 significant explanatory variables were selected out of about 30 ratios proxying for profitability, productivity, liquidity, financial structure, tension in credit relationships, growth, size and geographical location of the enterprises (Table 1).

For each model, two thirds of the firms were used for the estimation; the rest were used to test out of sample. Since in the estimation sample the proportion of sound and insolvent firms mimics that of the universe, the forecast values of the logistic regression can be regarded as the probability of default of the individual firms within one year. The overall correct classification rate - the fraction of firms that are correctly classified by the model as sound or insolvent - is around 74% on average (Table 2). Out of sample, similar percentages are observed for both sound and insolvent firms. The overall performance is also assessed using the power curve (or "Gini curve"), considering the results of the model out of the sample in the year of estimation and the full sample in other periods. This curve measures the discriminatory power of the function; that is, the overall ability of the model to distinguish sound from insolvent firms. A related measure is the accuracy ratio, the ratio of the area between the power curve and the random model to the area between a perfect model and the random model. The model produced an accuracy ratio of 65% for the control sample and of 66-67% for each of the years 2001, 1999 and 1998 (Graph 2). The value of accuracy ratios mentioned in studies regarding other countries normally ranges between 50 and 70%.[14] In the following application we consider 104,300 firms for which, in addition to an estimate of the probability of default, interest rates on credit lines are also available. The sample accounts for nearly 40% of total corporate loans of the banks for which interest rates are known. A set of about 255,000 credit relationships is considered.

The sample includes both large companies and SMEs (Table 3):[15]

- 1,900 firms with sales higher than €50 million account for 2% of those included in the sample and for 41% of lending to the sample;
- 20,000 firms with annual sales between €5 and €50 million represent 19% of the sample and 37% of loans. As mentioned above, in the new regulatory proposal capital requirements on loans to these firms are reduced by up to 20% relative to larger firms;
- 46,000 firms with sales of less than €5 million and an exposure higher than €1 million account for 44% of the total number of enterprises and for almost 19% of lending. Capital requirements on these exposures are reduced by 20%, other things being equal, relative to larger firms;

[12] In contrast to the rating systems normally adopted by international banks, which are nearly always determined according to both quantitative and qualitative information, this procedure relies on the first type of information only. Thus, it is only an approximation of the ratings that banks would normally estimate.

[13] This definition of default is narrower than that endorsed by the Basel Committee in the New Basel Capital Accord proposal, which also covers substandard loans and loans 90 days past due.

[14] For further details see Cannata et al (2002).

[15] For simplicity, the amount of bank debt is considered as a single item.

- finally, 36,000 firms accounting for 35% of the sample and 4% of total lending could be included in the retail portfolio provided their exposures are managed as part of a portfolio segment.

The probabilities of default of these firms have been estimated on the basis of company accounts for 2000 and credit relationships for 2001. Therefore their PDs represent an estimate of the default rate in 2002.

Their average value, weighted by the amount of lending, turns out to be 0.93%, lower than the default rate of all bank corporate borrowers recorded in the Credit Register (1.3% in 2002). The gap is mainly due to the overrepresentation of big firms in the sample in relation to smaller firms.

By applying the risk-weight functions contained in CP3, it is possible to obtain a proxy of the new capital requirements.

For the whole sample, the overall minimum capital requirements would be equal to 5.8% of total exposures (Table 4). However, results have to be interpreted with caution, given the data limitations, the narrower default definition adopted, and the bias of the sample towards better credit quality with respect to the average of banks' corporate borrowers.

3.2 Risk assessment and interest rates

As already mentioned, the impact of the New Basel Capital Accord on the pricing of bank loans can be checked by comparing the risk assessment of lending operations that is implicit in the CP3 document and the interest rates currently charged to borrowers. Since we assume that loans are senior unsecured, only the interest rates on short-term loans are considered; collateral should be less relevant in this case.

As a general point, Graph 3 shows an increasing relationship between firms' riskiness and the average interest rate on the loans granted to the firms in the same risk class, even though there is a substantial variation around the mean. A significant relationship is confirmed by a simple regression on cross section data. Similarly, the comparison between the average rate charged by each bank to its own borrowers and the average riskiness of the same borrowers also shows an increasing relationship between the two variables (Graph 4).

This evidence supports the reliability of our assessment of the financial conditions of bank borrowers and strengthens the results regarding the impact of the new capital adequacy framework on banks' lending policies to the sample of non-financial firms considered.

In the simplified formula we have adopted to measure the overall risk component of lending, that is, $EL/EAD + k(REQ - EL)/EAD$, the value of k, the rate of return on bank capital, is proxied by a weighted average of the pre-tax return on equity and of the interest rate on subordinated debt, net of the interest rate on risk-free assets (in which it is assumed that own funds are invested). As a result, for 2001 the rate of return on bank capital turned out to be equal to 10.3%. For the whole sample, the risk component ranges between 0.25 and 2% for the loans to borrowers with a PD not higher than 4%; its average value is 0.97%.

The risk component tends to be relatively low for larger firms, as a consequence of lower PDs; on the other hand, smaller borrowers can benefit from a favourable treatment in the definition of capital requirements. For firms with reported sales higher than €50 million the average risk component is equal to 0.85%, whereas it turns out to be 0.95% for firms with sales up to €5 million. The average risk component peaks at 1.2% both for corporate firms with sales lower than €5 million and for those included in the retail portfolio.

In Graphs 5-9 the change in the overall risk component of lending operations related to borrowers of different quality, as measured according to CP3, can be compared with the corresponding increase in the interest rates on loans. In this framework, we are not interested in explaining the level of the interest rates, which is influenced by several other factors. On the contrary, we are interested in their changes along the whole spectrum of the borrowers' PDs.

For corporate and retail portfolios the two variables, ie overall risk component and short-term interest rates, move together in response to an increase of the borrowers' PDs.

This evidence indicates that the risk-weight functions included in the new capital adequacy framework, for a given risk, are on average consistent with banks' pricing decisions. As a consequence, lending decisions are unlikely to be altered by the introduction of the regulatory scheme.

4. New capital requirements and procyclicality: a stress testing exercise

The potential impact of the new capital requirements on the pricing of bank loans has been assessed with reference to a period in which the overall quality of credit was particularly good in Italy. Indeed, in recent years the favourable trend of corporate profitability has been reflected in improved loan quality. Moreover, Italian banks have tended to direct their lending to less risky counterparties across borrowers of different sizes, economic sectors and geographical areas. Therefore, the result according to which the new capital adequacy framework is not likely to alter the lending decisions of banks needs to be made more robust by considering an unfavourable cyclical situation. For this purpose it is necessary to simulate a sudden deterioration of the financial condition of the corporate sector.

This amounts to dealing with the problem of procyclicality of loan supply in the framework of the New Accord, namely assessing the impact of the new capital requirements on lending decisions in the context of an economic slowdown. In fact, if capital requirements were to react too severely to an increase in the riskiness of lending activity, banks could decide to sharply restrict the supply of loans, thereby contributing to a further deterioration of the macroeconomic environment.

A certain degree of procyclicality in banks' lending decisions is a common experience of all countries: a slowdown in economic activity tends to be considered as an early indicator of increased financial fragility of the corporate sector and to be reflected in higher risk premiums on lending operations.

In order to perform a stress test with reference to the Italian economy, we tried to replicate the financial conditions of corporate borrowers in the recession that occurred in Italy at the beginning of the 1990s (a "worst case" scenario). The slowdown started in the second quarter of 1992; the percentage change of GDP with respect to the corresponding quarter of the previous year turned out to be negative in real terms in the last quarter of 1992 and in the first three quarters of 1993. In 1993 Italy's gross domestic product declined by 0.9% at constant prices, the first contraction since 1975.

The economic recovery took place in 1994, as a consequence of an acceleration of export growth driven by the fall in the exchange rate and wage moderation in the framework of increased world trade. However, in banks' balance sheets the volume of bad debts and substandard loans continued to increase substantially for some years: bad debts peaked at 10% as a ratio to total loans in 1997, although in the following years this ratio rapidly decreased, down to 4.5% at the end of 2002, partly as a result of loan securitisation.

A small number of financial ratios is sufficient to describe the severity of the financial situation of the industrial and commercial firms in the 1992-93 recession and the improvements recorded in the most recent period (Graph 10): (1) gross operating profit as a ratio to value added recorded its minimum value in the 1990-92 period: 37.7% as opposed to 40.6% in the second half of the 1980s and 44.4% in 2001; (2) net financial costs increased from a yearly average of 22.2% of gross operating profit in the 1985-89 period to a peak value of 29.7% in 1992; they were equal to 3.4% in 2001; (3) the return on assets was negative in 1992 and 1993, as opposed to 2.1% in the second half of the 1980s; it was equal to 1.1% in 2001; (4) leverage peaked at 60% in 1992 and 1993, from 56.5% in the second half of the 1980s; it was equal to 50.7% in 2001.

In order to set up a distressed scenario, we compute: (a) the PDs of individual firms consistent with the financial situation of the Italian corporate sector at that time; (b) the corresponding capital requirements according to CP3; and (c) the overall risk component of each credit relationship.

The increase of the overall risk component with respect to the present situation provides a proxy of the increase we should expect to observe in the interest rate (net of the risk-free rate) charged to each borrower.

4.1 Measuring PDs in a distressed scenario

Data on firms' balance sheets starts in 1993; we therefore use both financial ratios and credit relationships as of 1993 in order to simulate a sudden deterioration in the probabilities of default of the corporate borrowers included in our sample for 2002.

The assumptions underlying this calculation are: (1) the logit regression estimated for recent years is also suited to estimating probabilities of default for the past. In fact, we did not check whether there could be a better algorithm to proxy the financial health of Italian firms in those years. However, we compared the ex post effective default rate in 1994, relative to the 1993 sample of firms, with the ex ante estimates and the results were satisfactory for all risk classes; (2) the shocks affecting the macroeconomic scenario are completely incorporated in the micro-variables used in the exercise, the impact on which is different depending on the economic sector, geographical area and size of the firms considered.

Finally, we assume that the downturn materialises suddenly and abruptly, starting from the relatively good situation for banks' portfolios recorded in 2001, whereas usually a slowdown in economic activity unfolds gradually over time. Moreover, as a result of capital requirements directly linked to the probability of default on loans, banks should usually behave more proactively, continuously adjusting bank capital and loan loss reserves to changes in the quality of their portfolios.

In order to compute the PDs in the distressed scenario we use all the information contained in the original set of 188,000 non-financial firms. However, the final results in terms of lending policies refer to the sample of 104,000 corporate borrowers for which information on loan rates is also available.

About 64,000 firms out of the 188,000, accounting for 56% of the loans extended by banks to the firms in the 2002 sample, were recorded both in CERVED's Company Accounts Register and in the Credit Register in 1993. For these firms we simply used the 1993 data to calculate the PDs.

For the 124,000 firms which are not included in the 1993 sample, we have modified the 2000 balance sheet values and the 2001 credit relationship indicators so as to reproduce on average the values recorded in 1993 by economic sector, geographical area and size, thereby maintaining relative differences among firms.

The simulated deterioration in the financial conditions of the corporate sector can be better assessed on the basis of the transition matrices referring both to the number of the firms and to the overall amounts of their financing. The ratio of borrowers included in the first two classes (PDs not higher than 0.45%) shrinks from 27.2% to 16.5% as a number (Table 5) and from 31.3 to 19.9 as a percentage of total bank loans (Table 6). On the other hand, the number of firms for which the PD exceeds 1% increases from 29.4% to 43%, whereas their overall lending increases from 20.3% to 33.3% of the total.

The average PD, which was 1.27% in 2002, would increase to 1.79%, as opposed to 1.51% recorded in 1994 for all the firms included in CERVED's Company Accounts Register. The weighted average PD would increase from 0.97% to 1.4%, as compared with an actual default rate of 1.87% in 1994 in CERVED's database.

Such a sudden deterioration of the macroeconomic framework would involve a 16% increase in the overall minimum capital requirements. At the end of 2002 the overall capital buffer of the Italian banking system was equal to 40% of the minimum amount of capital required.

4.2 Capital requirements and interest rates in a distressed scenario

The results of the stress test provide a first indication regarding the reduction of potential procyclicality effects of the New Accord relative to the proposal issued in January 2001. Indeed, the application of the risk-weight function contained in CP2 would have implied a 24% increase in the overall minimum capital requirement for this set of loans.

However, we are more interested in assessing whether the risk measurement implicit in the new regulatory scheme would force banks to charge their customers exceptionally high loan rates when confronted with an adverse macroeconomic scenario. If this were true, we should conclude that the New Accord would anyway involve an increase in the procyclicality of banks' lending decisions. In the opposite case, the change in the capital regulation would turn out to be at least neutral in relation to the current situation.

Similarly to the exercise we performed on the more recent data, we can compare the overall risk component of lending operations as computed on the basis of the distressed PDs with the interest rates which were actually recorded at the time of the recession. For each credit line, the increase of the overall risk component provides a proxy of the increase we should expect to observe in the interest rates on credit lines.

Unfortunately, the comparison can be performed only to a limited extent, due to data limitations. Ten years ago only 20,000 firms out of the 104,000 that are included in our sample were financed by banks participating in the survey on interest rates. For this reason, we complement the comparison based on individual data with the observation of the differential between the average short-term lending rate and the rate of return on treasury bills.

Graph 11 shows that the risk premium on lending operations increased sharply during the 1992-93 recession. The increase in the second half of the 1990s was not connected with concerns regarding firms' financial situation, since banks' interest rates were decreasing and corporate profitability strongly improving. On the contrary, it was linked to the convergence of domestic monetary conditions towards the situation prevailing in the other leading European countries, consistent with the reduction of actual and expected inflation in Italy. Finally, some increase in the risk premium on bank loans was observable at the end of 2000, when a deterioration of the macroeconomic environment also took place.

By applying the risk-weight functions included in CP3 to the PDs referring to the 1993 situation, it is possible to compute the overall risk component of lending operations that is consistent with a distressed scenario. For the whole sample of 104,000 firms, the average risk component comes out a quarter of a percentage point higher than in 2002: 1.22% compared with 0.97%. Its increase ranges between 18 basis points for firms with reported sales higher than €50 million and 35 basis points for firms with sales between €50 and €5 million and for those included in the retail portfolio.

As a consequence, the lending rates (net of the risk-free rate) should increase on average by a quarter of a percentage point, starting from the value of 2.9 percentage points recorded at the end of 2002. The new level, although quite high, would not reach peak values relative to those recorded in the previous recession. As a matter of fact, it would be between the two values observed in 1992 and 1993, which were 2.6% and 3.6% respectively.

A more detailed analysis can be performed with respect to the 20,000 companies for which the interest rates on credit relationships in 1993-94 are available and for a subset of 7,000 firms with sales between €5 and €50 million.

Graphs 12 and 13 differ slightly from the corresponding Graphs 9 and 6. The low number of firms included in the first risk bucket made it difficult to use this class as a benchmark; we therefore considered the levels of the average interest rates corresponding to the firms included in each bucket.

The results achieved for the more recent period are basically confirmed: even in a distressed macroeconomic environment, the interest rates charged to borrowers in 1994 move together with the overall risk component of lending operations as credit quality deteriorates. This seems to imply that, as a consequence of the new regulatory scheme, banks are not induced to behave in a more procyclical way than in the past. Thus, the New Basel Capital Accord is unlikely to alter banks' decisions regarding the financing of the economy even under a distressed scenario.

Indeed, the new regulation will stimulate the banking industry to introduce more forward-looking elements in the assignment of ratings, in order to make judgments less correlated with the business cycle. Moreover, the role of provisions in offsetting expected losses, as well as the need for banks to continuously adjust their capital endowments to changing risk, could actually reduce the procyclicality of loan supply.

5. Conclusions

The New Basel Capital Accord can promote a vast improvement in the risk measurement and management practices of banks. The flexibility of the approach allows regulation to adapt to institutions of different size and sophistication.

Among the issues that emerged after the publication of the January 2001 consultative document was the need to prevent any difficulty in the financing of small and medium-sized firms and to balance the goal of a higher risk sensitivity of capital requirements with the potential amplification of business cycle fluctuations. As confirmed by the results of the Quantitative Impact Study recently conducted by the Basel Committee (QIS3), the changes made to the original proposal and contained in the third consultative document reduce the capital charges for almost any risk level and deal better with the peculiarities of credit risk in the case of SMEs.

As a result of the new treatment of credit risk, the exposures to a substantial share of borrowers will attract a capital charge for credit risk lower than 8%. However, this is not sufficient to argue that banks' lending decisions will not be distorted by the application of the new framework. What is relevant is whether the assessment of credit risk implicit in the new regulatory scheme (CP3) substantially differs from the banks' own evaluation, as embodied in their lending rates.

With reference to a large sample of Italian firms, for which the probabilities of default have been computed on the basis of their balance sheets and credit relationships, we compared the change in the overall risk component of lending operations, defined according to the foundation IRB approach, with the interest rates charged by banks on individual credit lines at a recent date. Since we find that the two variables move together in response to an increase of the borrowers' PDs, we tend to conclude that the CP3 approach to measuring credit risk is consistent with banks' risk assessment.

This result is supported by the finding that the same relationship also holds in a distressed scenario which replicates the financial condition of the Italian corporate sector in the 1993-94 recession. This provides an indication that the procyclicality of loan supply is not strengthened by capital requirements more strictly related to the riskiness of lending operations.

On the basis of this empirical evidence, we do not expect loan pricing to be distorted as a consequence of the new capital adequacy framework.

Table 1

Estimating the probability of default of non-financial firms

Significant explanatory variables[1]

	Economic sector			
	Manufacturing	**Trade**	**Construction**	**Services**
"Geographical" (dummy) variables				
Central Italy	**	*	**	–
Southern Italy	***	***	–	–
"Credit Register" variables				
Drawn/granted amount (y avg)	***	***	–	***
Overshoot (avg no)	***	***	***	***
Δ (Drawn/granted amount)	***	*	–	–
"Balance sheet register" variables				
Value added/total assets	–	**	–	–
Current assets/current liabilities	***	–	–	–
Cash flow/total assets	***	–	–	***
Coverage ratio	**	**	**	–
Leverage ratio	***	***	***	***
Long-term debt/total debt	–	–	–	***

[1] Significance levels (Wald chi-squared statistic): *** 0.1%, ** 1%, * 5%.

Table 2

Estimating the probability of default of non-financial firms

Performance of the logistic regression model

Economic sector	**Sample composition**			**"Correct classification" rate (%)**			
				In-sample		**Out-of-sample**	
	No of sound firms	**No of insolvent firms**	**% insolvent**	**Sound**	**Insolvent**	**Sound**	**Insolvent**
Manufacturing	46,683	585	1.2	74.7	74.9	74.6	71.7
Trade	28,949	387	1.3	75.1	74.9	75.7	73.7
Construction	17,282	323	1.8	72.7	72.4	72.1	70.7
Services	22,915	242	1.0	75.0	74.8	75.2	81.8
Total or average	115,829	1,537	1.3	74.6	74.4	74.6	73.6

Table 3

The sample: number of firms and bank debt by risk bucket

Risk buckets (PDs)	Total sales									
	>€50 million (share)		€5-50 million (share)		<€5 million (share)		"Retail" (share)		Total (share)	
	No	Debt	No	Debt	No	Debt	No	Debt	No	Debt
0.00-0.15%	11.4	5.2	7.5	3.5	5.3	2.5	3.8	2.2	5.3	4.0
0.15-0.45%	33.4	35.9	27.0	23.9	21.8	17.7	18.5	14.4	21.9	27.3
0.45-0.70%	26.2	30.3	25.4	25.6	21.9	18.8	18.7	15.8	21.5	25.9
0.70-1.00%	18.3	19.5	22.8	25.5	23.0	23.5	20.3	20.8	21.9	22.5
1.00-2.00%	8.0	6.7	12.8	16.1	18.1	23.5	21.0	24.4	17.9	13.9
2.00-4.00%	1.8	2.0	2.8	3.5	5.4	7.3	8.8	11.1	6.0	3.9
>4.00%	0.8	0.4	1.8	2.0	4.5	6.9	8.9	11.3	5.5	2.6
Number of firms	1,915		20,078		45,935		36,381		104,309	
Bank loans	79,605		71,802		36,354		7,333		195,093	
Average PD	0.62		0.89		1.49		1.91		0.93	

Table 4

Expected losses, capital requirements and risk components

Whole sample

Risk buckets	Bank loans		Expected losses (%)	Capital requirements (%)	Risk components (%)
	Amount	%			
0.00-0.15%	7,729	4.0	0.04	2.09	0.25
0.15-0.45%	53,201	27.3	0.14	4.06	0.54
0.45-0.70%	50,502	25.9	0.26	5.63	0.82
0.70-1.00%	43,898	22.5	0.38	6.41	1.00
1.00-2.00%	27,201	13.9	0.58	7.21	1.26
2.00-4.00%	7,529	3.9	1.24	9.27	2.07
>4.00%	5,033	2.6	3.83	16.38	6.40
Total	195,093	100.0	0.42	5.81	0.97

Table 5

Stress test: transition matrix for the whole sample

Number of firms

Risk buckets (2002)	Amount	Share	Risk buckets (simulated 1994)								Number of firms in simulated 1994
			0.00-0.15%	0.15-0.45%	0.45-0.70%	0.70-1.00%	1.00-2.00%	2.00-4.00%	>4.00%	Total	Shares
0.00-0.15%	5,534	5.3	44.2	35.6	9.2	5.2	3.8	1.4	0.8	100.0	2.3
0.15-0.45%	22,821	21.9	–	56.2	25.3	8.9	5.9	2.3	1.4	100.0	14.2
0.45-0.70%	22,450	21.5	–	–	55.3	28.8	10.1	3.4	2.4	100.0	17.9
0.70-1.00%	22,852	21.9	–	–	–	64.5	27.5	4.4	3.6	100.0	22.6
1.00-2.00%	18,701	17.9	–	–	–	–	76.9	17.7	5.4	100.0	23.4
2.00-4.00%	6,258	6.0	–	–	–	–	–	73.9	26.1	100.0	9.9
>4.00%	5,693	5.5	–	–	–	–	–	–	100.0	100.0	9.7
Total	104,309	100.0	2.3	14.2	17.9	22.6	23.5	9.9	9.7	100.0	100.0

Table 6

Stress test: transition matrix for the whole sample

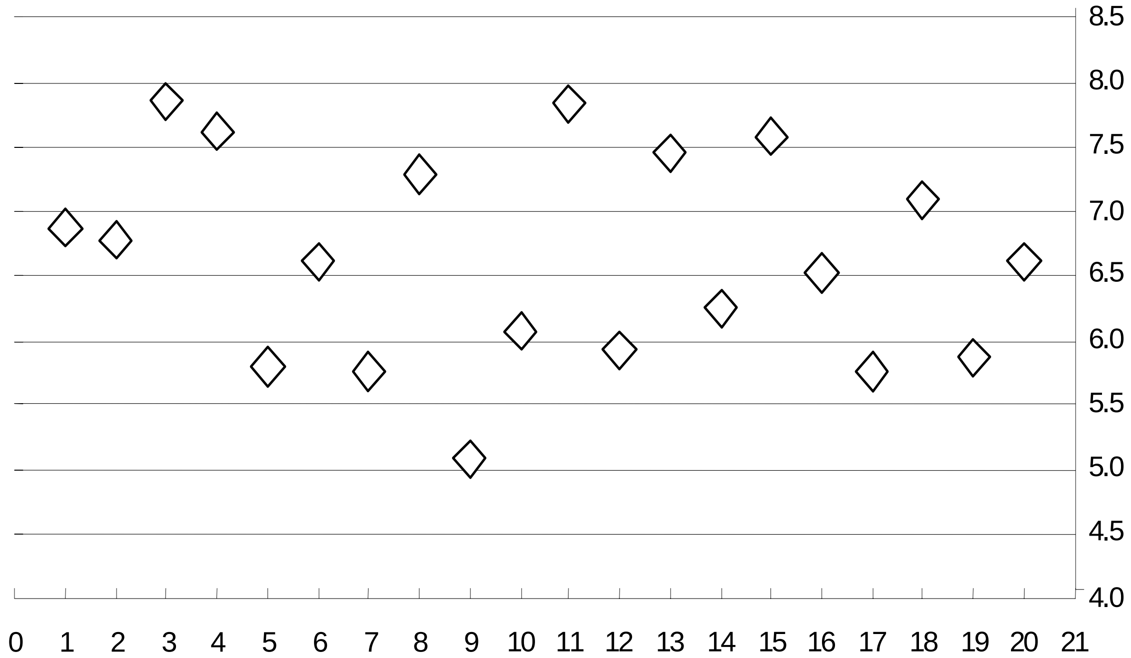
Bank debt

Risk buckets (2002)	**Amount**	**Share**	**Risk buckets (simulated 1994)**								**Bank debt in simulated 1994**
			0.00-0.15%	**0.15-0.45%**	**0.45-0.70%**	**0.70-1.00%**	**1.00-2.00%**	**2.00-4.00%**	**>4.00%**	**Total**	**Shares**
0.00-0.15%	7,729	4.0	35.0	44.5	13.3	2.6	2.8	1.6	0.3	100.0	1.4
0.15-0.45%	53,201	27.3	–	61.5	25.4	6.3	4.6	1.3	1.0	100.0	18.5
0.45-0.70%	50,502	25.9	–	–	53.8	33.6	9.4	2.0	1.3	100.0	21.4
0.70-1.00%	43,898	22.5	–	–	–	65.7	25.5	5.0	3.8	100.0	25.3
1.00-2.00%	27,201	13.9	–	–	–	-	83.1	11.4	5.5	100.0	21.1
2.00-4.00%	7,529	3.9	–	–	–	–	–	80.5	19.5	100.0	6.8
>4.00%	5,033	2.6	–	–	–	–	–	–	100.0	100.0	5.5
Total	195,093	100.0	1.4	18.5	21.4	25.3	21.1	6.8	5.5	100.0	100.0

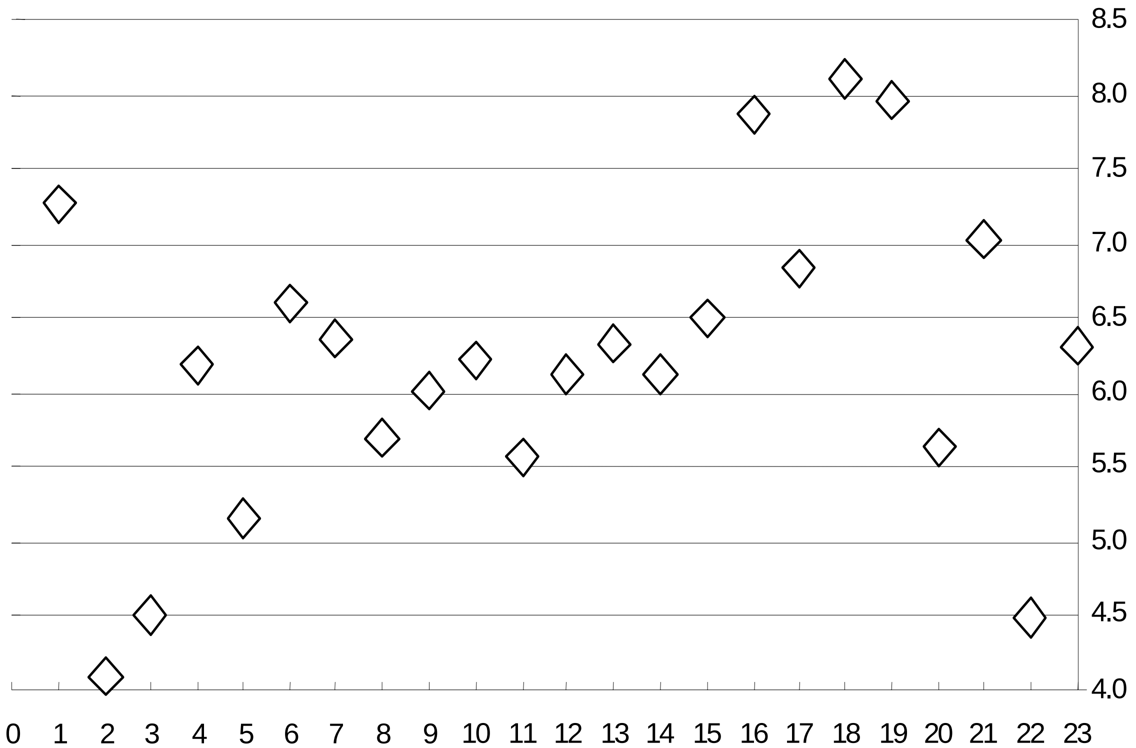
Graph 1

Interest rates on bank loans

(a) by region



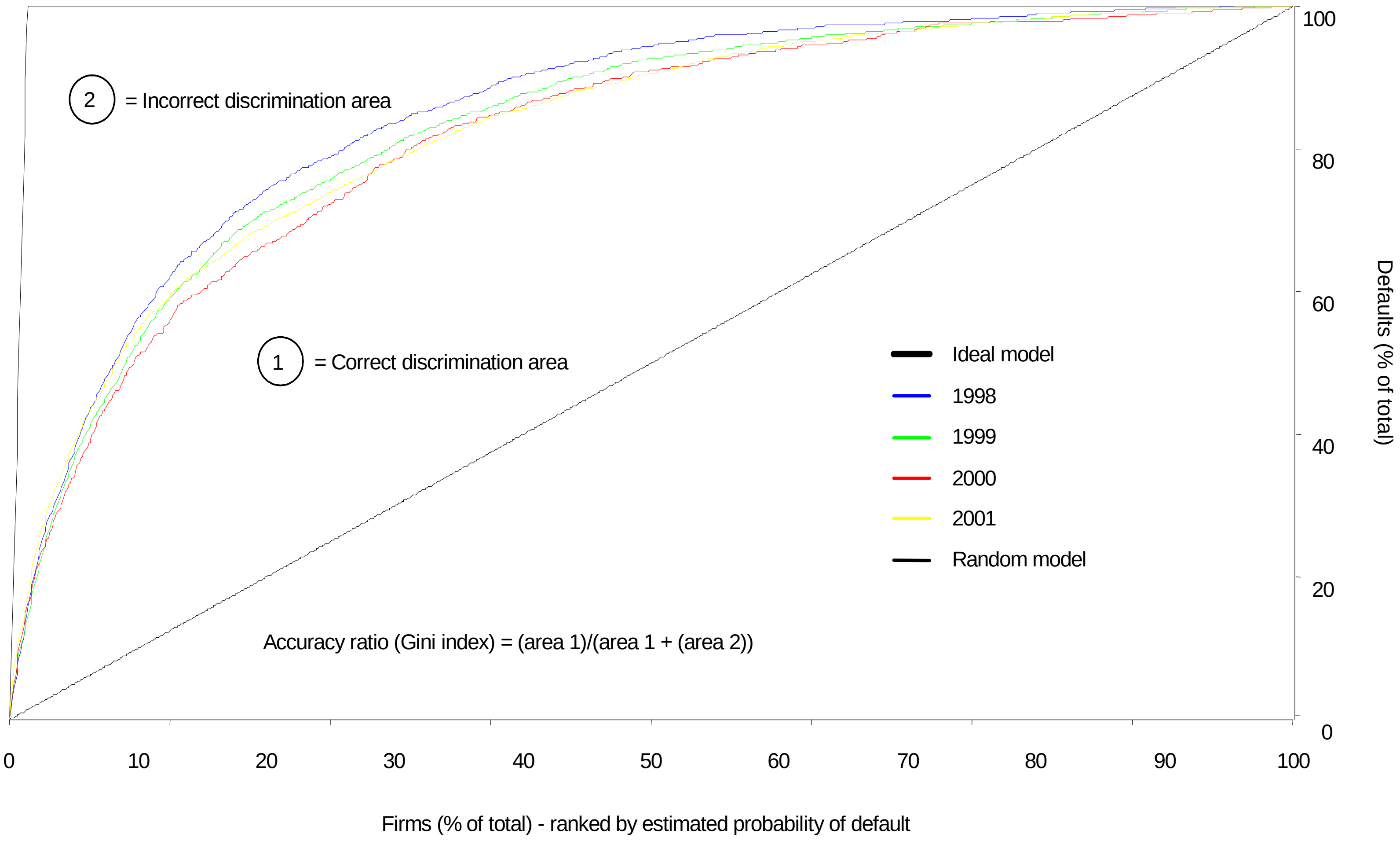
(b) by economic sector



Graph 2

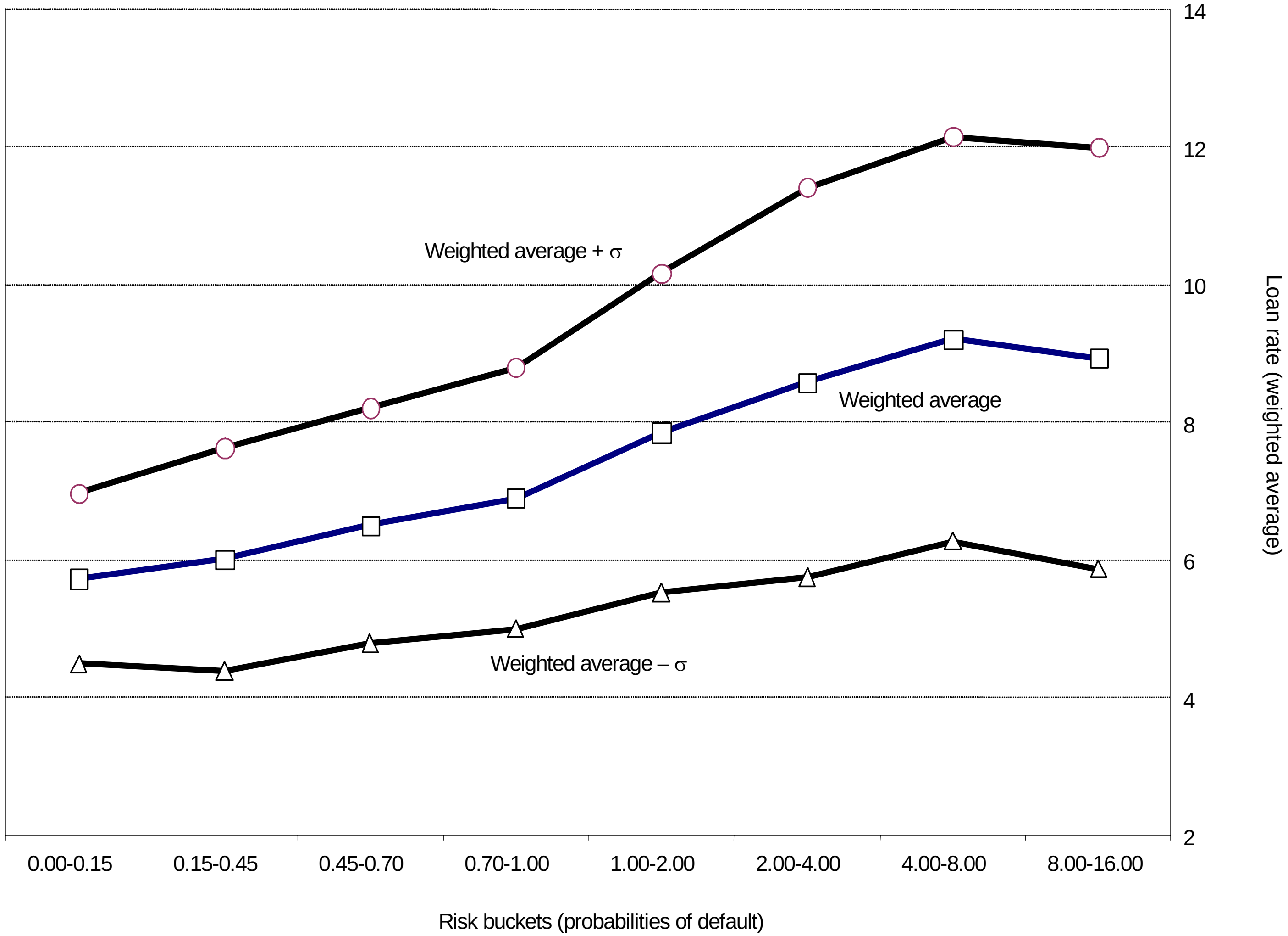
Estimating the probability of default of non-financial firms

Model accuracy



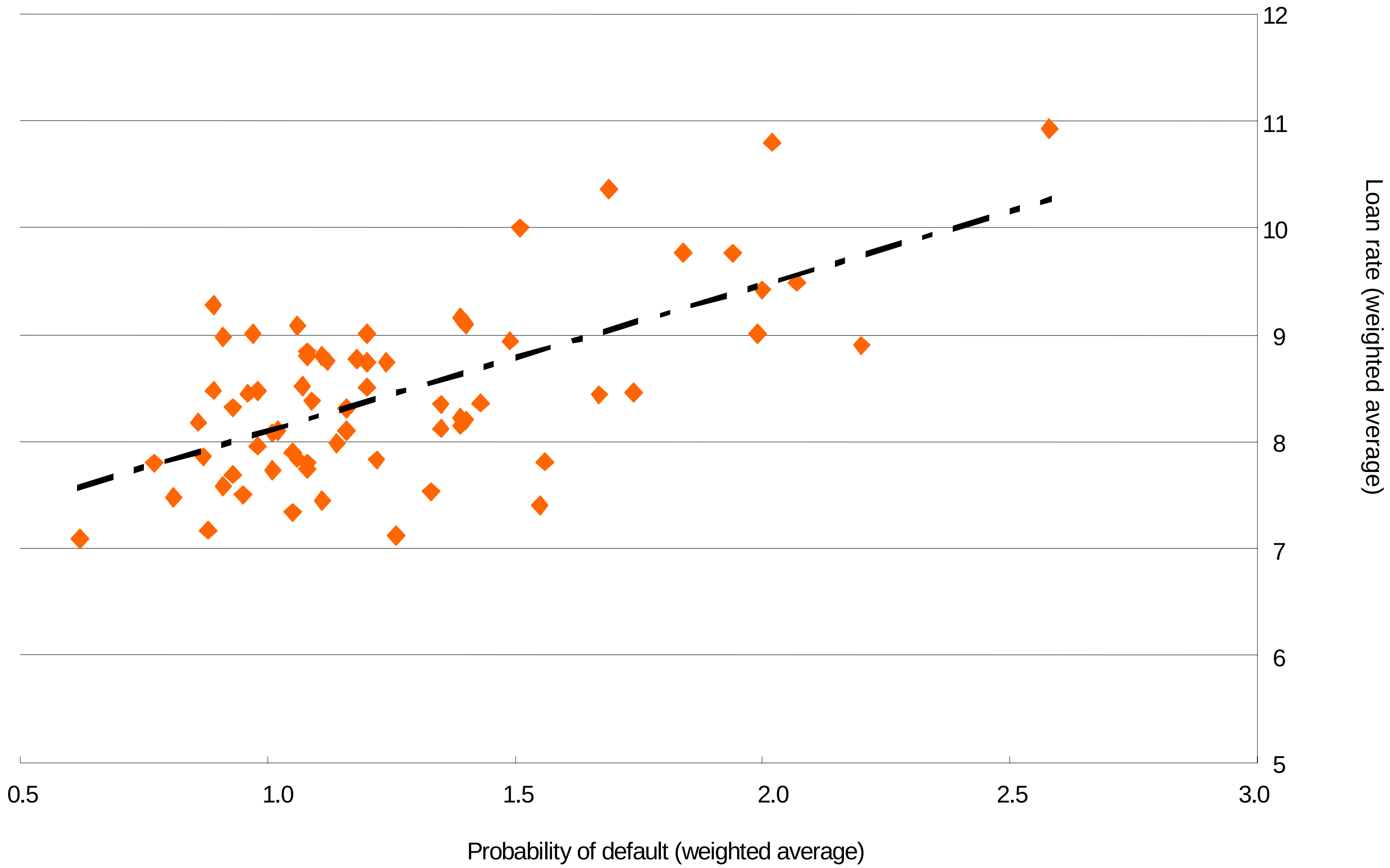
Graph 3

Probabilities of default and loan rates



Graph 4

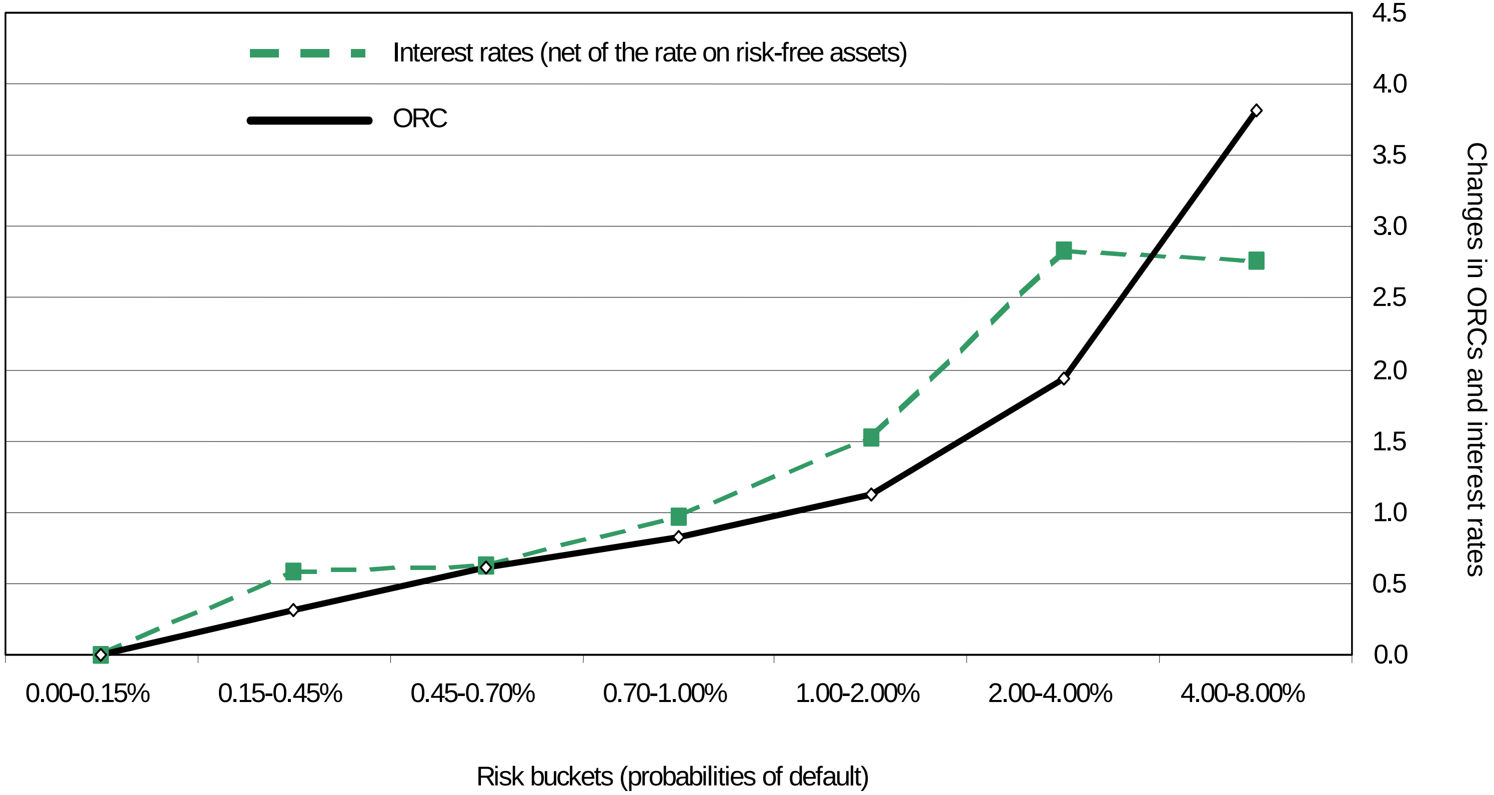
Distribution of banks by loan rate and firm riskiness



Graph 5

Changes in ORCs and interest rates by risk bucket

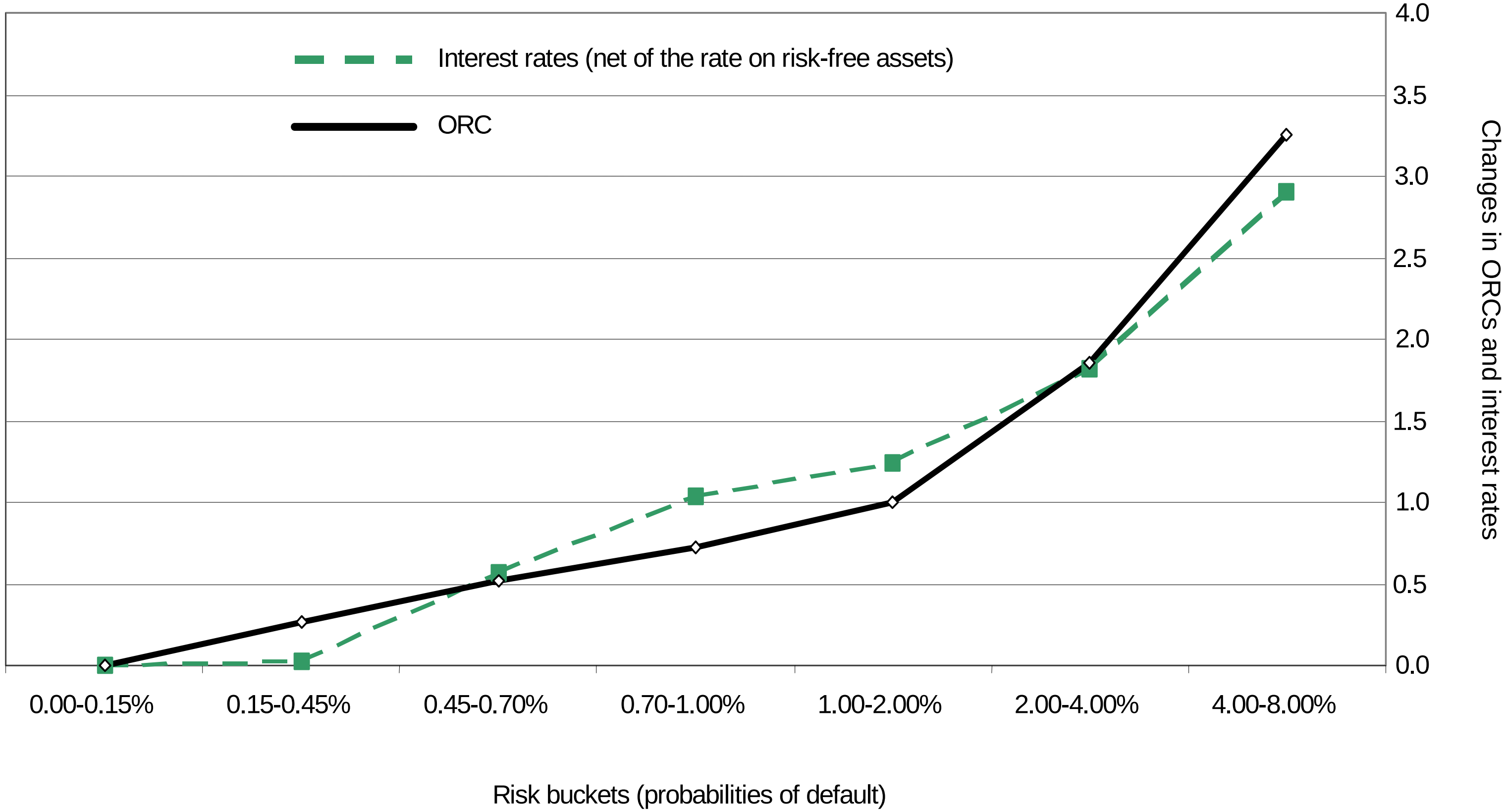
Total sales >€50 million



Graph 6

Changes in ORCs and interest rates by risk bucket

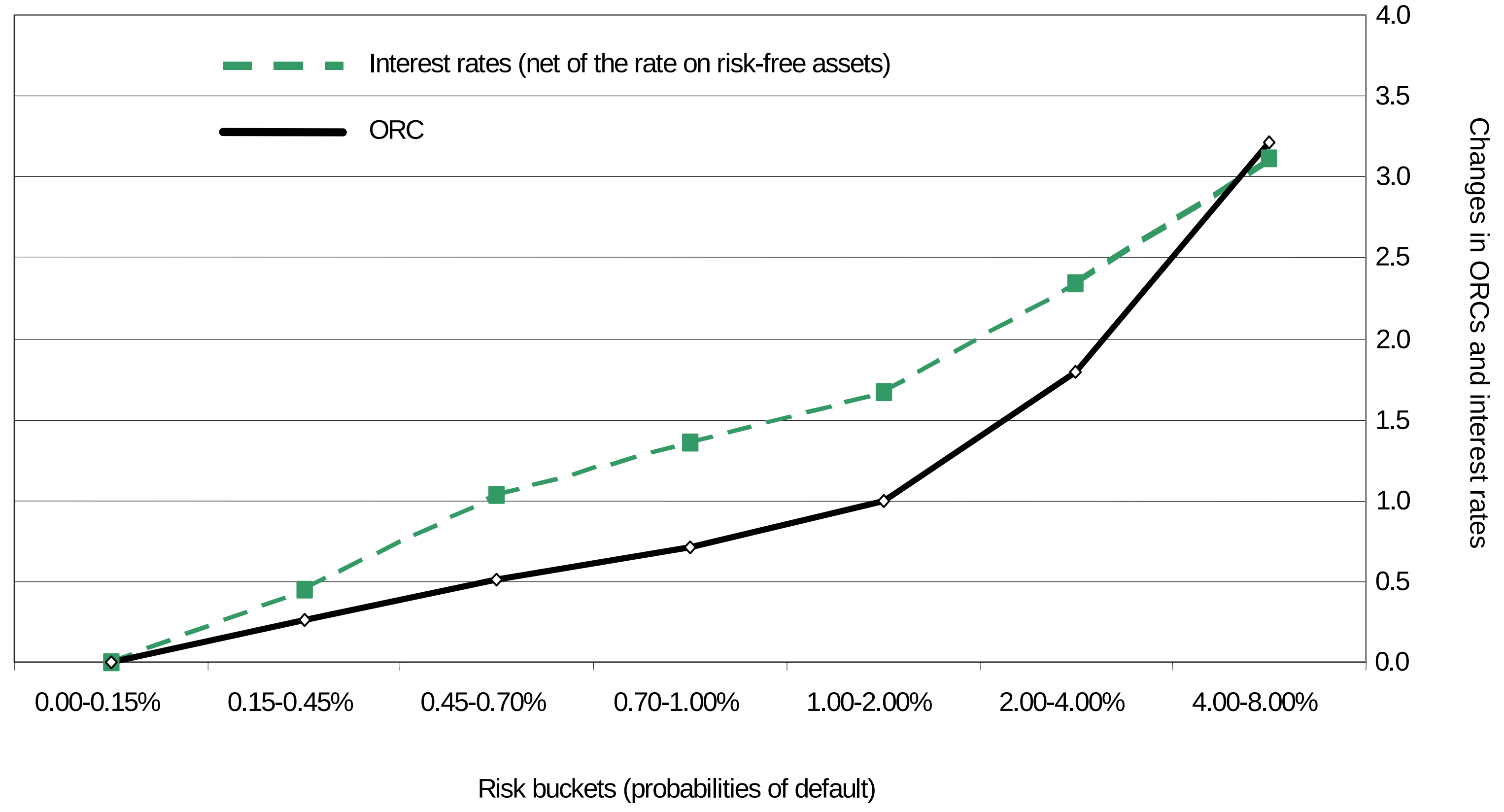
Total sales €5-50 million



Graph 7

Changes in ORCs and interest rates by risk bucket

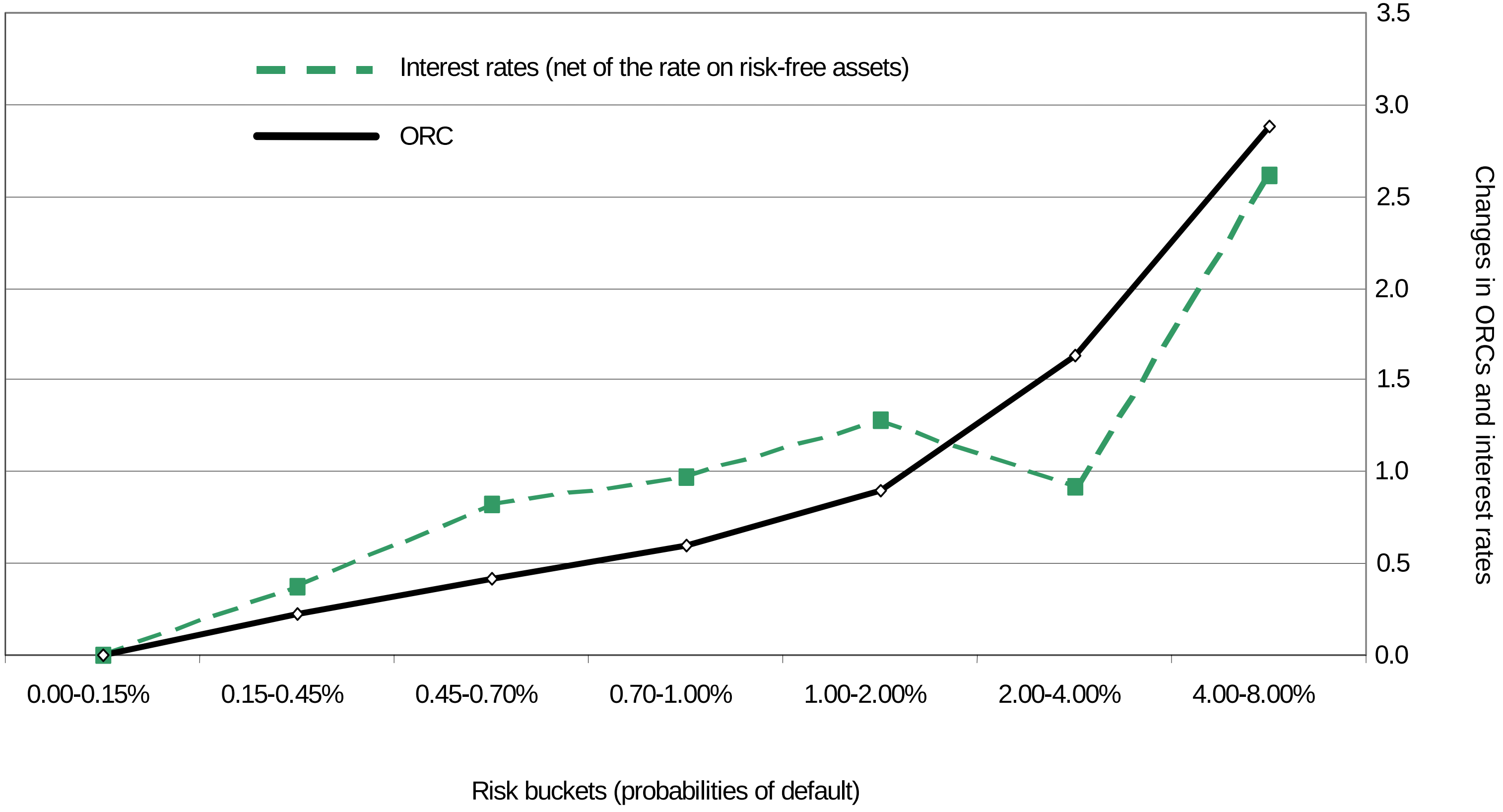
Total sales <€5 million



Graph 8

Changes in ORCs and interest rates by risk bucket

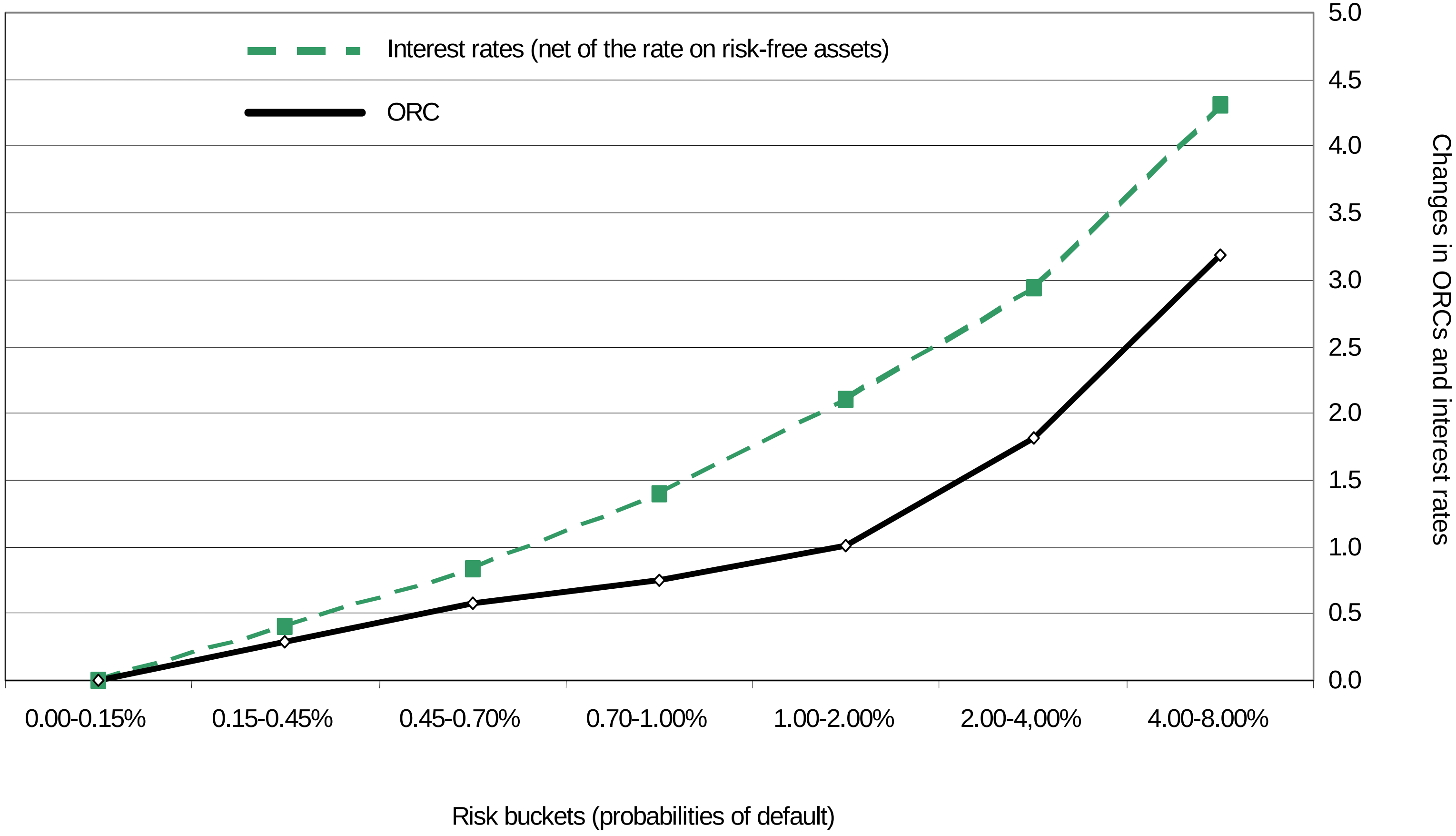
Retail portfolio



Graph 9

Changes in ORCs and interest rates by risk bucket

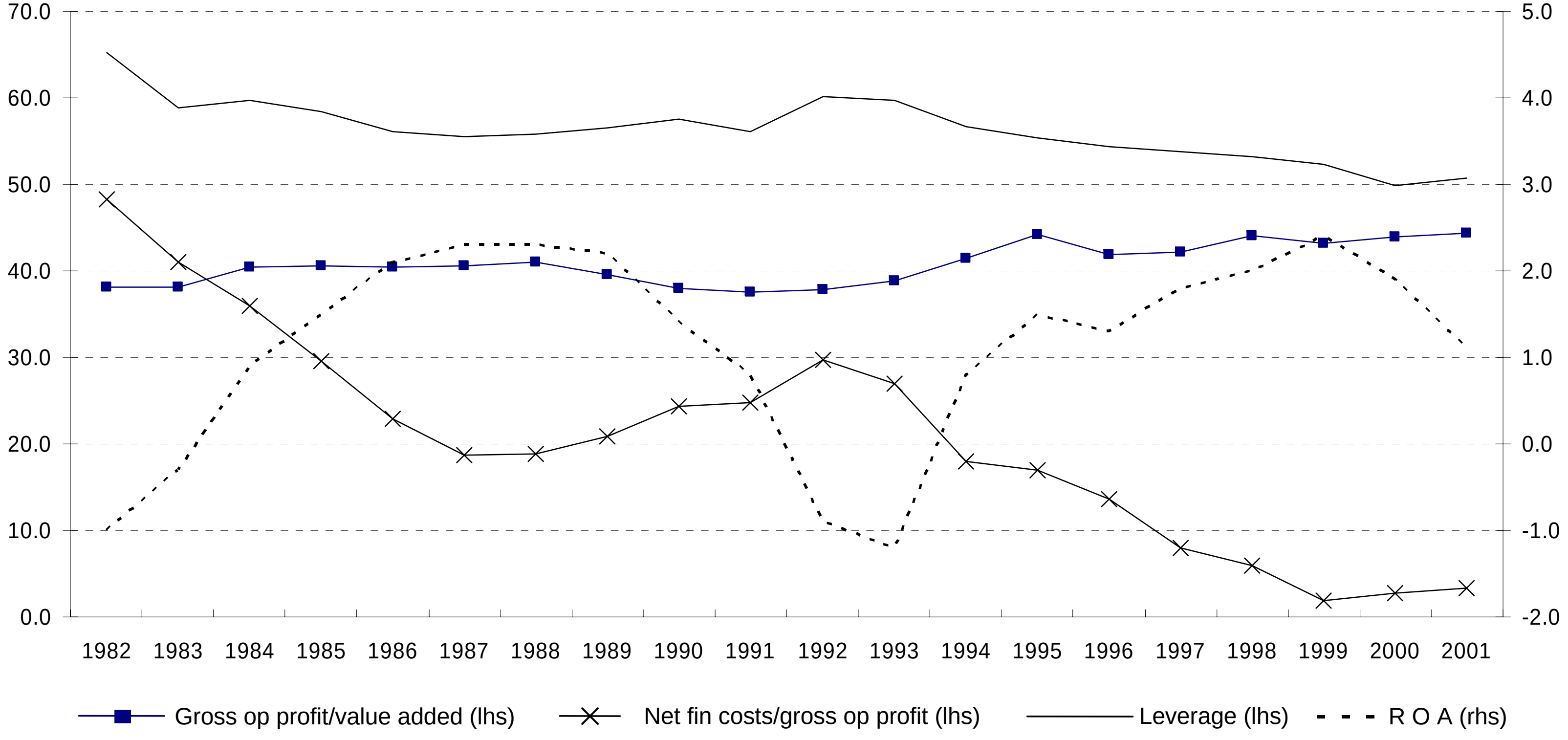
Whole sample



Graph 10

Industrial and commercial firms

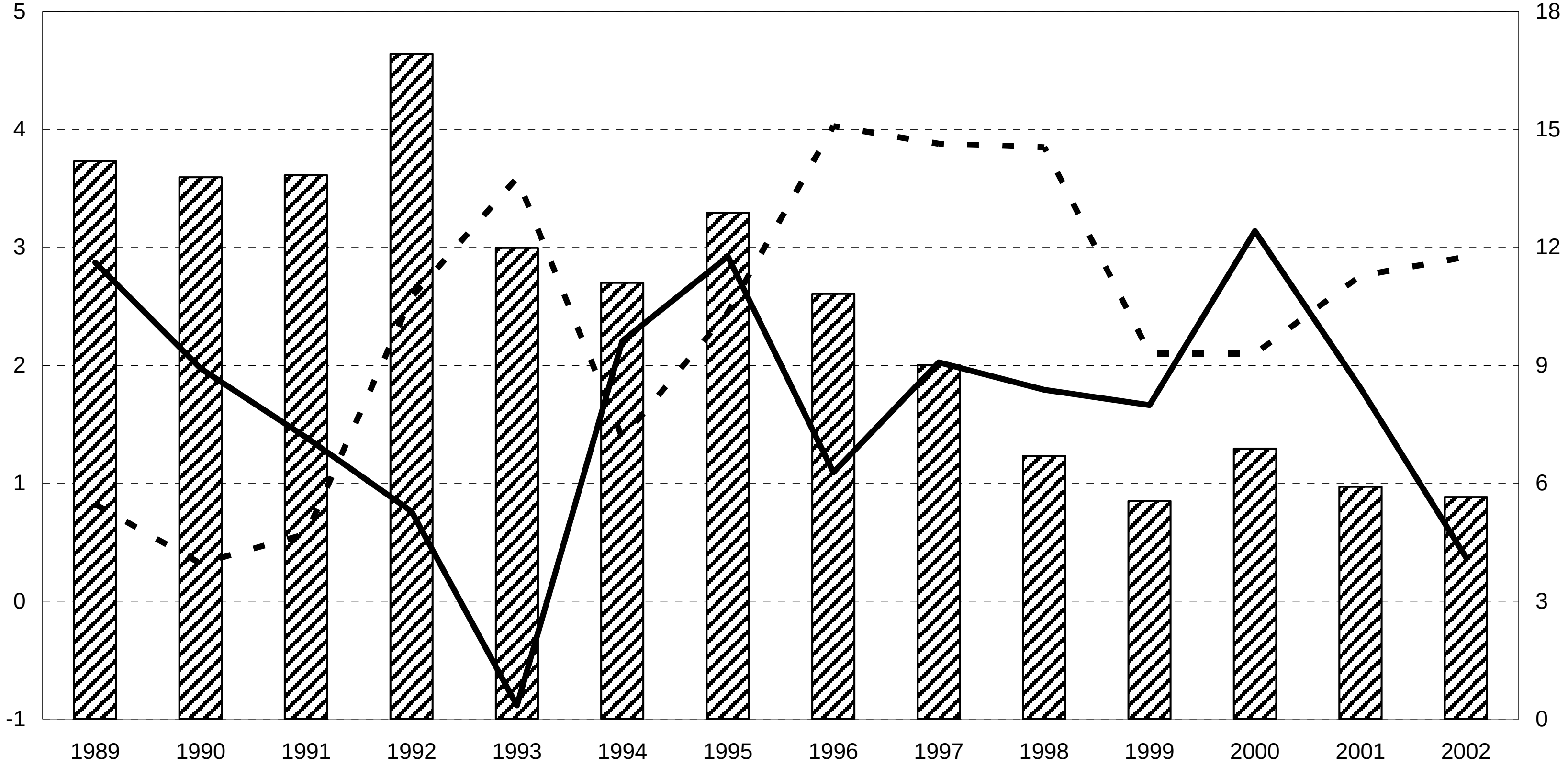
Accounting ratios



Source: Credit Register.

Graph 11

Gross domestic product and interest rates on bank loans



Interest rate on short-term lending (rhs)

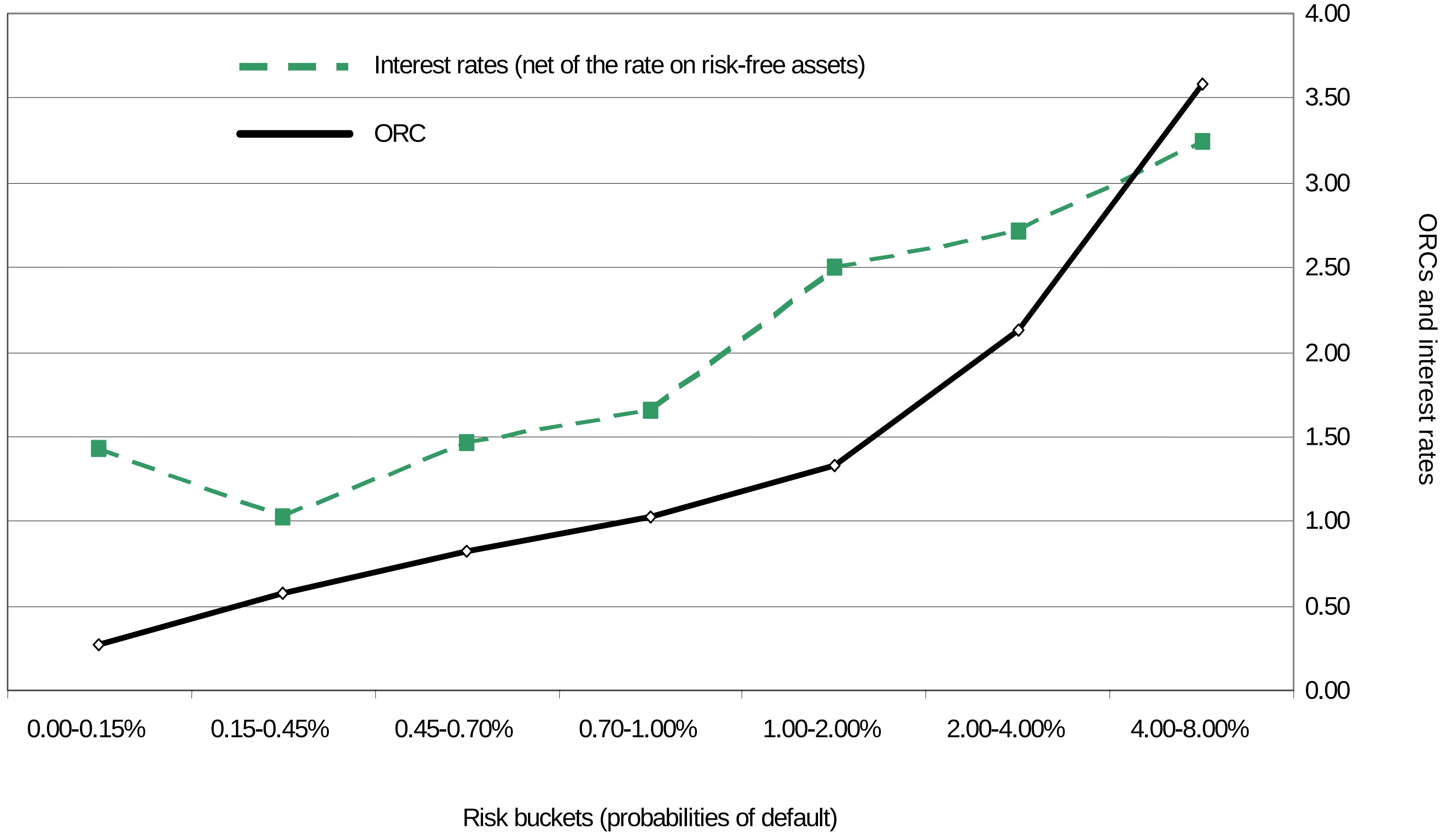
Percentage changes of GDP at constant prices (lhs)

Interest rate differential: short-term loans - treasury bills (lhs)

Graph 12

ORCs and interest rates by risk bucket in a distressed scenario

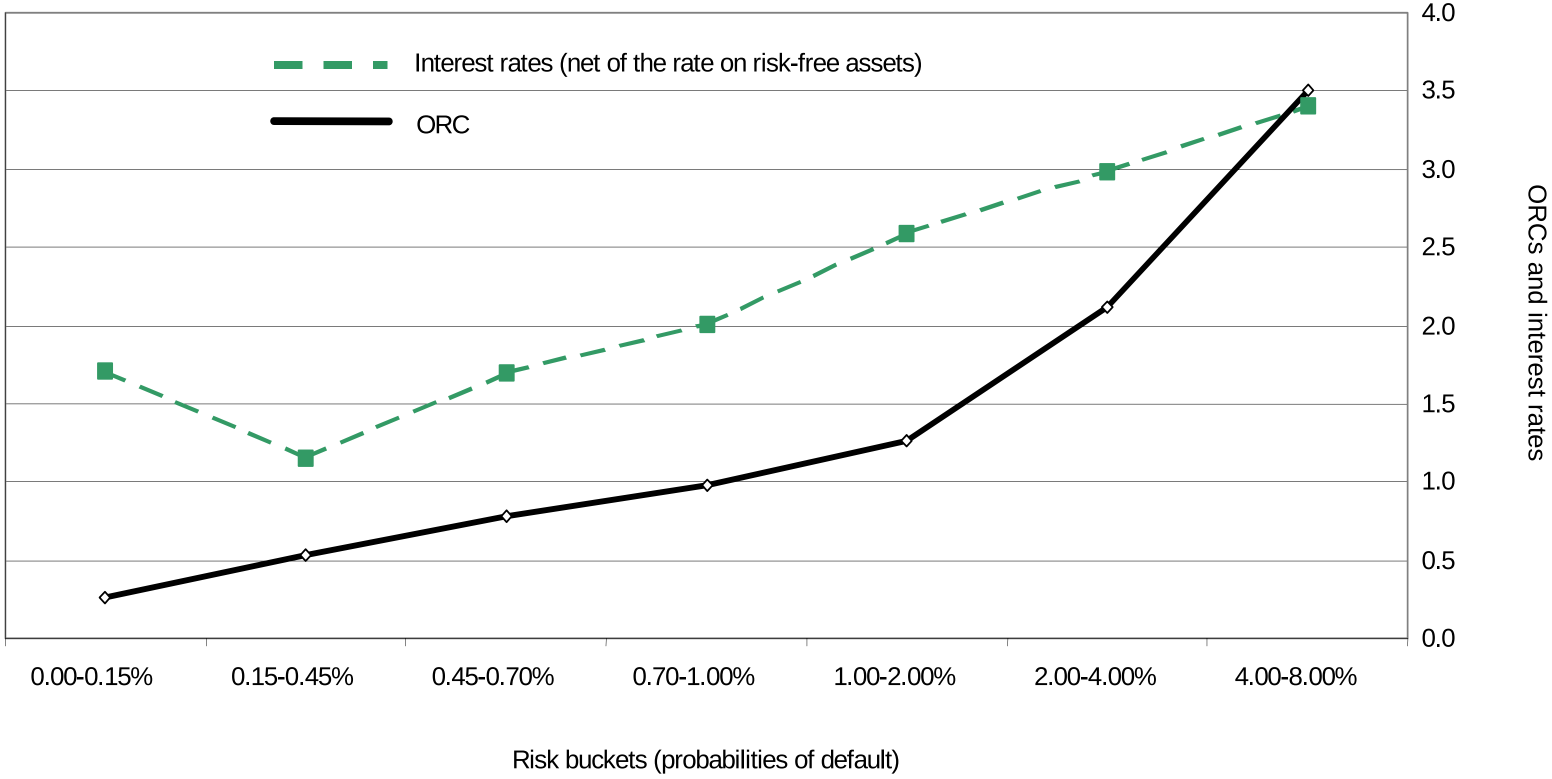
Whole sample



Graph 13

ORCs and interest rates by risk bucket in a distressed scenario

Total sales €5-50 million



6. References

Allen, L and T Saunders (2002): "A survey of cyclical effects in credit risk measurement models", presentation at the conference on "Changes in risk through time: measurement and policy options", BIS, Basel, 6 March.

Altman, E, A Resti and A Sironi (2002): "The link between default and recovery rates: effects on the procyclicality of regulatory capital ratios", paper presented at the conference on "Changes in risk through time: measurement and policy options", BIS, Basel, 6 March.

Associazione Bancaria Italiana (2001): *Italian banks' position*, paper on capital adequacy revision, May.

Ayuso, J, D Pérez and J Saurina (2002): "Are capital buffers pro-cyclical?", paper presented at the conference "Basel II: an economic assessment", BIS, May.

Basel Committee on Banking Supervision (2001a): *The New Basel Capital Accord*, consultative document, January.

——— (2001b): *Results of the Second Quantitative Impact Study*, November.

——— (2001c): *Review of procyclicality*, mimeo.

——— (2001d): *Potential modifications to the Committee's proposals*, November.

——— (2002a): *Results of Quantitative Impact Study 2.5*, June.

——— (2002b): Press statement, 10 July.

——— (2002c): *QIS3 technical rules*, October.

——— (2003a): *The New Basel Capital Accord*, consultative document, April.

——— (2003b): *Quantitative Impact Study 3 - overview of global results*, May.

——— (2003c): *Supplementary information on QIS3*, May.

Borio, C, C Furfine and P Lowe (2001): "Procyclicality of the financial system and financial stability: issues and policy options", *BIS Papers*, no 1.

Cannata, F, F Fabi and S Laviola (2002): "Rating interni e gestione del rischio di credito: la misurazione della probabilità di insolvenza", *Bancaria*, no 4.

Cannata, F, P Grippa and S Laviola (2001): *Calibration of benchmark risk weights in the internal ratings based approach*, Bank of Italy, mimeo, October.

Cannata, F and S Laviola (2001): "Il nuovo Accordo sul Capitale delle banche: i commenti dell'industria bancaria", *Rivista Bancaria - Minerva Bancaria*, no 4.

Carosio, G (2001): "Il Nuovo Accordo sul Capitale: aspetti generali", *Bancaria*, no 4.

——— (2002): *Il nuovo Accordo sul Capitale: aggiornamento sui lavori del Comitato di Basilea*, Associazione per lo Sviluppo degli Studi di Banca e Borsa, Sadiba (Perugia), 15 marzo.

Catarineu-Rabell, E, P Jackson and D Tsomocos (2003): "Procyclicality and the new Basel Accord - banks' choice of loan rating system", *Bank of England Working Paper Series*, no 181, March.

Danielsson, J, P Embrechts, C Goodhart, C Keating, F Muennich, O Renault and H Song Shin (2001): "An academic response to Basel II", *Special Paper Series*, no 130, Financial Markets Group, London School of Economics.

Dietsch, M and J Petey (2003): *Should SME exposures be treated as retail or corporate exposures? A comparative analysis of default probabilities and asset correlations in French and German SMEs*, mimeo.

Dullmann, K and H Scheule (2003): "Asset correlation of German corporate obligors: its estimation, its drivers and implications for regulatory capital", paper presented at the Basel Committee's Research Task Force Workshop on "Banking and Financial Stability", Bank of Italy, Rome, March.

Ervin, W and T Wilde (2001): "Procyclicality in the new Basel Accord", Credit risk special report, *Risk Magazine*, October.

Estrella, A (2001): *The cyclical behaviour of optimal bank capital*, Federal Reserve Bank of New York, mimeo.

European Central Bank (2001): "The new capital adequacy regime, the ECB perspective", *ECB Monthly Bulletin*, May.

Institute of International Finance (2001): *Report of the Working Group on Capital Adequacy*, May.

International Swaps and Derivatives Association (2001): *ISDA's response to the Basel Committee on Banking Supervision's consultation on the New Capital Accord*, May.

Jackson, P, C Furfine, H Groeneveld, D Hancock, D Jones, W Perraudin, L Radecki and M Yoneyama (1999): "Capital requirements and bank behaviour: the impact of the Basel Accord", *Basel Committee on Banking Supervision Working Papers*, no 1.

Jordan, J, J Peek and E Rosengren (2002): "Credit risk modeling and the cyclicality of capital", paper presented at the conference on "Changes in risk through time: measurement and policy options", BIS, Basel, 6 March.

Laviola, S (2001): "Il Nuovo Accordo sul Capitale: il metodo dei rating interni", *Bancaria*, no 4.

Lopez, J (2002): "The relationship between average asset correlation, firm probability of default and asset size", paper presented at the conference "Basel II: an economic assessment", BIS, May.

Lowe, P (2002): "Credit risk measurement and procyclicality", paper presented for the conference "Banking Supervision at the Crossroads", held at the Netherlands Bank, Amsterdam, 24-25 April.

Masschelein, N (2003): *The Basel II Capital Accord, SME loans and implications for Belgium*, mimeo.

Perraudin, W and Y T Hu (2002): *The dependence of recovery rates and defaults*, Birkbeck College, UK, mimeo.

Resti, A (2002): *The New Basel Capital Accord: structure, possible changes, micro and macroeconomic effects*, mimeo.

Saurina, J and C Trucharte (2003): *The impact of Basel II on lending to small and medium-sized firms. A regulatory policy assessment based on the Spanish Credit Register*, Bank of Spain, mimeo.

Segoviano, M and P Lowe (2002): "Internal ratings, the business cycle and capital requirements: some evidence from an emerging market economy", *BIS Working Papers*, no 117, September.

Macro stress tests of UK banks

Glenn Hoggarth, Andrew Logan and Lea Zicchino
Bank of England

1. Introduction

Stress testing the vulnerability of financial institutions to adverse macroeconomic events is an important tool in assessing financial stability. Central banks and financial regulators increasingly use this approach in calibrating the risks facing the financial system. A number of recent policy initiatives also aim to formalise a role for stress tests. One of these has been the inclusion of stress tests in the IMF Financial Sector Assessment Programmes (FSAPs). Stress testing is also important as part of Pillar 2 of the New Basel Accord. For example, with regard to the procyclicality debate, macro stress testing might give some indication of how the impact on bank capital during a recession would vary depending on the *type* of recession (eg whether it is consumer- or export-led).

This paper describes a number of approaches used in the financial stability area of the Bank of England to stress test banks and draws on our experience from last year, when stress tests were carried out as part of the IMF's FSAP on the United Kingdom. We also outline some of our future proposed work.

2. Possible approaches to stress tests

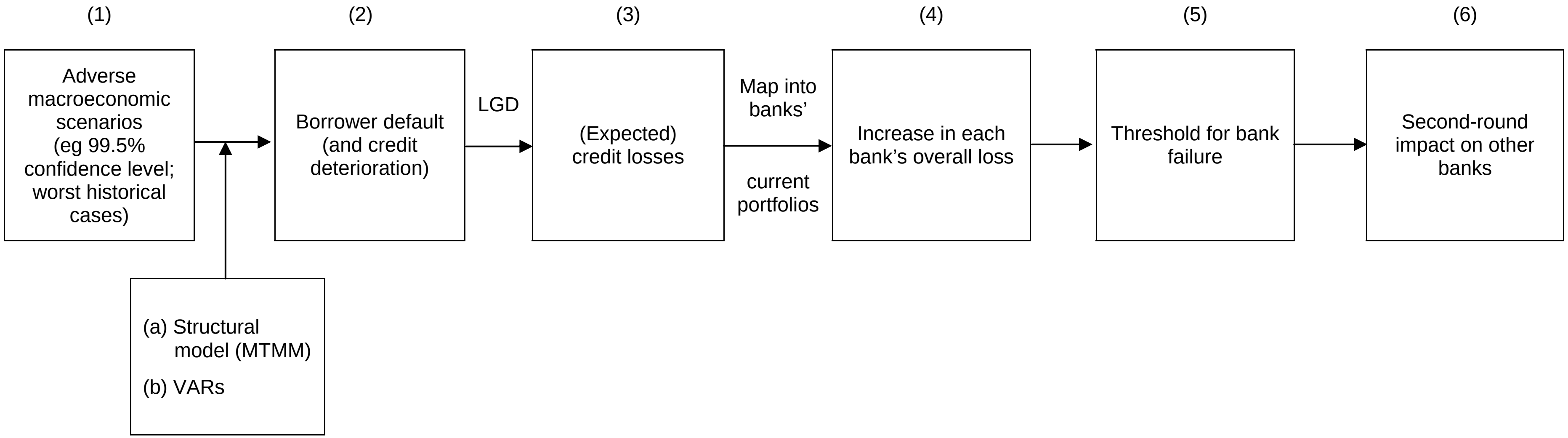
Stress tests involve a number of elements. These are illustrated in Figure 1. First, plausible and internally consistent but "challenging" macroeconomic scenarios or single factor sensitivity tests need to be devised to illustrate possible extreme downside risks - so-called "tail events" (Box (1)). Whereas the former assess the impact on credit risk of a combination of changes in macroeconomic variables, the latter focus on the change in one variable and assume that other variables remain unaffected. Second, these scenarios (or sensitivity tests) need to be mapped into measures of increases in credit default by loan type or borrower (Box (2)). Third, changes in borrower default need to be translated into bank credit losses, ie allowing for recoveries, by loan type (Box (3)).

In a "bottom-up" approach, each bank would estimate the increase in credit losses on its entire portfolio (allowing for the possibility that losses are interdependent). This was one of the approaches adopted in the FSAP exercise (see below and also Hoggarth and Whitley (2003)). Such an approach has the advantage of evaluating banks' portfolios at a detailed level of disaggregation. It also provides information on how banks themselves assess the likely impact of adverse events on the quality of their loan book. However, such estimates are not based on applying a consistent framework across banks and, in any case, would not be practical for the authorities to carry out on a frequent basis. An alternative approach is to adopt a "top-down" methodology. Here macroeconomic scenarios are linked to banks' aggregate sectoral losses.

The various approaches described below aim to estimate the impact of a variety of common macro shocks on the credit losses of the UK banking system (steps (1) to (3) in Figure 1). There are a number of approaches that can be used to carry out macroeconomic stress tests, and we have adopted an eclectic approach building upon the stress testing exercise conducted last year for the UK FSAP.

Figure 1

Framework for macro stress testing UK banks



UK FSAP:	bottom-up:			banks estimated their own losses - the major UK banks were given simulations from an extended version of the BoE's macroeconometric model (MTMM) 1(a). They gave us back (4), having done (2) and (3) themselves.
	top-down:	equations on banks' aggregate provisions	-	direct from MTMM simulations (1)(a) to (4) without intermediate steps.
	top-down sectoral:	equations on banks' sectoral write-offs	-	linking equations to an extended version of the MTMM (1(a)==>3)
		VAR model including sectoral write-offs	-	1(b)==>3

3. UK FSAP[1]

In the UK FSAP of 2002 we constructed specific macroeconomic scenarios derived using an extension of the Bank of England's then current Medium-Term Macroeconometric Model (MTMM). The outputs from these scenarios were supplied to 10 large UK banks as inputs to their own assessments (the "bottom-up" approach). The UK-owned institutions were asked to consider the effects on a consolidated basis. However, the results do not, in all cases, capture the impact on all their non-bank and foreign operations. The tests were conducted in spring 2002, and firms assessed the impact on their profit and loss account and regulatory capital during the first year (until March 2003) - compared with their own internal forecast or base line.[2]

The "bottom-up" results were returned to us and compared with our own analysis of the impact of the scenarios on UK banks (the "top-down" approach). The latter used aggregate reduced-form relationships linking changes in macroeconomic variables to banks' aggregate loan loss provisions.

The scenarios

Four scenarios were chosen in the UK FSAP exercise to include both domestic and global events, and shifts in both the demand for and supply of goods and services in the economy:

1. Decline of 35% in world and UK equity prices. The macroeconomic transmission is largely through household balance sheets, whereby lower household sector wealth reduces household consumption and hence aggregate GDP. But the impact on demand and output is partly offset by an easing in monetary policy in the United Kingdom and elsewhere. The main adverse consequences for the financial system are predicted to occur in the corporate sector, as a result of lower GDP and profits.

2. Decline of 12% in UK house and commercial property prices. Since housing accounts for one half of UK households' net worth, the personal sector's balance sheet deteriorates and UK household consumption is reduced. Output is lower than otherwise, but the adverse effect is a little smaller than under the first scenario. Similarly, the monetary authorities are assumed to respond by cutting UK interest rates. Nonetheless, the net effect is that mortgage arrears increase relative to base, even though they remain low by historical standards. Corporate sector income is expected to fall relative to base as a consequence of weaker aggregate demand, and capital gearing rises because of the decline in commercial property prices. This shock is expected mainly to hit banks with a high concentration of property loans.

3. A one and a half percentage points unanticipated increase in UK average earnings growth (reflecting a step increase in real reservation wages). This supply shock boosts personal incomes and consumption. But the transmission to higher inflation induces a rise in official interest rates. Overall there is a marginal decline in GDP compared with the base case. Both corporate and household sectors are adversely affected. Despite higher household incomes, there is a rise in income gearing, which implies an increase in household mortgage and credit card arrears. Corporate profits fall relative to base and corporate liquidations increase.

4. A 15% (initial) unanticipated depreciation in the trade-weighted sterling exchange rate. This results in higher inflation and, in response, nominal interest rates increase. Nonetheless, since wages and prices adjust only gradually, there is a temporary depreciation in the real exchange rate, which, in turn, boosts net export volumes. On balance, GDP growth is higher than otherwise. The corporate sector benefits from higher net exports, and profits rise relative to base, although aggregate corporate liquidations increase because of the increase in interest rates and therefore gearing. However, this scenario also hurts the household sector through the shift in the terms of trade and the rise in interest rates. Consequently, mortgage arrears increase substantially.

---

[1] This section draws on Hoggarth and Whitley (2003).

[2] Some banks could not provide quantitative estimates beyond a one-year horizon.

The error variances from the equations in the Bank of England's MTMM were used in order to calibrate the initial shocks. The equations were estimated from 1987, so the conditional variances include the early 1990s recession. But this approach could not be applied for the shocks to the exchange rate and equity prices.[3] In these two cases, historical variances and peak-to-trough estimates were used.

In choosing the threshold probability for the shock to be regarded as a scenario worthy of analysis, a balance needs to be struck. On the one hand, if the probability were set too high - and thus the size of shocks too low - there would be little impact. Nothing would be learnt about how the banking system would fare in a period of stress. On the other hand, if the size of shocks were extremely large, there would be almost no possibility of the event occurring. The size of the events chosen broadly corresponds to an event three standard deviations away from the mean.[4]

All the scenarios were estimated relative to a base case that was broadly consistent with the central outlook underlying the Bank's Inflation Report for November 2001. The impact of the shocks was estimated over a 12-month period (2002 Q2 to 2003 Q1) to provide an internally consistent set of outcomes for key macroeconomic variables, as well as for components of corporate and household sector balance sheets. The alternative scenarios also assumed that UK monetary policy (interest rates) reacted to the shocks according to a Taylor rule, which sets interest rates as a function of inflation and the output gap.[5] The assumed policy responses were intended to be broadly consistent with an inflation targeting monetary policy regime (but they should not be interpreted as indicating how the Bank of England's Monetary Policy Committee would respond in practice). This assumption played an important role in the scenarios in stabilising some of the macroeconomic responses to the events.

Results

Bottom-up approach

Panel (i) in Table 1 shows the overall impact of the four scenarios on the UK-owned banks' P&L account, while Graph 1 shows details of the effects on individual banks.[6] Panels (ii) to (iv) in Table 1 show the impact of the scenarios as a percentage of the banks' annual operating profits (averaged over the previous three years), risk-weighted assets and Tier 1 capital, respectively.

Overall, the effects on UK banks were estimated to be quite small in all the scenarios. Aggregating across the major UK-owned banks, the adverse impact on profits varies from an average in scenario 1 (fall in world equity prices) of £432 million (23% of annual profits) to £146 million (6% of profits) in scenario 3 (rise in wage pressure). Looking at individual banks, only one was estimated to have suffered a loss of more than 50% of average annual profits (over the past three years) or 10% of Tier 1 capital. This happened in the first scenario (panels (b) and (c) in Graph 1): the marked fall in equity prices reduces profits in a range of activities - loans and trading income, and, in some cases, income on asset fund management and insurance business. Overall, the results suggest that under all scenarios the major UK banks would have a sufficient cushion in profits to absorb the shocks without depleting their capital. The size of the impacts (after allowing for tax) is also small in relation to UK-owned banks' risk-weighted assets - the biggest adverse impact, under scenario 1, is in the range of 0.12 to 0.56% of risk-weighted assets (1.5 to 10% of Tier 1 capital).

---

[3] Although the macroeconomic model has rules of thumb for the determination of equity prices and the exchange rate, the equations do not have standard error distributions.

[4] Assuming a normal distribution, multiplying the standard deviation of the variable by 2.8 would imply a 5 in 1,000 occurrence (ie 99.5% confidence level) - suggesting an extreme but still plausible event. However, applying a normal distribution will understate the likelihood of extreme events if the tails of the distribution are fat.

[5] See Taylor (1993).

[6] The impact of the scenarios on the foreign-owned institutions are not reported since they only cover a part of their business and are therefore not estimated on a comparable basis.

Table 1

Impact of stress scenarios performed by major UK-owned banks on profits[1, 2]

(i) In millions of pounds sterling

	Scenario 1	**Scenario 2**	**Scenario 3**	**Scenario 4**
Mean	–432	–252	–146	–214
Median	–408	–195	–57	–81
Standard deviation	305	219	270	359

(ii) As a percentage of banks' annual pre-tax profits[3]

	Scenario 1	**Scenario 2**	**Scenario 3**	**Scenario 4**
Mean	–22.7	–15.0	–6.3	–1.8
Median	–18.4	–8.1	–6.1	–3.4
Standard deviation	21.2	18.1	8.3	18.4

(iii) As a percentage of (end-2001) risk-weighted assets

	Scenario 1	**Scenario 2**	**Scenario 3**	**Scenario 4**
Mean	–0.2	–0.2	–0.1	–0.1
Median	–0.2	–0.1	–0.1	–0.1
Standard deviation	0.2	0.1	0.1	0.2

(iv) As a percentage of (end-2001) Tier 1 capital

	Scenario 1	**Scenario 2**	**Scenario 3**	**Scenario 4**
Mean	–4.9	–2.9	–1.5	–1.5
Median	–4.4	–2.8	–1.2	–0.9
Standard deviation	3.3	2.2	2.1	3.6

[1] Negative implies stress test reduces profits, positive implies an increase in profits (relative to base). [2] On a group basis other than HSBC which relates to HSBC Bank. [3] Measured, on average, over previous three years.

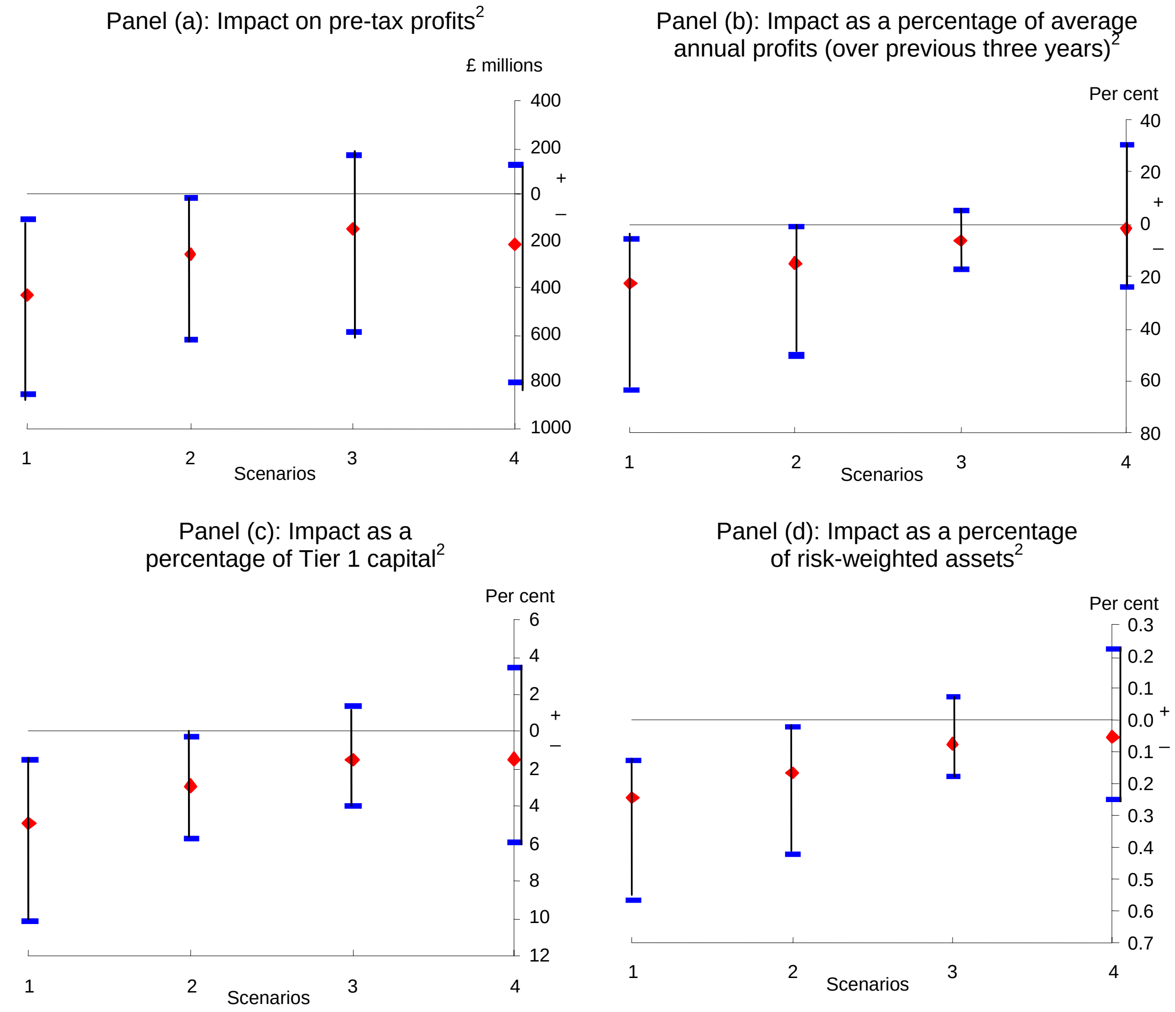
Source: Major UK-owned banks.

Aggregate top-down approach

As a complement to the stress test results provided by the large banks, as part of the FSAP we also estimated the effects on the provisions made against aggregate credit losses by the major UK-owned commercial banks measured on a consolidated basis using a single equation econometric model. These top-down simulations compared the model-based predictions for banks' new provisions charged against profits under each scenario relative to a base case.

Graph 1

Impact of stress scenarios[1] on UK-owned banks - bottom-up approach



[1] For any given scenario the rank ordering of banks varies across the four measures shown above. [2] The blue line represents the range across individual banks, the pink diamond shows the mean.

Source: Major UK-owned banks.

The econometric model for banks' provisions is a reduced form showing the relationship between key macroeconomic (and bank-specific) variables and banks' new provisions on their total loan book (see Pain (2003) for a further explanation). An advantage of this top-down approach is that the impact of the scenarios can be estimated beyond the one-year horizon.[7]

One of the preferred equations estimated using a small panel dataset on the UK bank is

$$\ln \frac{prF_{it}}{1-prF_{it}} = -6.3 - 0.07\Delta gdp_t - 0.08\Delta wgdp_t + 0.09\Delta RR_{t-1} + 0.04\Delta M4L_{it-3} + 0.04 propsh_{it-1} + 3.3 herf_{it-1} \qquad (1)$$

$$\bar{R}^2 = 0.75$$

[7] However, a potential disadvantage of this approach is that it is based on the average historical relationships rather than on the impact on banks' current loan portfolios.

where:

- *prF* is the new provisions charge against profits relative to loans and advances
- Δgdp is annual growth in real GDP
- $\Delta wgdp$ is annual growth in world real GDP
- ΔRR is a measure of ex post real interest rates based on base rates and the GDP deflator
- $\Delta M4L$ is the annual growth in M4 lending
- *propsh* is the share of total (sterling) lending to domestic commercial property companies
- *herf* is the Herfindahl measure of concentration of the domestic (sterling) loan portfolio
- Δgdp is significant at the 5% level, all other variables significant at the 1% level

Using the equation, the impact of a shock was calculated as the difference between the "shocked" value and a base case.

Table 2 summarises the average impact on provisions for the top-down simulations for those UK-owned commercial banks that also provided individual bottom-up estimates for the effects on provisions.

As in the case of the bottom-up approach, the largest effect on UK banks' provisions occurs in scenario 1: the 35% fall in world equity prices. Under this scenario, reductions in two of the key macroeconomic variables in equation (1) - UK and world GDP growth - increase the new provisions charge, more than offsetting the impact of lower real interest rates.

Overall, the top-down simulations also suggest that the likely increases in credit losses arising under all scenarios are quite small - all scenarios would result in an increase in banks' new provisions charges, both in the first year and cumulatively after three years, of less than £200 million on average (less than 10% of annual profits or 2% of Tier 1 capital).

4. Sectoral top-down approach

One drawback with the top-down approach used in the FSAP is that provisions are only available on UK banks' ***total*** loan book. Actual write-offs (losses) on loans to UK residents, on the other hand, are available at a (broad) sectoral level on a quarterly basis back to the early 1990s (Graph 2).[8] These more disaggregated data can be used to assess the impact of adverse shocks on different components of banks' loan portfolios. Bank write-offs relate to the losses (net of recoveries) made by UK-owned banks on loans initiated from their UK-resident banking operations.[9]

Two approaches have been adopted to stress testing banks' sectoral write-offs: (1) we have integrated sectoral write-offs with a version of the Bank's extended Medium-Term Macroeconometric Model (see Benito et al (2001) for details of the latter); and (2) we have included sectoral write-offs in a small VAR model.

8 Quarterly data at a sectoral level (households, corporates, etc) are not reported for all banks. For banks that only report annual sectoral data, the quarterly data have been derived by applying the annual sectoral shares to the aggregate quarterly data.

9 Therefore, the data exclude losses made by overseas branches and subsidiaries of UK-owned banks and losses made by domestically located non-bank businesses.

Table 2

Potential impact of stress test scenarios on UK commercial banks' provisions charge against profit[1]

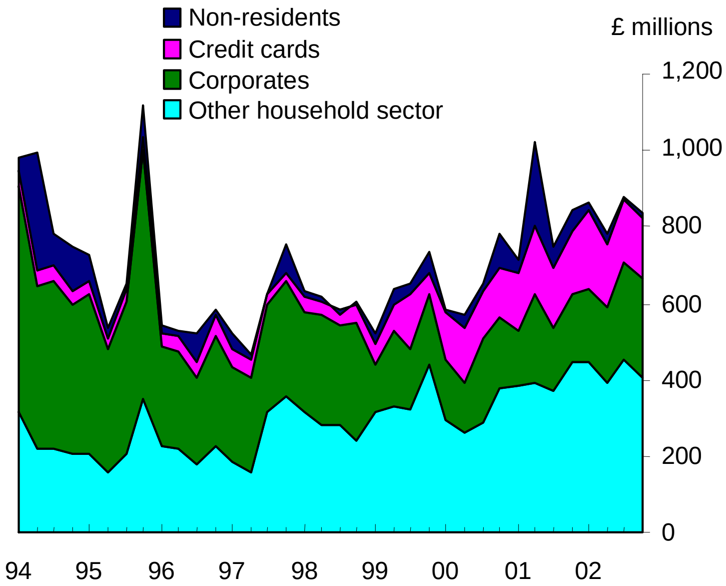
	Scenario 1			Scenario 2			Scenario 3			Scenario 4		
	£m	**% of profits**[2]	**% of Tier 1 capital**[3]	**£m**	**% of profits**[2]	**% of Tier 1 capital**[3]	**£m**	**% of profits**[2]	**% of Tier 1 capital**[3]	**£m**	**% of profits**[2]	**% of Tier 1 capital**[3]
First year												
Mean	–172	–5.7	–1.6	–47	–1.6	–0.4	–4	–0.1	0.0	–31	1.0	0.3
Median	–182	–6.1	–1.6	–50	–1.7	–0.4	–4	–0.1	0.0	–32	1.1	0.3
Standard deviation	39	0.8	0.3	11	0.2	–0.1	1	0.0	0.0	7	0.1	0.0
After three years[4]												
Mean	–130	–4.3	–1.2	–3	–0.1	0.0	–53	–1.8	–0.5	–110	–3.7	–1.0
Median	–138	–4.6	–1.2	–4	–0.1	0.0	–56	–1.9	–0.5	–116	–3.9	–1.0
Standard deviation	29	0.6	0.2	6	0.0	0.0	12	0.2	–0.1	25	0.5	0.2

[1] A negative sign means a decrease in profits, a positive sign an increase in profits. Banks were Barclays, Lloyds TSB, HSBC and Royal Bank of Scotland. [2] Percentage of previous three years' annual profits. [3] End-2001 Tier 1 capital. [4] Cumulative impact. Assumes that the key macroeconomic variables return to base by 2004 Q4.

Source: Bank of England calculations.

Graph 2

UK-owned banks' write-offs[1]

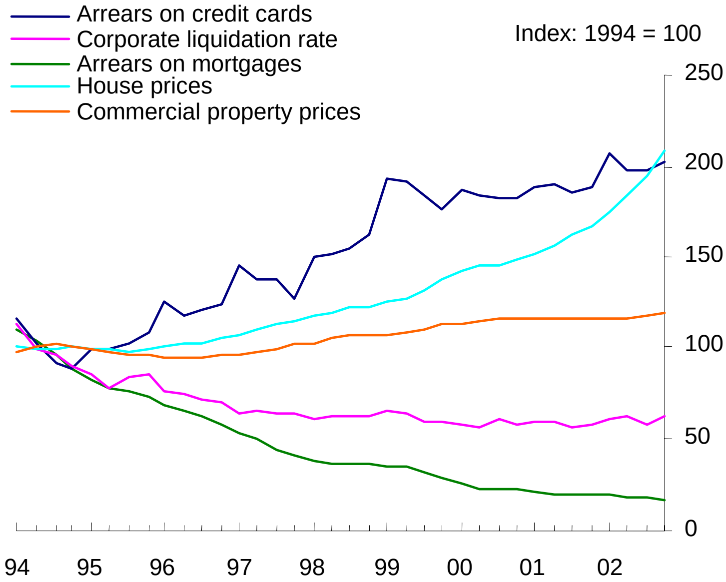


[1] Corporates include both financial and non-financial companies. Other household sector includes unincorporated businesses and non-profit organisations.

Source: Bank of England.

Graph 3

Sectoral delinquency rates and asset prices



Source: Bank of England.

A Extending the Medium-Term Macroeconometric Model (MTMM) for sectoral write-offs

The aim here is to extend the Bank's MTMM to include equations for sectoral write-offs. Only variables that are currently available in the MTMM are used to ensure that the impact of any initial shock can be traced through using an internally consistent scenario.

Bank $\text{losses}_i = p_i{*}\text{lgd}_i{*}\text{loans}_i$

where i refers to the sector, p is the probability of default and lgd is the percentage written off given default (ie 1 minus the recovery rate). Rearranging then

bank $\text{losses}_i/\text{loans}_i \equiv \text{write-off rate}_i = p_i{*}\text{lgd}_i$

Actual sectoral defaults or credit deteriorations are used to proxy p_i. There are no UK data on lgd/recovery rates, so we use variables that are likely to ***affect*** the recovery rate, in particular sectoral asset values. So the modelling strategy is:

write-off $\text{rate}_i = f(\text{default proxy}_i, \text{recovery rate proxy}_i)$

In the ***corporate*** sector, default is proxied by the corporate liquidation rate (the number of insolvencies in the period/number of registered firms). In turn, in the MTMM the corporate liquidation rate depends positively on corporate income gearing, changes in real interest rates and changes in net corporate debt/GDP and negatively on the growth in UK output and commercial property prices. The recovery rate is proxied by commercial property prices.

For the ***household*** sector, the proportion of credit card debt in arrears is used as the default proxy in the equation for credit card write-offs. The recovery rate is assumed to be zero. Credit card arrears, in turn, depend on household income gearing and the number of active credit card balances. As discussed in Cox et al (2004), the latter is used as a proxy for supply side influences such as UK banks' recent move down the credit quality spectrum, the adoption of more aggressive marketing techniques and generally the increase of competition in the UK credit card market during the past decade.

There is no further breakdown of household write-offs by loan type available on a consistent basis back to the first half of the 1990s, implying that non-credit card household write-offs ("other household sector") include write-offs on both secured debt (ie housing loans) and unsecured consumer debt (other than credit cards). Therefore, both mortgage and consumer credit arrears are included in the equation for other household sector write-offs to capture the likelihood of default. In the MTMM, in turn, mortgage arrears depend positively on mortgage income gearing and unemployment and negatively on undrawn housing equity and the loan-to-value (LTV) ratio of first-time buyers (as a proxy for the credit risk of new borrowers).[10]

House prices were included in the initial specification for other household sector write-offs to capture the impact of changes in loss-given-default on mortgage debt but were not found to be statistically significant. This may be attributable to house prices and mortgage arrears being dependent on the same factors. So that in periods when mortgage defaults decline, house prices increase. As seen from Graph 3, mortgage arrears have been on a steep downward trend since the early 1990s while over the same period house prices have been on a steep upward trend.[11] Therefore, the impact of mortgage arrears on other household write-offs may not only be capturing the impact of changes in default but also changes in loss-given-default.

Results

The equations linking variables of sectoral fragility to bank write-offs are shown in Table 3. All variables enter contemporaneously, other than credit card arrears, which have a four-quarter lag. This suggests that as households become fragile, they first delay paying their consumer debt and only later

---

[10] Cox et al (2004) argue that banks undertake high LTV mortgage lending with customers they judge to be of high credit quality.

[11] The simple correlation coefficient between house prices and mortgage arrears over the period is –0.81.

their mortgage debt.[12] These simple equations seem to explain past movements in bank write-offs quite well, especially on corporate and other household loans (panels (a) to (c) in Graph 4). The equations capture the steady decline in corporate write-off rates throughout the past decade, the gentler decline in other household write-offs and, to some extent, the initial decline and then rise over the past five years in credit card write-offs.[13]

Table 3

Sectoral write-off rate linking equations, 1994 Q1 to 2002 Q4

Explanatory variables	Corporate sector	Household sector	
		Credit cards	Other
Corporate liquidation rate$_t$	1.275 (0.00)		
Commercial property prices$_t$	–0.002 (0.00)		
Mortgage arrears$_t$			0.038 (0.00)
Credit card arrears$_{t-4}$		1.133 (0.00)	0.107 (0.00)
1995 Q4 dummy	0.207 (0.00)		0.075 (0.00)
R-bar squared	0.94	0.59	0.80
DW	1.5	1.4	2.0
Number of observations	36	36	36

Note: Corporates include both non-financial and financial companies. Other household sector consists of secured household, unsecured household (other than credit cards), unincorporated businesses and non-profit organisations. Corporate liquidation rate is the number of corporate insolvencies as a percentage of the number of registered companies. Mortgage arrears are the number of mortgage arrears more than six months as a percentage of the number of mortgages outstanding. Credit card arrears are the value of credit card balances in arrears by more than three months as a percentage of the value of all credit card balances.

p-values in parenthesis. All variables are significant at the 1% level.

The sectoral linking equations can only be estimated from 1993, since when sectoral write-off data have been available. However, the equations explaining the default proxies are estimated back to the late 1980s. This implies that the scenarios for sectoral defaults, at least, are based on relationships that include the last boom and bust in the United Kingdom in the late 1980s/early 1990s.

We then repeated the four scenarios used in the FSAP and traced through the impact on banks' sectoral write-offs. The results are shown in Table 4 below.

As seen in Table 4, again the impact on banks' balance sheets is estimated to be quite small. None of the scenarios results in write-offs increasing (relative to base) in the first year or cumulatively after three years by more than 2% of the banking system's Tier 1 capital. However, there are differences across the scenarios. Since income gearing is an important determinant of sectoral default, particularly for the household sector, the assumed interest rate response has an important impact on write-offs in the simulations.

[12] It may also partly reflect differences in the definition of when a late payment is categorised as an arrear. For mortgages the variable is measured as arrears of more than six months, and for credit cards it is arrears of three months or more.

[13] However, credit card arrears seem to overstate credit card write-offs somewhat in 1997-98 and understate them in 2001.

Graph 4

Sectoral write-off rate linking equations -
actual vs fitted

[1] Financial and non-financial companies. [2] Non-credit card household plus unincorporated businesses and non-profit making organisations.

Source: Bank of England.

Monetary policy is assumed to ease in response to the sharp fall in equity and property prices (scenarios 1 and 2 respectively). The consequent fall in household income gearing implies that the net effect is to ***reduce*** household sector write-offs albeit slightly. In scenario 2, although mortgage arrears (and thus implicitly mortgage write-offs) rise relative to base, this is more than offset by an implied reduction in (non-credit card) unsecured write-offs due to the fall in household income gearing. However, corporate sector write-offs increase in both these scenarios despite a decline in corporate income gearing. This is partly attributable to the initial fall in output growth (relative to base). Also in scenario 2, the large fall in commercial property prices increases both corporate liquidations and loss-given-default.

In contrast, scenarios 3 and 4 - an increase in earnings growth and a depreciation of sterling respectively - lead to higher inflation, which is met by a tightening of monetary policy. Under both scenarios, there is a rise in households' income gearing - interest payments increase and disposable incomes fall. The impact of sterling depreciation (scenario 4) on the fragility of the corporate sector is partially offset by an increase in export volumes and output (relative to base) over the simulation period. Consequently, in this scenario the write-off rate for companies rises by less than for households.

Table 4

Impact of stress test scenarios on UK banks' sectoral write-offs[1]

(a) First year

Sector	Scenario 1		Scenario 2		Scenario 3		Scenario 4	
	£m	**% of Tier 1 capital**	**£m**	**% of Tier 1 capital**	**£m**	**% of Tier 1 capital**	**£m**	**% of Tier 1 capital**
Corporates	115	0.1	545	0.5	40	0.0	125	0.1
Credit cards	0	0.0	0	0.0	0	0.0	0	0
Other household sector	–15	0.0	20	0.0	0	0.0	90	0.1
Total	100	0.1	565	0.5	40	0.0	215	0.2

(b) After three years[2]

Sector	Scenario 1		Scenario 2		Scenario 3		Scenario 4	
	£m	**% of Tier 1 capital**	**£m**	**% of Tier 1 capital**	**£m**	**% of Tier 1 capital**	**£m**	**% of Tier 1 capital**
Corporates	270	0.3	1,845	1.8	930	0.9	470	0.5
Credit cards	–105	–0.1	–70	–0.1	115	0.1	250	0.2
Other household sector	–350	–0.4	–50	–0.1	465	0.5	1,050	1.0
Total	–185	–0.2	1,725	1.6	1,510	1.5	1,770	1.7

[1] A positive sign implies an increase in write-offs, a negative sign a reduction in write-offs compared with the base case.
[2] Cumulative impact. Assumes that the key macroeconomic variables return to base by 2004 Q4.

B VAR approach

We also adopted another approach to derive the scenarios and to apply the shocks directly to UK banks' actual losses (write-offs). We produced a vector autoregressive (VAR) model consisting of a limited number of macroeconomic variables and bank write-offs.

The choice of macroeconomic variables included in the VAR was motivated by the existing literature on reduced-form macro models, for example Blake and Westaway (1996), Ball (1998) and Batini and Haldane (1999). So the VAR consisted of UK output (relative to a simple trend), nominal short-term interest rate, the real exchange rate, the annual RPIX inflation rate and banks' write-off rate (net write-offs divided by the value of loans outstanding).

Since quarterly data on bank write-offs are available only from 1993 Q1, the data period covers only the recovery phase of the early 1990s economic cycle. It also implies that some of our variables show little variation over the period - in particular retail price inflation and the banks' base rate, which have remained in a relatively narrow range of between 1.75 and 3.5% per annum and between 4 and 7.5% respectively over the past decade. We experimented with including house price inflation in the VAR since it shows more movement over the past decade and might be expected to affect bank write-offs. As a check on our results, we also used annual data on the main UK banks' consolidated published accounts to derive aggregate banking system data back to 1988 (ie to capture the economic downturn). Our data are spliced in 1993 Q1, and the annual data before 1993 are interpolated onto a quarterly basis.

We tested for stationarity using the augmented Dickey-Fuller (ADF) test. Though the tests were not always able to reject a unit root at the 10% level, the *p*-values were never far from 10%. Given that it is well known that the ADF test suffers from low power and we expect that the series should be mean-reverting, we treat them as such.

In order to ensure that the shocks are uncorrelated, we applied a Cholesky decomposition (with a degrees-of-freedom correction). The variables in the model were ordered in ascendance according to the likely speed of reaction to any particular shock. Variables at the front end of the VAR are assumed to affect the following variables contemporaneously but only to be affected themselves by shocks to the other variables after a lag. Variables at the bottom of the VAR, on the other hand, only affect the preceding variables after a lag but are affected themselves immediately. The financial variables - interest rates and the exchange rate[14] - were ordered at the bottom of the VAR, implying that they react instantaneously to shocks in the real-side variables, whereas the other variables react only after a lag following shocks to the financial variables. Output was ordered after write-offs, reflecting priors that the economic cycle affects bank losses in the United Kingdom only after a lag (Hoggarth and Pain (2002)).

In principle, inference in VAR models is sensitive to the choice of lag length based on the different information criteria and appropriate lag length can be critical. If a large number of lags is included, degrees of freedom are eroded. If the lag length is too small, important lag dependencies may be omitted. We used both the Akaike and the Schwarz information criteria to set the lag length equal to 2 for all the various specifications reported below.

Results

Using post-1993 data, none of the shocks had a statistically significant impact at the 95% confidence level on write-offs either in the basic aggregate VAR or where house prices are included. As mentioned above, this might reflect a lack of variation in a number of the variables. However, once the estimation period is taken back to 1988, then some shocks have a statistically significant impact. In particular, shocks to output always had a negative and statistically significant impact on write-offs.[15]

In the sectoral VAR for private non-financial companies (PNFCs) we also included PNFCs' income and capital gearing,[16] in addition to the macroeconomic variables discussed above, since, as discussed earlier, there is evidence that these types of financial variables also affect corporate liquidations in the United Kingdom. The maximum impact seems to occur more quickly than suggested by the VAR including aggregate write-offs - after nine months for changes in output (relative to trend) and six months for changes in interest rates.

But in both the aggregate and the corporate VARs, the economic impact was quite modest - the impact of a 1% adverse shock to output on write-offs never exceeded 2% of Tier 1 capital.

---

[14] Although the real exchange rate is included in the VAR, short-term movements are driven by the nominal exchange rate.

[15] We also experimented with including world output in the VAR, since it was found important by Pain (2003) in affecting UK bank provisions. But this variable did not have a significant impact on write-offs. One reason that may explain the different result is that provisions data relate to the consolidated entity, including overseas branches and subsidiaries, whereas the (post-1993) write-off data relate only to the UK-based operations. The latter are likely to be less affected by adverse shocks abroad.

[16] Income gearing is defined as interest payments as a percentage of PNFC pre-tax profit and capital gearing is PNFCs' net debt as a percentage of net debt plus net equity.

5. Why do the stress tests not have a bigger impact on UK banks' balance sheets?

One factor helping to explain the small size of the effects is the higher quality of UK banks' loan books than in the late 1980s. Over the past decade, there has been a widespread decline in the ratio of "risk-weighted" assets (used by regulators to calculate capital requirements) to total assets and an increase in geographical asset diversification. Also, aggregate sectoral data on domestic loans suggest that the composition of the large UK-owned banks' retail loan book has shifted away from riskier unsecured lending to relatively safer mortgage lending over the past decade.[17] And within the mortgage market, loan-to-value ratios (LTVs) are now much lower than in the late 1980s. For example, the proportion of UK banks' new mortgages with LTVs over 90% has fallen since the mid-1990s, from almost 50% to below 30%.[18] Consequently, it would probably take a marked decline in house prices to cause a significant increase in losses on housing loans. UK banks' corporate loan portfolios also appear to be of a relatively high quality. Estimates indicate that almost half of major UK banks' corporate exposures have internal ratings equivalent to A or above.

Second, the impact of the scenarios used in the "bottom-up" approach in the FSAP was estimated only over a one-year horizon. In practice, it takes longer than one year for the full impact of the shock to work through. Some of the defaults caused by an overall credit deterioration will not occur until later years. One bank extended the simulations beyond the one-year horizon. This analysis suggested that its provisions for retail credit losses could be on average six times higher in the second year than in the first. And, as a rough ready reckoner, another bank suggested that the peak effect on retail provisions was around three times the first-year effect and was likely to occur three years after the initial shock.

Also, in the MTMM scenarios at least, the policy reaction tempers the impact of two of the shocks (scenarios 1 and 2). Monetary policy is assumed to adjust partly to *offset* declines in output as well as rises in inflation (given the Taylor reaction function). So, for example, the decline in house prices is followed by a reduction in interest rates that moderates the impact on output, and thus on corporate liquidations and housing arrears. The large losses that UK banks incurred following asset price deflation in the early 1990s were accompanied by a sharp *increase* in nominal interest rates, and hence income gearing. In consequence, output fell substantially and liquidations and arrears rose sharply.

The analysis also ignores how banks and their creditors, including other banks, would react faced with a weakened bank. Although individual bank actions might be designed to reduce potential losses, the collective results might intensify economic stress - through a credit crunch, for example - and weaken banks' positions further. In extremis, if the shock were big enough to cause the failure of a large bank, this might have a direct impact on the capital, or even solvency, of other (counterparty) banks.[19] Wells (2002) provides the back end of this analysis through estimating the impact via the interbank market of a single bank failure on other banks. But this analysis assumes implicitly that the initial shock is specific to a particular bank.

It might also be the case that in order to maintain a high credit rating and to have access to interbank funding, the large UK banks hold capital in case of more extreme events than are considered here (Jackson et al (2002)).

6. Extensions and future work

The above top-down analysis focuses on the impact of adverse macroeconomic scenarios on the UK banking system as a whole. One planned extension is to compare the impact across the major UK banks at a bank by bank level. The size of the impact on any individual bank will depend on both

---

17 These changes reflect the impact of demutualisation as well as shifts in banks' portfolios.

18 See Bank of England (2002), Part III.

19 See Elsinger et al (2002) and Wells (2002).

the composition and quality of its portfolio (Box (4) in Figure 1) and the amount of capital it has to withstand the shock (Box (5)). An important aspect of the latter will be to assess the threshold beyond which a decline in capital would be likely to result in a bank "failure". This top-down analysis could also be bolted onto the interlinkages work to estimate the second-round effects of a bank failure on other banks (step (6) in Figure 1). The impact on sectoral losses discussed above focused on loans to UK residents. This analysis could be extended to include loans to non-residents.

The above approach has concentrated on accounting measures of bank losses. We also plan to complement this work through estimating the impact of adverse macroeconomic shocks on financial market measures of credit losses. This analysis will involve first generating macroeconomic scenarios either from the Bank of England's macroeconomic model or from a more parsimonious VAR model. The macroeconomic variables from this first stage will be included together with industry-specific (and firm-specific) variables in a model to explain firm equity returns.[20] The forecast equity returns will then be plugged into a Merton model to provide estimates of the conditional probability of default for each firm. The final stage will be to use information on loss-given-default and the pattern of banks' corporate exposures to generate projected bank-specific losses for different adverse macroeconomic scenarios.

7. Conclusions

We have carried out a range of stress tests on the UK banking system using a number of approaches building upon the analysis carried out as part of the UK FSAP. The estimated potential losses in no case exceeded annual profits or represented a large fraction of banks' capital. However, some caution needs to be exercised with these results.

The results are likely to be sensitive to the nature and specification of the macroeconomic stress tests. The size of the shocks is based largely on historical experience averaged over normal times and periods of stress, rather than taken from stress periods alone. The latter, by definition, occur infrequently and may be conditioned by the precise circumstances at the time. There may be sharp discontinuities in economic behaviour and relationships in crisis periods. The analysis also ignores how banks and their creditors, including other banks, would react faced with a weakened bank. Although individual bank actions might be designed to reduce potential losses, the collective results might intensify economic stress - through a credit crunch, for example - and weaken banks' positions further. It might also be the case that banks set capital as an insurance against more extreme events than have been considered here.

An important factor explaining the relatively modest impact of some of the scenarios derived from the MTMM on UK banks' profits is the assumed monetary policy reaction in response to a change in the outlook for inflation. Although the particular numerical results may depend on the precise specification of the interest rate reaction rule, to the extent that inflation targeting serves to stabilise some of the macroeconomic responses to unanticipated shocks, it will have beneficial implications for the stability of the UK financial system.

Overall, these estimates suggest that the stability of UK banks is unlikely to be threatened by a range of plausible adverse shocks, especially given that most UK banks are currently very profitable by international standards and have capital ratios well in excess of the regulatory minimum. Nonetheless, this exercise emphasises the importance for the authorities, and for banks themselves, of continuing to develop quantitative techniques which can be used to assess the resilience of the financial system to potential shocks.

[20] Pesaran et al (2003) adopt a similar approach. They use a global VAR in combination with an equity returns equation to produce estimates of defaults for 119 firms worldwide.

References

Altman, E, A Resti and A Sironi (2002): "The link between default and recovery rates: effects on the procyclicality of regulatory capital ratios", *BIS Working Papers*, no 113, July.

Ball, L (1998): "Policy rules for open economies", *NBER Working Papers*, no 6760.

Bank of England (1999): *Economic models at the Bank of England*.

——— (2002): "The financial stability conjuncture and outlook", *Financial Stability Review*, issue 13, December.

Basel Committee on Banking Supervision (2003): *Overview of The New Basel Accord*, Consultative Document, Bank for International Settlements, April.

Batini, N and A Haldane (1999): "Forward-looking rules for monetary policy", *Bank of England Working Paper Series*, no 91, January.

Benito, A, J Whitley and G Young (2001): "Analysing corporate and household sector balance sheets", *Bank of England Financial Stability Review*, December, pp 160-74.

Blake, A P and P J F Westaway (1996): "Credibility and the effectiveness of inflation targeting regimes", *The Manchester School Supplement*, pp 28-50.

Boss, M (2002): "A macroeconomic credit risk model for stress testing the Austrian credit portfolio", *Financial Stability Review* 4, Oesterreichische Nationalbank, October, pp 64-82.

Cox, P, J Whitley and R Windram (2004): "An empirical model of household arrears", *Bank of England Working Paper Series*, no 214, March.

Elsinger, H, A Lehar and M Summer (2002): "A new approach to assessing the risk of interbank loans", *Financial Stability Review* 3, Oesterreichische Nationalbank, July, pp 75-86.

Enders, W (1995): *Applied econometric time series*, John Wiley & Sons.

Hoggarth, G and D Pain (2002): "Bank provisioning: the UK experience", *Bank of England Financial Stability Review*, June, pp 116-27.

Hoggarth, G and J Whitley (2003): "Assessing the strength of UK banks through macroeconomic stress tests", *Bank of England Financial Stability Review*, June, pp 91-103.

International Monetary Fund (2003): *UK: financial system stability assessment*, February.

Jackson, P, W Perraudin and V Saporta (2002): "Regulatory and 'economic' solvency standards for internationally active banks", *Bank of England Working Paper Series*, no 161, August.

Nickell, P, W Perraudin and S Varotto (2001): "Stability of ratings transitions", *Bank of England Working Paper Series*, no 133, May.

Pain, D (2003): "The provisioning experience of the major UK banks: a small panel investigation", *Bank of England Working Paper Series*, no 177, January.

Pesaran, M H, T Schuermann, B-J Treutler and S Weiner (2003): "Macroeconomic dynamics and credit risk: a global perspective", *Wharton Financial Center Working Paper* 03-13.

Taylor, J (1993): "Discretion versus policy rules in practice", *Carnegie-Rochester Series on Public Policy*, vol 39, pp 195-214.

Wells, S (2002): "UK interbank exposures: systemic risk implications", *Bank of England Financial Stability Review*, December, pp 175-82.

Monetary and financial stability in Norway: what can we learn from macroeconomic stress tests?

Snorre Evjen, Arild J Lund, Kjersti Haare Morka,
Kjell B Nordal and Ingvild Svendsen[1]
Norges Bank

1. Introduction

Over the past few years, the discussion among academics and central bankers about the relationship between monetary and financial stability has intensified. The discussion has particularly focused on whether inflation targeting is consistent with financial stability, and if an inflation targeting regime contributes to financial stability. Furthermore, is there a conflict between monetary and financial stability, and if so, in what situations do such conflicts typically occur?[2]

The traditional view has been that a monetary policy regime preserving low and stable inflation tends to facilitate financial stability. Low and stable inflation provides households and enterprises with a clear indication of changes in relative prices, thereby making it easier for economic agents to make the correct decisions. Low and stable consumer price inflation also contributes to price stability in financial and property markets. An unexpected decline in inflation increases the real value of outstanding debt, making defaults more likely. Furthermore, the vulnerability of the financial system tends to rise when inflation is high, particularly if monetary policy needs to be tightened significantly to reduce inflation or restore economic stability. Hence, the traditional view has been that low and stable inflation provides a sound foundation for financial stability and that the two objectives normally underpin each other.

However, financial imbalances can build up in a low-inflation environment. This relates to the fact that high credibility in the policymakers' commitment to price stability, or stable inflation expectations, may enhance price rigidity at the mean level. As a result, overall inflation may be under control even in a macroeconomic environment with high and increasing demand, and where demand pressure results in higher asset prices and credit growth. The same may ensue from supply side developments putting downward pressure on prices.

It has therefore been argued that inflation targeters should more explicitly consider developments in financial variables such as equity and bond prices, credit and property prices when setting interest rates. Some argue that central banks' key interest rates should also respond to these variables in situations where inflation pressures seem to be under control.

Financial imbalances may build up in a low-inflation environment without threatening the inflation target in the short to medium term. However, these imbalances may be a threat to nominal stability in the somewhat longer run when a burst of the bubble could imply strong deflationary pressure and bring inflation below target. Consequently, it has been argued that monetary policy in some situations should adopt a somewhat longer policy horizon allowing inflation to undershoot the target for some time in order to dampen credit growth and the rise in asset prices and thus reduce the risk of a burst of the bubble which may threaten an even more substantial undershoot of the inflation target in the future. The build-up of financial imbalances may also constrain the use of monetary policy. High levels of debt and overvalued asset prices may prevent the central bank from taking adequate steps because of the risk of turmoil in the financial sector.

---

1 Arild J Lund and Kjell B Nordal work in the Financial Analysis and Structure Department, Norges Bank (Central Bank of Norway). Snorre Evjen, Kjersti Haare Morka and Ingvild Svendsen work in the Monetary Policy Department, Norges Bank. Questions concerning the paper can be addressed to: Arild-J.Lund@norges-bank.no or Ingvild.Svendsen@norges-bank.no. The views expressed are those of the authors, and do not necessarily represent those of Norges Bank. We would like to thank Eivind Bernhardsen, Kjersti-Gro Lindquist and conference participants for helpful comments and discussions.

2 See for instance Bernanke and Gertler (1999), Borio and Lowe (2002), Cecchetti et al (2000) and Borio et al (2003).

The costs to society associated with a crisis in the financial system can be large. But keeping inflation below the target for a certain period also involves costs. In some cases, substantial increases in interest rates may be required in order to curb the build-up of financial imbalances. Unemployment may rise, inflation expectations may fall below target and central bank credibility may be jeopardised. Furthermore, not all situations involving a build-up of financial imbalances result in financial crisis. We therefore need good indicators to show whether financial imbalances are emerging and the danger they impose on the macroeconomic balance. Some promising steps have been taken in this field.[3]

The aim of this paper is (1) to investigate the effects on financial institutions' losses of different monetary responses to supply and demand side shocks and discuss how stress tests may assist in monetary policymaking, and (2) to present the model used to conduct the stress tests.

The paper is organised as follows: we first discuss important characteristics of macroeconomic stress tests. In Section 3, we present the methodology, ie how the stress tests are implemented. In Section 4, the results of the stress tests are presented and discussed. The major findings are discussed in Section 5.

2. Macroeconomic stress tests of the Norwegian financial sector

In essence, a stress test is a what-if analysis. What-if analyses are undertaken to gain an insight into the mechanisms of the economy by analysing the effects of certain shocks to the economy. In this paper we focus on shocks in demand and wage growth and study the impact on banks' loan losses of different monetary policy responses. This may be viewed as a post-shock analysis. The results from these analyses are particularly relevant to monetary policymaking in an ex ante perspective if they give insight into how today's monetary policy decision influences the probability and nature of future instability in the financial sector.

Financial stability is often defined as the absence of financial instability. Financial instability is typically characterised by large and abrupt changes in property prices and securities markets and by financial institutions or financial markets that do not function adequately. Disturbances occur in the credit supply or in the flow of capital. In most cases, this will have consequences for output, employment and inflation.

Increases in banks' provisioning for bad debt may be used as an indicator of the degree of financial stability. This indicator typically summarises the financial situation for both households and enterprises and their implications for the financial sector. The macroeconomic environment is crucial for the debt servicing capacity of households and enterprises and for the level of prices of those assets which often serve as collateral. Macroeconomic shocks have an impact on these variables and hence on banks' loan loss provisioning.

We apply macroeconomic stress testing to illustrate the financial sector's robustness to adverse macroeconomic shocks and to analyse whether a monetary policy reaction to the same shocks will mitigate or amplify banks' credit losses. The stress test approach in Norges Bank is model based. Output from a macroeconomic model - the RIMINI model of Norges Bank[4] - is used as input when forecasting loss provisioning. Losses are forecast separately for the household sector and the corporate sector. For the corporate sector, a micro model, based on firms' accounts, is used. Combining predicted bankruptcy probabilities with information about each firm's bank loans and general property prices as a proxy for the value of the collateral enables us to compute expected bank losses at an aggregate level. The variation in risk structure across lenders is explicitly taken into account. For the household sector, we use a single loan loss function where loan losses depend on the initial debt to income ratio, the level of interest rates and the unemployment rate. The methodology is described in Section 3.

---

3 See Borio and Lowe (2002).

4 See Eitrheim and Gulbrandsen (2001) and Olsen and Wulfsberg (2001) for an overview of key aspects of the model.

We consider both a demand and a supply shock. The demand shock stems from a sudden drop in public spending, while a strong rise in wage costs is the source of the supply side shock. We model the shocks with and without a monetary policy response. For simplicity, we chose to model the monetary response by a standard Taylor rule.[5] According to the Taylor rule, the interest rate is set as a function of the neutral long-term rate of interest, excess production (the output gap) and excess inflation (the inflation gap). In a situation where inflation is on target and the output gap is zero, the Taylor rule interest rate will be equal to the neutral nominal interest rate (the neutral real interest rate plus the inflation target); see Taylor (1993). We have applied a backward-looking Taylor rule, which normally gives a somewhat slower monetary policy reaction than forward-looking rules.[6] The scenarios with a Taylor rule response are compared with scenarios without a monetary policy reaction (ie monetary policy as in the reference scenario). Few, if any, inflation targeting central banks follow a Taylor rule. However, a Taylor rule has in many cases proved to be useful as an empirical description of an inflation targeting regime.

Stress tests at two different points in time

The initial situation for enterprises and households is important. For the individual households and enterprises, their ability to service their loan is a result of both the general economic situation and individual characteristics.

We have stress-tested the economy at two different points in time, in 1996 and 2001. The purpose is to see how different economic conditions influence the impact of the shocks. In particular, we are interested in situations with different levels of indebtedness and different levels of asset prices. The vulnerability of the household sector to increases in the unemployment rate depends positively on the initial debt burden and how debt is dispersed among different groups of households. In general, a firm’s bankruptcy probability, given a drop in new orders, depends on operating income and expenses, own funds, debt structure and other individual characteristics.

In 1996, the macroeconomic environment in Norway was relatively balanced. Around three years of growth above trend had closed a negative output gap. Inflation seemed to be under control. Norwegian enterprises had gradually built up their capital reserves. The level of debt and asset prices was low. The banking crisis was over. We would have expected the financial system to be quite robust if faced with a negative shock to the economy.

The latest observation in the dataset of Norwegian enterprises’ accounts is 2001. The situation for enterprises and households that year is comparable to their financial situation today, although the macro fundamentals and corporate key variables have changed negatively from 2001 to 2003. In 2001, capacity utilisation in the Norwegian economy was very high after several years of high growth. The financial situation of the corporate sector was still very sound, but the indebtedness of firms had increased compared to 1996. Also the indebtedness of the household sector had increased, but to a lesser extent. House prices had risen considerably. See Table 1 for a summary of key variables in 1996 and 2001.

---

5 The Taylor-rule applied: $i_t = i^* + 1.5 \cdot (\pi_t - \pi^*) + 0.5 \cdot y_t$. y_t is output gap at time t, i is the nominal interest rate, π is inflation, π^* is the inflation target and i^* the neutral real interest rate.

6 It should be noted that assuming a Taylor rule is completely different from actual monetary policy in Norway in the 1990s. In the 1990s monetary policy aimed at stabilising the exchange rate. Since March 2001, the government has defined an inflation target for monetary policy in Norway. The operational objective is an inflation rate of 2½% over time. See www.norges-bank.no for more information about Norwegian monetary policy.

Table 1

Summary of key variables describing the state of the Norwegian economy and the corporate and financial sectors in 1996 and 2001

Per cent

	1996	2001
Macroeconomics		
GDP growth (mainland economy)	4.2	1.7
Output gap	0	2
Unemployment rate (registered)	4.2	2.7
Annual wage growth	4.4	5.8
CPI inflation	1.2	3.0
Households		
Credit growth	4.8	10.4
House prices, annual rise	9.1	7.3
Interest rate on loans	7.2	8.9
Annual real disposable income growth	3.4	0.5
Saving ratio	2.3	4.0
Enterprises[1]		
Return on capital	9.1	6.3
Return on equity	20.0	7.3
Interest paid/debt	4.3	5.8
Equity ratio	30.7	34.3
Growth in bank debt	1.3	2.9
Banks[2]		
Return on equity	17.5	12.0
Non-performing loans/gross loans	3.0	1.3
Equity/total capital	6.5	6.5
Tier 1 and 2 capital/risk-weighted assets	12.9	12.6
Tier 1 capital/risk-weighted assets	9.9	9.7

[1] Information based on accounts for all joint stock companies. [2] The numbers apply to Norwegian banks. Norwegian banks' branches abroad are not included.

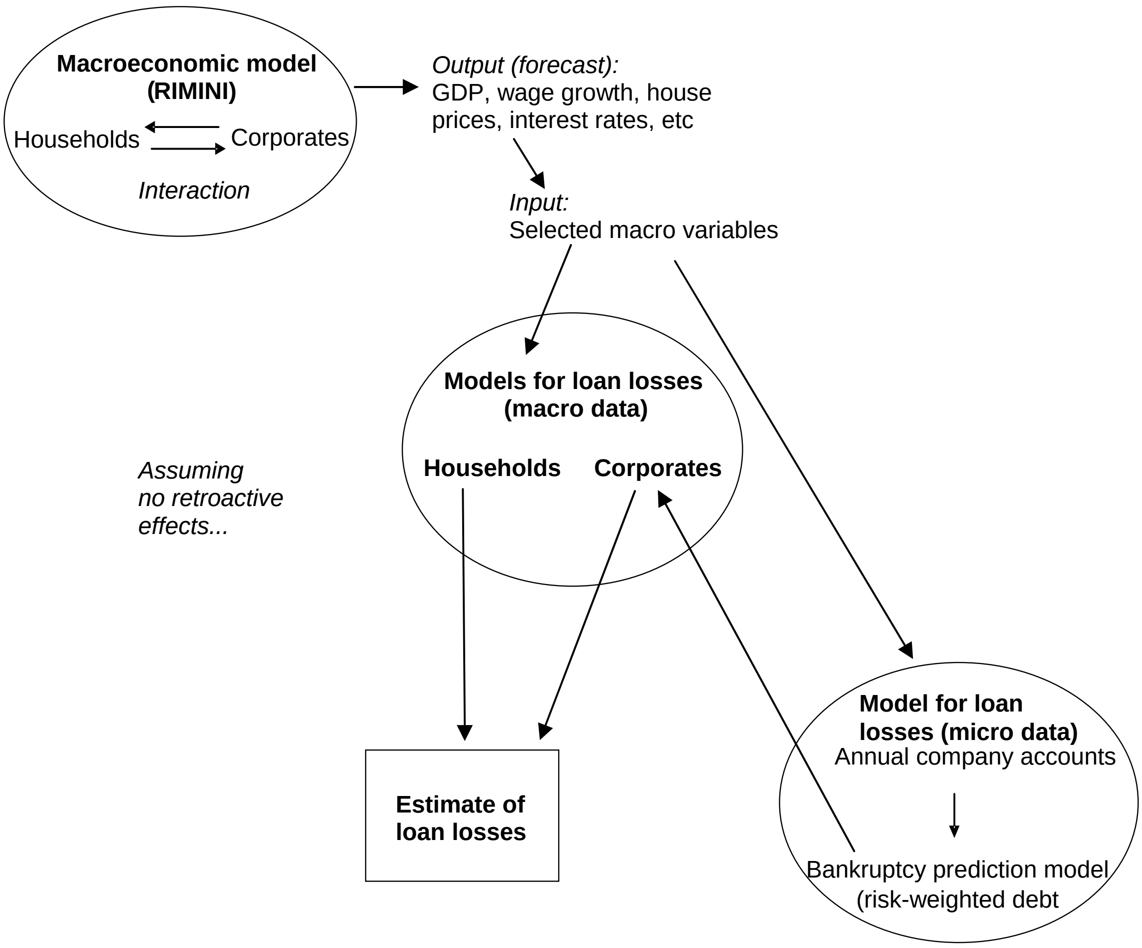
Sources: Norges Bank; Statistics Norway.

3. The methodology used to estimate loan losses

Estimation of losses on loans to both the household and corporate sectors is based on macroeconomic variables such as GDP growth, wage growth, interest rates and changes in house prices; see Figure 1. The macroeconomic variables reflect the interaction between firms and households as both sectors are included in the macroeconomic model RIMINI. There is, however, no feedback from estimated bank losses to the macroeconomic scenario.

Figure 1

Outline of methodology used to calculate loan losses



For a lender, the *expected* loss on a portfolio of loans is the product of the probability of default or bankruptcy, the borrower's outstanding debt and the level of loss in the event of default or bankruptcy.[7] The probability of bankruptcy, debt and loss-given-default is a function of both macroeconomic developments and microeconomic conditions associated with the individual borrower. To analyse loan losses, all these factors should be assessed.

Losses are estimated separately for the household and corporate sectors. These sectors have specific risk characteristics and they are treated as different segments by financial institutions.

[7] Expected loss is computed as $\sum_{i=1}^{n} p_t^{(i)} D_t^{(i)} LGD_t^{(i)}$, where $p_t^{(i)}$ is the probability of borrower *i* defaulting or going bankrupt, $D_t^{(i)}$ is borrower *i*'s debt and $LGD_t^{(i)}$ is the level of loss-given-default or bankruptcy at a point in time, *t*. By aggregating the figures for all borrowers, we obtain an estimate of the overall expected loan loss in the economy.

Household sector

The model for financial institutions' provisioning for bad debt in the household sector is solely based on macroeconomic variables. The equation for losses in per cent of outstanding debt, *LOSSREL*, is given by

$$lossrel_t = 3.31 dburd_t - 1.45 rhous_t + 13.55 R_t + 31.55 UMP_t - 7.05 DUM97_t \qquad (1)$$

where *dburd* is the debt burden measured as debt in per cent of disposable income, *rhous* is the real value of private houses, *R* is the interest rate, and *UMP* is the unemployment rate. The use of lower case letters indicates that the variables are in logarithmic form. Equation (1) is estimated on actual losses for the time period 1978-2001. For the model summary, see Appendix 1.

The partial effects of the variables on provisioning are intuitive. An increase in the debt burden, higher unemployment and higher interest rates increase financial institutions' losses. Losses may also increase as a result of reduced values of private houses, which result in lower values of collateral.

This analysis does not reflect the fact that households are a heterogeneous group. Debt burden, for example, varies widely across income deciles in the household sector and has developed differently over time. This implies that changes in interest rates may have a very different effect on households in different income deciles. In a more micro-based approach, financial institutions' loan losses could be modelled for the various income categories in the household sector.

Corporate sector

The provisioning for bad debt in the corporate sector is modelled according to the equation

$$loss_t = 0.95 rwd_{t-1} - 13.34 \Delta rph_t \,, \qquad (2)$$

where *LOSS* is financial institutions' losses on loans to enterprises, *RWD* is the sum of risk-weighted debt for all enterprises and *RPH*[8] is the real price of existing dwellings.[9] The collateral pledged by enterprises to lenders consists mainly of real estate, operating assets and inventories. Information about the realisation value of these assets is, however, not available. The annual change in real house prices is therefore used as a proxy for the change in the realisation value of the lenders' collateral.

According to equation (2), a 1% increase in risk-weighted debt will increase loan losses by 0.95%. A 1 percentage point reduction in the value of financial institutions' collateral will increase losses by 13%.

Risk-weighted debt for a company is defined as the product of the company's debt and its bankruptcy probability. It is an estimate of how much the lender can expect to lose in the absence of collateral. The risk-weighted debt will vary across firms according to the level of their debt, and according to their individual bankruptcy probabilities. The bankruptcy probabilities are estimated using Norges Bank's bankruptcy prediction model (SEBRA). For a description of the model, see Appendix 2. In SEBRA, the bankruptcy probabilities are a function of selected accounting variables (operating income, operating expenses, interest expenses, long-term debt and overdraft debt), company age and size and industry characteristics.

For actual and modelled losses in per cent of outstanding debt in Norway for the years 1989-2002, see Figure 2. This period covers the peak of the banking crisis in 1990-92, the following consolidation phase and the recent period from 2001 with increasing losses. The modelled losses are based on historical figures.

---

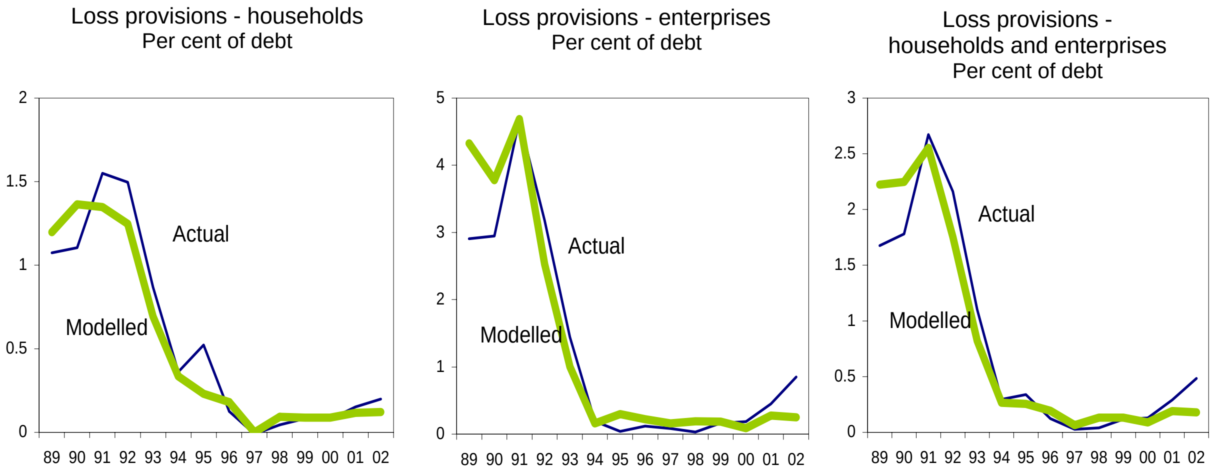
8 The variable *RPH* is an output from RIMINI in the stress tests.

9 Lower case letters indicate logarithmic form and Δ indicates the first difference of the variable.

Figure 2

Losses as a percentage of debt, by sector and aggregate

Loss in sector/debt in sector



For the corporate sector, risk-weighted debt is computed in three steps. First, each company's annual accounts are projected for the scenario period. This is done by assuming that key revenue and expense items in the accounts will vary in tandem with estimated changes in key macroeconomic variables. See Table 2 for a summary of the modelled relationship between macro and accounting variables.

Second, a bankruptcy probability is estimated for each company based on the projected accounts. Finally the risk-weighted debt for all companies is computed and aggregated.

The heterogeneity between companies is reflected in the variable risk-weighted debt. Risk-weighted debt is computed based on actual accounting figures. Hence, sectoral and regional differences in the profitability, liquidity and solvency of individual firms are reflected in their bankruptcy probabilities. Differences in debt growth between companies will also be reflected in the aggregate.

Table 2

Modelled relationship between accounting variables for companies and macro variables

	Accounting variable (at the company level)	Macro variable (output from RIMINI)
1	Operating income	Mainland GDP
2	Operating expenses excl wage expenses	Mainland GDP
3	Wage expenses	Wages
4	Interest expenses	Interest rate
5	Long-term debt and overdraft debt	

Variables 1-3 in the left-hand column are assumed to have the same yearly percentage increase (decrease) as the accompanying macro variables in the right hand column. Variable 4 is based on the level of the interest rate. Variable 5 is not an output of the RIMINI model.

However, some of this heterogeneity is lost when we project the accounts for the scenario period as we assume that all companies develop similarly. As an example, consider the case of operating income. The percentage growth in operating income will be equal to the growth in mainland GDP, irrespective of the industry. The year prior to the first scenario year also influences the projections. If a company has a particularly low operating income in the year in question, the results for the whole scenario will be influenced. The simplified modelling of company accounts is motivated by tractability. We do not, however, lose all the heterogeneity between companies. The debt and bankruptcy probability is still computed for each company individually.

4. Stress test results

4.1 The macroeconomic demand side shock

We study the impact on banks' loan losses of a considerable adverse macroeconomic demand side shock, initiated by a significant decline in public expenditure. Public consumption and investments are reduced permanently by 6 percentage points compared to a reference scenario. Note that the reference scenario that has been used to calculate the changes in macro variables due to the shocks is the forecast presented in Norges Bank's inflation report at that time, and not the actual outcome.

This drop in demand leads to a reduction in public sector employment, which also gives rise to other changes in the macroeconomic environment. In the scenario with no monetary policy response, the unemployment rate increases by around 1.5 percentage points in the first year, and after three years it is 2-2.5 percentage points higher than the unemployment rate in the reference scenario. Furthermore, we assume that these changes are followed by a decrease in the rise in house prices of around 10 percentage points per year in the first two of the three years involved in the forecasts, which means that house prices actually fall. Moreover, the inflation rate drops by 1 percentage point compared to the reference scenario after two years and 1.5 percentage points after three years. As these shocks yield substantial effects on both inflation and aggregate output, the results from no monetary response and a Taylor rule response are expected to be appreciably different. The results, which are summarised in Tables 3 and 4, show that monetary policy easing according to a Taylor rule mitigates the negative effects of the demand shock on the variables presented.

The shocks illustrated in this analysis are substantial. However, they are probably not necessarily unrealistic. The substantial macroeconomic instability and volatility experienced in the 1980s illustrate that large oscillations in macroeconomic variables can occur. For example, house prices in Norway fell by almost 30% from early 1988 to early 1993. The unemployment rate was 2% cent in 1986/87, before it increased and reached around 6% in 1993; see Figure 3.

Figure 3

Unemployment rate (per cent) and house prices (index) in Norway

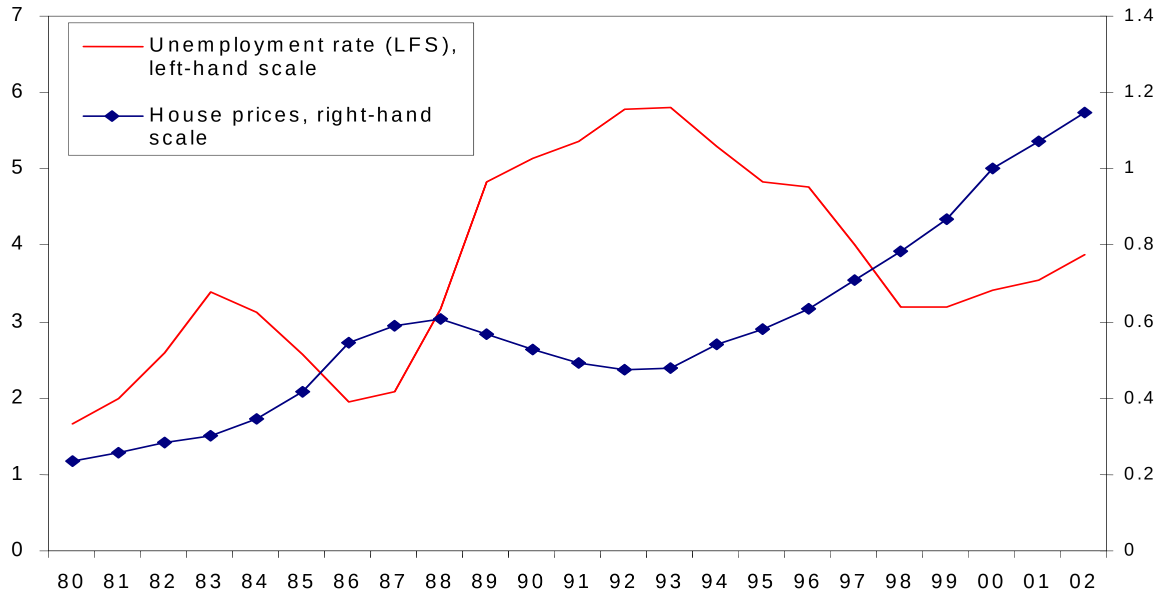


Table 3

Demand shock with no monetary policy response[1]

	***t* + 1**	***t* + 2**	***t* + 3**
Macroeconomic variables			
Mainland GDP	–2.5	–2.0	–1.7
Unemployment rate (change in level, percentage points)	+1.3	+1.8	+2.2
Wages	–0.3	–1.8	–2.0
CPI	–0.2	–1.0	–1.6
Household variables			
Credit growth households	–0.2	–2.8	–4.6
House prices	–10.0	–10.0	0.0
Value of house capital	–10.0	–10.0	0.0
Interest rate on loans (change in level, percentage points)	0.0	0.0	0.0
Interest expenses	0.0	–2.0	–4.7
Disposable income	–2.8	–1.9	–2.2

[1] Effect on growth rates (percentage points) unless otherwise stated. Shock occurs in year *t* + 1.

Table 4

Demand shock with monetary policy response according to a Taylor rule[1]

	***t* + 1**	***t* + 2**	***t* + 3**
Macroeconomic variables			
Mainland GDP	–2.3	–0.8	+0.3
Unemployment rate (change in level, percentage points)	+1.3	+1.7	+1.8
Wages	–0.4	–1.7	–1.6
CPI	–0.2	–0.9	–1.3
Household variables			
Credit growth households	–0.2	–1.9	–2.3
House prices	–9.0	–7.0	+3.0
Value of house capital	–9.0	–7.0	+3.0
Interest rate on loans (change in level, percentage points)	–0.9	–2.8	–3.5
Interest expenses	–10.0	–32.2	–8.2
Disposable income	–2.5	–1.1	–1.2

[1] Effect on growth rates (percentage points) unless otherwise stated. Shock occurs in year *t* + 1.

A motivation for a shock initiated by a drop in public spending could for example be the fact that approximately 25% of the Norwegian government's revenues stem from petroleum activities.[10] A large drop in the oil price that is perceived by policymakers as permanent could enforce a reduction in public sector expenses in order to balance the expected public revenues and expenses in the longer term.[11]

Loan losses with the demand side shock

Estimated losses for households in the cases involving demand shocks are higher than estimated losses in the baseline scenario; see Figure 4. The household sector is hit by the demand shock in the form of increased unemployment, reduced growth in disposable income and a reduction in households' housing wealth. These factors contribute to higher losses on loans to households. The monetary response partly reverses the changes in unemployment, disposable income and housing wealth. The net effect is that losses are higher compared to the baseline scenario, but lower than in the case with no monetary response.

Also, estimated losses in the corporate sector increase with the demand shock; see Figure 5. Losses in the corporate sector are larger than losses in the household sector. As expected, corporate loans are more risky. The effect of the demand side shock on the corporate sector's risk-weighted debt only influences estimated losses in years two and three of the scenarios. The reason is that risk-weighted debt is lagged by one year in the loan loss equation.

The demand effect of the demand side shock on the value of collateral, proxied by the change in value of housing, is negative and causes losses to increase in 1996 and 2001. Higher risk-weighted debt and a further fall in house prices contribute to increased losses in year two of the scenarios. In the final year, estimated losses fall. This is primarily due to a stabilisation of property prices.

The average bankruptcy probabilities for the different shocks, ie, risk-weighted debt per unit of debt, are illustrated in Figure 7. The demand shock increases risk-weighted debt primarily because it reduces sales in the corporate sector. Low wage growth contributes to a reduction in risk-weighted debt, but this effect is not sufficiently strong to dominate the effect of the sector's fall in revenues.

The monetary policy response according to the Taylor rule reduces the fall in property prices, thereby reducing the losses in the first year of the scenarios. The growth in risk-weighted debt is reduced, contributing to reduced losses in years two and three of the scenarios.

Risk-weighted debt increases more slowly because of lower interest rates and because of the smaller reduction in sales. These positive effects are not outweighed by the smaller decrease in wage growth.

Estimated losses in the cases involving demand shocks are higher than estimated losses in the baseline scenario in both sectors. Applying a Taylor rule for monetary policy implies reduced losses in both the household and corporate sector. Hence, with a demand shock, there is no conflict between inflation targeting and financial stability.

4.2 The macroeconomic supply side shock

If the economy is hit by a cost-push shock, there may be a trade-off between stabilising output and stabilising inflation. As often illustrated in the inflation targeting literature, a cost-push shock may lead to an increase in both inflation and unemployment. A tightened monetary policy aiming at stabilising inflation will then lead to a further increase in unemployment. Such a monetary policy reaction increases the burden on the financial system, due to both increased interest rates and an extra reduction in employment.

We have analysed the effects on banks' losses of a macroeconomic supply side shock. In this scenario, growth in annual wages increases by 4 percentage points per year compared to the baseline scenario. The results are summarised in Tables 5 and 6.

---

[10] Estimate for 2003; see the Government's Revised National Budget 2003 (Ministry of Finance).

[11] According to the fiscal policy rule in Norway, over time, the use of petroleum revenues over the government budget should be equal to the expected real return on the capital of the Petroleum Fund, stipulated at 4% per annum. Hence, a substantial fall in oil prices that is perceived as permanent will probably lead to a reduction in public spending.

The increase in wages leads to higher consumer price inflation. This is a result of both higher costs for enterprises/employers and higher domestic demand. In the short term, ie within a two-year horizon, higher wages lead to higher private consumption. Consequently, GDP growth increases. In turn, higher demand and production lead to lower unemployment. Usually, in the literature, a positive cost-push shock leads to an increase in both inflation and unemployment. When we as a result get an increase in inflation but a fall in unemployment, it is a result of the way we have designed the shock (ie as a wage shock) and a quite strong link from wage growth to private consumption in our model. In addition, expectations are not explicitly modelled, and households' and firms' current decisions are not affected by the long-run consequences of higher wage growth.

In the longer term, one would expect a wage shock to cause a deterioration in conditions for enterprises. As wage costs rise dramatically, many enterprises will be forced to cut back on their stocks of employees. In addition, the bankruptcy rate would increase. It normally takes some time before these effects on employment are exhausted. In a perspective of about one year, it is not clear whether the positive aggregate demand effect or the negative cost effect of a large wage rise dominates. In our scenarios, the total effect on employment from such a wage shock is slightly negative after two years, so that unemployment is higher. In the longer term we would expect the negative effects on employment to dominate more clearly. (A parallel to this is the situation in Norway where wage growth was high in the period 1998-2002. This has probably had a negative impact on employment growth, especially in the internationally exposed industries.)

Table 5

Supply shock with no monetary policy response[1]

	***t* + 1**	***t* + 2**	***t* + 3**
Macroeconomic variables			
Mainland GDP	+0.2	+1.0	+1.6
Unemployment rate (change in level, percentage points)	0.0	+0.3	+0.5
Wages	+4.0	+4.0	+4.0
CPI	+0.6	+2.2	+2.3
Household variables			
Credit growth households	+0.1	+1.0	+2.1
House prices	+1.2	+4.3	+5.5
Value of house capital	+1.2	+4.4	+5.8
Interest rate on loans (change in level, percentage points)	0.0	0.0	0.0
Interest expenses	0.0	+0.7	+2.0
Disposable income	+3.0	+3.3	+3.4

[1] Effect on growth rates (percentage points) unless otherwise stated. Shock occurs in year *t* + 1.

Table 6

Supply shock with monetary policy response according to a Taylor rule[1]

	t **+ 1**	***t*** **+ 2**	***t*** **+ 3**
Macroeconomic variables			
Mainland GDP	+0.3	–0.1	+0.1
Unemployment rate (change in level, percentage points)	0.0	+0.4	+1.0
Wages	+4.0	+3.0	+2.0
CPI	+0.6	+2.0	+1.4
Household variables			
Credit growth households	+0.1	+0.2	–0.8
House prices	0.0	–1.4	+2.4
Value of house capital	0.0	–1.5	+2.3
Interest rate on loans (change in level, percentage points)	+0.7	+3.0	+2.3
Interest expenses	+7.7	+36.5	–13.1
Disposable income	+2.8	+1.7	+1.6

[1] Effect on growth rates (percentage points) unless otherwise stated. Shock occurs in year $t + 1$.

Moreover, in the scenario with no monetary policy response, the rise in wages contributes to higher house prices. This increases the value of the collateral of banks, which in turn reduces banks' losses.

Under inflation targeting, a sudden increase in labour costs will prompt an increase in interest rates to counteract the build-up of inflationary pressures. If the response pattern of the central bank is well known, we expect the monetary policy regime to have a disciplinary effect on wage growth. The labour unions will foresee that high wage growth results in higher inflation and then higher interest rates, reducing the disposable income of households with debt. Higher interest rates will typically lead to an appreciation of the krone, with a further reduction in earnings and employment for the exposed businesses. In line with these arguments, we have assumed in our cost-push scenario with a monetary policy reaction that the central bank's response pattern is gradually internalised by trade unions. When a Taylor-rule monetary policy response is implemented, we assume that wage growth only increases relative to the reference scenario by 4 percentage points the first year, 3 percentage points in the second, and 2 percentage points in the third year (see Table 6).

As wage increases lead to higher inflation, the Taylor rule yields higher interest rates. This in turn curbs aggregate demand and house prices. Unemployment increases.

Loan losses in the supply side shock case

The supply shock causes a reduction in estimated losses compared to the baseline scenario in the household sector; see Figure 4. Higher wage growth increases disposable income and house prices. The effect of these positive factors is not reversed by the increase in unemployment.

The monetary policy response causes higher interest rates, increased unemployment and a more moderate development in housing wealth in the household sector. Due to the hike in interest rates, house prices fall in year two, relative to the reference scenario. The result is that losses increase. The level of estimated loss is higher than the loss in the baseline scenario.

The supply shock causes a reduction in estimated corporate losses compared to the baseline scenario; see Figure 5. With a Taylor-type monetary policy response, losses increase to a level above the losses in the baseline scenarios.

The supply shock increases risk-weighted debt because of the large increase in companies' wage costs. The increase in wages is, however, also accompanied by an increase in sales, but the increase in wages dominates the latter effect. Risk-weighted debt increases and contributes by itself to increased losses in years two and three of the scenarios. Increased wage growth is accompanied by higher property prices. This increases the banks' value of collateral. Estimated losses are marginally reduced in the first year of the scenarios. This effect is, however, reversed by the rise in property values, leaving the estimated losses well below the losses in the baseline scenario.

A shortcoming in the way we model the loan losses in the corporate sector is that companies exposed to foreign competition and companies sheltered from foreign competition are equally influenced by the rise in domestic demand caused by increased domestic consumption. In general, internationally competing companies will be severely hit by the supply shock through increased wages. The result is that losses in the corporate sector are underestimated.

With a Taylor-like monetary policy response, the interest rate increase causes lower wage growth, reduced sales growth and lower property prices. In the first year of the scenario, property prices are unchanged compared to the baseline scenario. Accordingly, the estimated losses are unchanged. An increase in risk-weighted debt contributes to increased losses in year two of the scenario. The effect of a further increase in risk-weighted debt in year three of the scenario is counteracted by an increase in property values. Property prices increase in the third year because households' disposable income growth is high (due to a high wage increase and somewhat lower interest rates in this year). Estimated losses are accordingly reduced in the final year.

The supply shock with no monetary policy response causes a reduction in estimated losses compared to the baseline scenario in both the corporate and household sector. With a Taylor-like monetary policy response, however, losses increase above the level in the baseline scenario in both sectors. Of all the scenarios we consider, the supply shock with a monetary response increases risk-weighted debt the most. The combination of high wage growth and high interest rates severely worsens the cost burden of the corporate sector. In addition, with a supply shock, there may be a potential conflict between the objective of monetary policy and financial stability. In the relatively short time frame analysed here, monetary policy aimed at achieving the inflation target leads to higher losses in both the household and corporate sector. However, in a longer time perspective, this trade-off might be somewhat different (for further discussion see Section 5).

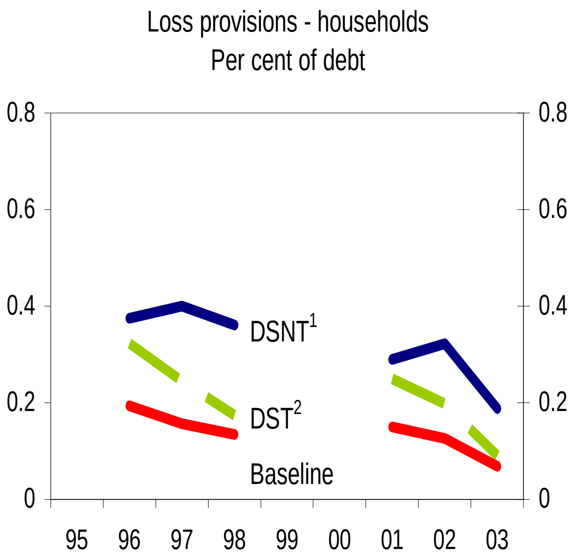
The difference between 1996 and 2001

The household sector was marginally better positioned in 2001 than in 1996, as measured by estimated losses in the first year of the baseline scenarios. Estimated losses (measured as a percentage of loans) fell by 0.04 percentage points from 1996 to 2001; see Figure 8. This reduction was caused by lower unemployment and an increase in housing wealth. During this period, the debt burden and interest rates rose, but not sufficiently to increase the household sector's losses. With the household sector in approximately the same condition along the baseline scenarios during the two time periods, the effects of the identical shocks are also almost identical irrespective of whether the shocks occur in 1996 or 2001.

Figure 4

Losses in the household sector by type of shock

Per cent of debt (loss in sector/debt in sector)

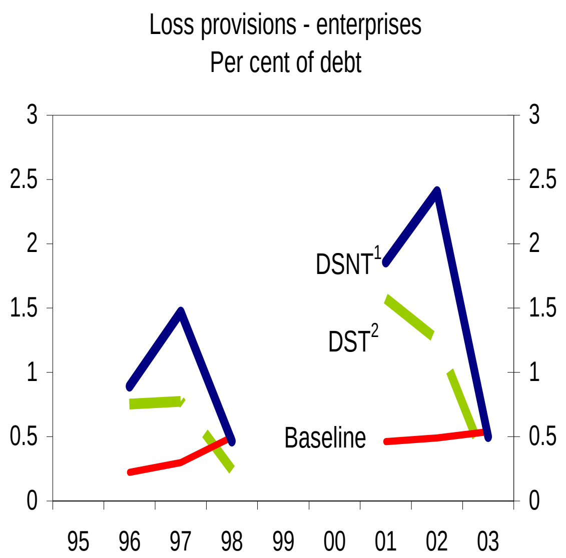


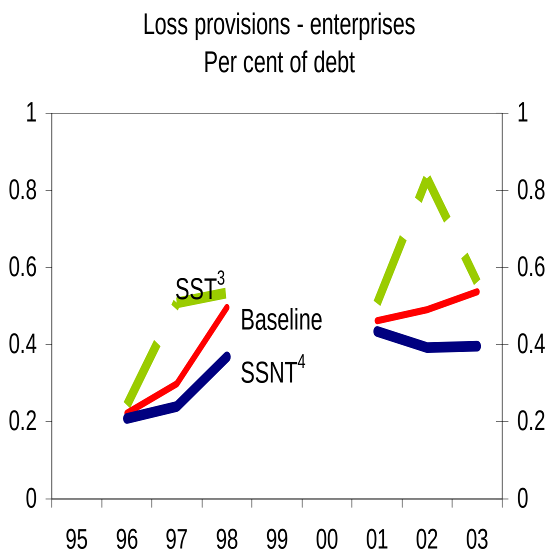
[1] Demand shock no Taylor rule (DSNT). [2] Demand shock with Taylor rule (DST). [3] Supply shock with Taylor rule (SST). [4] Supply shock no Taylor rule (SSNT).

Figure 5

Losses in the corporate sector by type of shock

Per cent of debt (loss in sector/debt in sector)



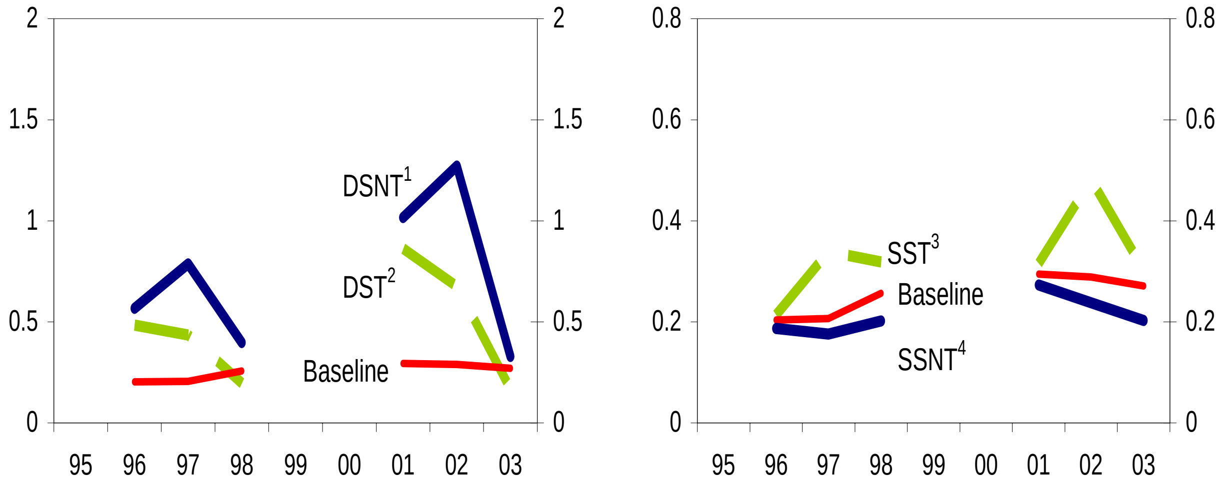


[1] Demand shock no Taylor rule (DSNT). [2] Demand shock with Taylor rule (DST). [3] Supply shock with Taylor rule (SST). [4] Supply shock no Taylor rule (SSNT).

Figure 6

Total losses by type of shock
(household and corporate sector)

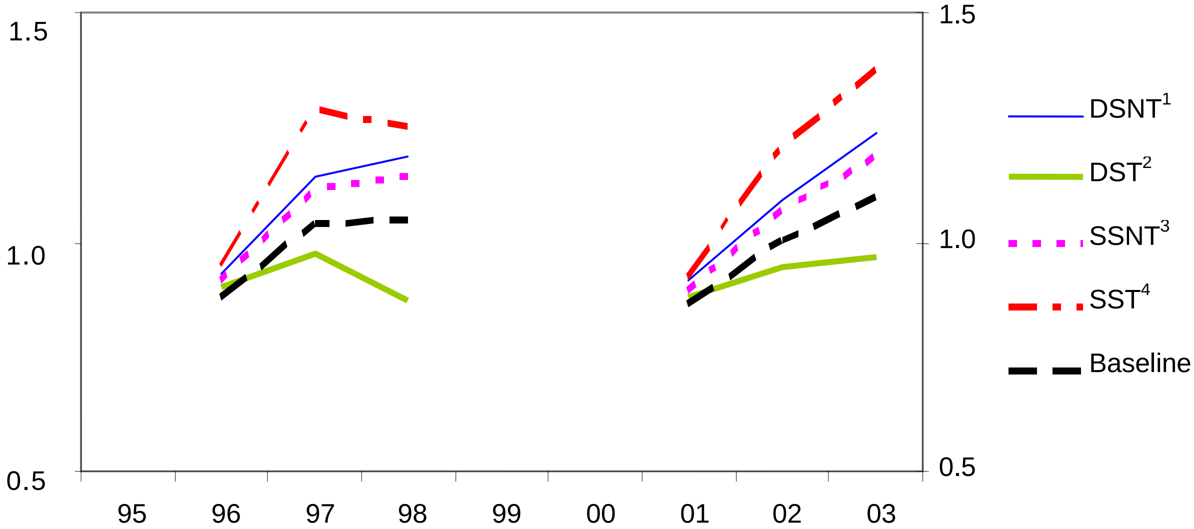
Per cent of debt



[1] Demand shock no Taylor rule (DSNT). [2] Demand shock with Taylor rule (DST). [3] Supply shock with Taylor rule (SST). [4] Supply shock no Taylor rule (SSNT).

Figure 7

Average bankruptcy probabilities in per cent
for various shocks (risk-weighted debt/debt)

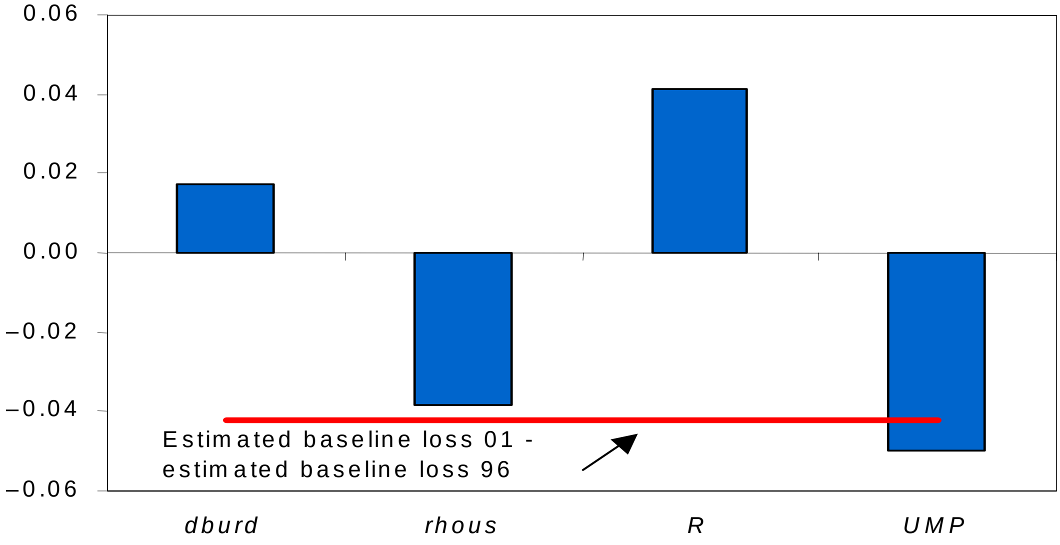


[1] Demand shock no Taylor rule (DSNT). [2] Demand shock with Taylor rule (DST). [3] Supply shock no Taylor rule (SSNT). [4] Supply shock with Taylor rule (SST).

Figure 8

Partial effects of the explanatory variables causing a lowering of the estimated losses in the household sector from baseline scenario in 1996 to baseline scenario in 2001

Red line shows total effect. Percentage points of loans (*dburd*: debt burden, *rhous*: real value of houses, *R*: interest rate, *UMP*: unemployment rate)

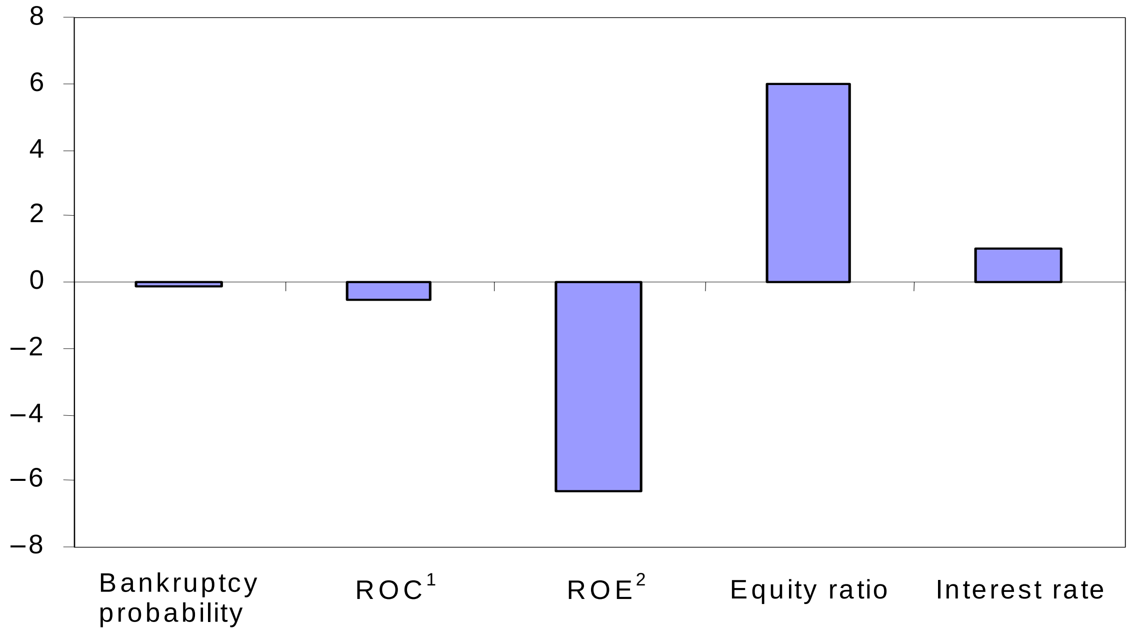


In spite of a higher debt burden, the corporate sector was also in a better position in 2001 than in 1996 measured by the average bankruptcy probability. The corporate sector had been operating profitably for five years and the equity ratio had increased by approximately 6 percentage points. This was the main factor behind the drop in bankruptcy probability between the two periods; see Figure 9.

Figure 9

Changes in key variables influencing companies' risk-weighted debt from 1996 to 2001

Changes measured in percentage points



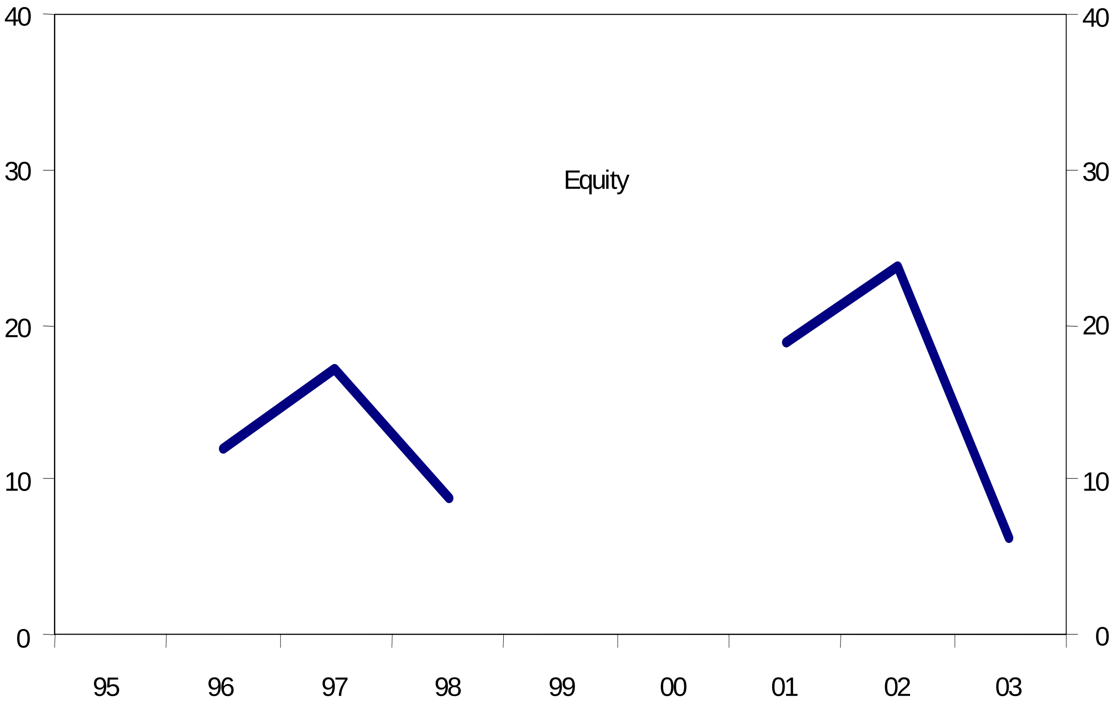
[1] ROC: return on capital. [2] ROE: return on equity.

As can be seen from the loan loss equation (2) for the corporate sector, property prices play an important role in the estimation of loan losses. The growth in house prices along the baseline scenario is lower in 2001 than in 1996. This outweighs the impact from the lower bankruptcy probability and causes the estimated losses to be higher in 2001 than in 1996.

The estimated demand shock losses as a percentage of banks' equity were small; see Figure 10. The shock starting in 2001 would have reduced equity more than the shock starting in 1996. The shocks would probably not have caused a banking crisis, especially since banks could have raised additional capital to improve their capital ratios.

Figure 10

Aggregate losses in the demand shock scenario without a monetary response[1]



[1] Aggregate losses for DSNT as a percentage of banks' equity at beginning of 96 and 01.

5. Conclusions and implications

Should monetary policy pay special attention to asset prices and the build-up of financial imbalances? The answers to this question differ somewhat in the international debate, but it has been pointed out that there need not be a conflict between the objectives of maintaining both financial and monetary stability.

First, flexible inflation targeting, where the central bank puts emphasis on smoothing variability in both inflation and output, reduces the scope for conflict between the monetary policy objective and financial stability. After a shock has brought inflation away from its target, a central bank may choose to bring inflation back to target rapidly. Such a strategy, which can be termed strict inflation targeting, would typically imply instability in output and employment. By contrast, flexible inflation targeting involves applying a somewhat longer horizon to achieve the inflation target. This would normally represent a smaller threat to financial stability than a strict inflation targeting regime, as it involves smaller fluctuations in production, employment, asset prices and interest rates.

Second, when assessing how financial stability issues should be dealt with in the conduct of monetary policy, it is useful to distinguish between the short and the long term. In this paper we have calculated the short-term effects of interest rate changes on the financial sector. Lower interest rates will reduce debt servicing costs and thus reduce the risk of higher loan losses in the short term. In the long term, the isolated effect of an expansionary monetary policy will be a faster rise in indebtedness and asset prices, which may increase future financial fragility. Higher interest rates have the opposite effect.

As illustrated in this paper, there seems to be no short-term conflict between financial and monetary stability when the economy is facing a typical negative aggregate demand shock. In this case, a monetary policy reaction following a standard Taylor rule, which may be interpreted as the response of a central bank with a flexible inflation target, would dampen the drop in inflation and production, but also reduce banks' loan losses. Even in the longer term a monetary policy aimed at stabilising inflation and output would most likely have a positive impact on financial stability by improving the robustness of the banks and their borrowers.

However, there may also be a risk that in the longer term the lower interest rate may stimulate excessive indebtedness and asset prices, so that financial fragility increases. Many firms will have excess capacity during an economic downturn, making it less probable that the corporate sector will react to an expansionary monetary policy by sharply increasing its debt exposure. The risk may be higher for the household sector. A heavily indebted household sector will be vulnerable to adverse shocks that may hit the economy in the future. Some households may also face financial problems when the economy recovers and interest rates return to their neutral level. The risk of increased financial fragility should be weighed against the consequences for activity, inflation and financial stability in the short run, if monetary policy is not eased sufficiently to counteract a negative demand shock.

With regard to cost-push shocks stemming from a sudden boost in wages, a conflict between monetary and financial stability may arise in the short term. The higher interest rate needed to maintain monetary stability will increase debt servicing costs, which may increase credit losses. In our scenario, aggregate demand increases immediately due to a positive effect on households' disposable income. Consequently, companies producing for domestic markets will experience reduced pressure on operating profits and loan losses are lower than in the baseline scenario. However, in the somewhat longer run increased wage costs have a negative impact on the operating results of all companies.[12]

Due to the higher inflationary pressures, the monetary policy reaction to such cost shocks would be increased interest rates. The positive demand effect is thus partly counteracted, while increased interest rates (and a potential appreciation of the currency due to increased interest rate differentials) at the same time place an extra burden on enterprises' expenses. Regarding the cost-push shock, we showed that a Taylor-rule monetary policy reaction raises the level of credit losses above the baseline scenario.

In the longer term, this conclusion may be turned around. Without monetary tightening, a continued increase in wage growth will have to stop at a later stage, due to longer-term economic dynamics. However, the consequences, if not curbed at an early stage, may be higher unemployment due to reduced competitiveness of exposed industries, and hence higher credit losses. As debt levels, asset prices and hence financial fragility most likely would have increased further in the meantime, the consequences for financial stability could even be more severe than if monetary policy was tightened immediately.

The appropriate central bank response when a cost-push shock occurs would of course depend on the magnitude of the forecast short-term losses. We found that the losses in these cases were rather small from a historical perspective and they would probably not have caused a banking crisis. The banks' buffer capital would have been sufficient to absorb the losses. The costs of not raising the interest rate would be related to the deviation of inflation from its target for a longer period, which could reduce monetary policy credibility.

We have run the same set of stress tests on what seemed to be two different periods with regard to financial vulnerability. When concerned with financial stability, the main focus is on the level and

---

[12] In particular, companies in exposed industries may experience a deterioration in profitability and competitiveness. Unfortunately, the difference between the sheltered and exposed sectors is not modelled within the micro-based SEBRA framework. If it were, we could have separated the effects of increased private consumption between domestic/sheltered industries and exposed industries.

increase in debt and asset prices. But other variables are also important. The household sector is particularly affected by lower unemployment in 2001 relative to 1996. This offsets the negative impact of other factors like the higher debt burden.

Growth in the debt burden was stronger for enterprises than for households in the period between 1996 and 2001. As with households, this was accompanied by an improving economy that increased the equity ratio of firms and thus strengthened their ability to withstand shocks. The level of the average bankruptcy probability fell slightly.

These results show that a sole focus on debt and asset prices may be too narrow when assessing the financial fragility of households and enterprises. It is important to include other factors that may have an impact on the different sectors' debt servicing capacity.

Also, the chosen years may not capture precisely the trough and the peak of the credit cycle. The results might have been different if the current year (2003) had been chosen instead of 2001. Debt levels have continued to rise and the unemployment rate is now significantly higher. It is possible that the financial situation in the corporate sector has deteriorated somewhat. In general, when an economy is recovering from a recession, a rise in debt levels and asset prices is not necessarily worrying. It is the excessive build-up of debt over time - and clearly over a time period when unemployment cannot continuously fall - that gives cause for concern.

It is clear from our loan loss equations that losses will increase with the level of indebtedness. Hence, the weight of preserving the soundness of the banking sector in monetary policy decisions should increase with the level of indebtedness.

Analyses of alternative scenarios are important as part of the monetary policymaking process. With the help of stress testing we may analyse the effects on the banking system of different macroeconomic scenarios. Within the SEBRA framework, the basis for making this analysis is the accounting data for all Norwegian joint stock companies. Also in the future, situations may arise in which the financial sector is vulnerable to adverse shocks. Stress testing may give us an early warning and monetary policy authorities may then assess whether this should give cause for particular concern.

One shortcoming of using this framework as an "early warning tool" is the fact that the accounting data for Norwegian companies lag. For example, data are only available up to 2001. This can, however, be partly solved by using projections for companies' operating results and other key variables.

Stress testing is a useful tool when analysing developments in the economy and financial stability. Such analyses improve the understanding of the interaction between "traditional" macroeconomics and financial stability issues. It is, however, important to understand the shortcomings of the method. As with all models, we may fail to include important variables. It is also not obvious that models calibrated on historical data are relevant for forecasting.

Finding better indicators for assessing the vulnerability of households, enterprises and financial institutions is probably the most important step to improve this kind of analysis. One important step in this regard would be to find indicators of bubbles in the property market. For example, it is reasonable to expect that the change in property prices following a negative macroeconomic shock would be larger the higher property prices are above their "equilibrium" values. The use of house prices as a proxy for commercial property prices and the value of collateral may also be solved in a better way.

An important question is how, and at what cost, monetary policy can curb the build-up of financial imbalances (ex ante). A "leaning against the wind" policy requires indicators which give information on the build-up of financial imbalances. Although our framework has some shortcomings, it may add to the suite of indicators we use in order to detect looming financial instability.

Appendix 1:
Model for losses in the household sector

The model is a re-estimation of the model presented in Frøyland and Larsen (2002). The time series underlying this model has been revised. We tested various model specifications with alternative variable specifications. These alternative specifications did not, however, give any new insights into the effects of the shocks.

Summary

The estimation sample is: 1978-2002

	Coefficient	**Std error**	***t*-value**	***t*-prob**	**Partial R^2**
dburd	3.31404	0.8116	4.08	0.001	0.4547
rhouse	–1.44635	0.2499	–5.79	0.000	0.6261
R	13.5534	3.002	4.51	0.000	0.5047
UMP	31.5508	8.068	3.91	0.001	0.4333
DUM97	–7.04948	0.3576	–19.7	0.000	0.9511

sigma	0.322947	RSS	2.08589821
log-likelihood	–4.42751	DW	1.96
no of observations	25	no of parameters	5
mean(lossrel)	–1.63586	var(lossrel)	3.33868

AR 1-2 test:	$F(2,18) = 0.75655$ [0.4836]
ARCH 1-1 test:	$F(1,18) = 0.026134$ [0.8734]
Normality test:	$X^2(2) = 9.0003$ [0.0111]
hetero test:	$F(9,10) = 0.88651$ [0.5670]
RESET test:	$F(1,19) = 0.70515$ [0.4115]

Appendix 2:
The bankruptcy prediction model SEBRA

The bankruptcy prediction model SEBRA is a logistic model. For each joint stock company in the database (in 2001 the number of companies is approx 140,000) the model produces an estimate of the bankruptcy probability. The model is presented in Eklund et al (2001). The explanatory variables reflect primarily company-specific information, like earnings, liquidity, financial strength and age, but industry-specific information is included in the model, like the average equity ratio and dispersion in earnings. A summary of the variables is given below.

Earnings

- Earnings as a percentage of total assets

Liquidity

- Liquid assets less short-term debt as a percentage of operating revenues
- Unpaid indirect taxes as a percentage of total assets

Financial strength

- Equity as a percentage of total assets
- Dummy variable for book equity less than paid-in equity capital
- Dummy variable for dividend payments the last accounting year

Industry

- Industry average for the variable "equity as a percentage of total assets"
- Industry average for the variable "trade accounts payable as a percentage of total assets"
- Industry standard deviation for the variable "earnings as a percentage of total assets"

Age

- Dummy variable for number of years since establishment

Size

- Total assets

References

Bernanke, Ben and Mark Gertler (1999): "Monetary policy and asset price volatility", *Economic Review*, fourth quarter, 1999.

Bank for International Settlements (2001a): *The new Basel Capital Accord*, Consultative Document, 16 January.

——— (2001b): "Stress testing in practice: a survey of 43 major financial institutions", *BIS Quarterly Review*, June 2001.

Borio, Claudio and Philip Lowe (2002): "Asset prices, financial and monetary stability: exploring the nexus", *BIS Working Papers*, no 114, Basel, July 2002.

Borio, Claudio, William English and Andrew Filardo (2003): "A tale of two perspectives: old or new challenges for monetary policy", *BIS Working Papers*, no 127, Basel, February 2003.

Cecchetti, Stephen G, Hans Genberg, John Lipsky and Sushil Wadhwani (2000): "Asset prices and Central Bank Policy", *Geneva Reports on World Economy 2*, International Center for Monetary and Banking Studies, Geneva.

Eitrheim, Øyvind and Bjarne Gulbrandsen (2001): "A model based approach to analysing financial stability", *BIS Papers*, no 1, 2001, Basel.

Eklund, Trond, Kai Larsen and Eivind Bernhardsen (2001): "Model for analysing credit risk in the enterprise sector", *Economic Bulletin*, 3/2001, Norges Bank.

Frøyland, Espen and Kai Larsen (2002): "How vulnerable are financial institutions to macroeconomic changes? An analysis based on stress testing", *Economic Bulletin*, 3/2002, Norges Bank.

Olsen, Kjetil and Fredrik Wulfsberg (2001): "The role of assessments and judgement in the macroeconomic model RIMINI", *Economic Bulletin*, 2/2001, Norges Bank.

Taylor, J B (1993): "Discretion versus policy rules in practice", *Carnegie-Rochester Conference Series on Public Policy*, 39, 195-214.

Measuring and forecasting stress in the banking sector: evidence from Switzerland

Elke Hanschel and Pierre Monnin[1]

1. Introduction

Central banks and supervisory authorities regularly assess the situation in the banking sector, which is a vital element in the financial system. Two main questions are at the centre of such assessments: what is the present condition of the banking sector, and how will it evolve in the medium term? The first aim of this paper is to develop a "stress index", summarising the current condition of the Swiss banking sector in one single measure. The second goal is to forecast the stress index with the information drawn from the economic environment and macroeconomic imbalances, which have the potential to influence the condition of the banking sector in the medium term.

Banking crises and their determinants have been the subject of widespread empirical research in the last few years. Binary variables, which signal whether a banking sector is in a crisis or not, are frequently used in the literature to describe the condition of banking sectors in emerging markets.[2] But, as banking crises are rare birds in industrialised countries, binary variables are less suitable to depict the condition of their banking sectors. Yet, the absence of full-blown crises does not mean that the condition of the banking sector is always equally sound and stress-free.

The index developed in this paper is an attempt to discern the fluctuations in the banks' stress. The index represents a continuum of states, describing the banking sector's condition ranging from low levels of stress, where the banking sector is tranquil, to high levels of stress, where the banking sector is in a severe crisis. To our knowledge, it is the first time that a stress index focusing exclusively on the banking sector has been developed. Four different types of variables are used to build the stress index: market price data, balance sheet data, non-public data of the supervisory authorities, and other structural variables. The stress index for Switzerland is calculated on a yearly basis for the period from 1987 to 2002. The final stress index shows two episodes of high stress: at the beginning of the 1990s, when the industry saw major restructuring and takeovers among regional banks, and in 2001-02, during the stock market crash. Financial experts in Switzerland generally agree that these two episodes were the most stressful in the period under consideration. We also find that indices based only on one single type of variable do not detect all stress episodes. This suggests that several variables should be incorporated in the stress index in order to capture the different ways in which a banking crisis can show up.

After computing the stress index, we explore the impact of the economic environment and macroeconomic imbalances on stress. Previous studies on early warning systems (EWSs) for banking crises have empirically established a link between the real economy and the financial sector. This suggests that the economic environment corresponds to a common risk to all financial institutions and that it has the potential to forecast the stress. Thus, if macroeconomic imbalances are prevailing and the economy is weak, the banking system is more prone to experience crises or stress in the near future.

Instead of taking the level or the growth rate of the variables, as most previous studies do, we use deviations from their longer-term trend, ie so-called gaps. The advantage of the gaps is that they underline the cumulative process of the imbalances: a large trend deviation can develop either in one period with strong above (or below) trend growth or through a sequence of years with above (or below)

---

[1] Pierre Monnin, Swiss National Bank, and Elke Hanschel, Credit Suisse (the research was conducted while the author was working as an economist at the Swiss National Bank). Both authors would like to thank Bertrand Rime and Jürg Blum for their helpful comments. The opinions herein are our own and do not necessarily reflect those of the Swiss National Bank or the Credit Suisse. E-mail: pierre.monnin@snb.ch and elke.hanschel@credit-suisse.com.

[2] See, for example, Goldstein et al (2000).

trend growth. To calculate the trend, we apply a "rolling" technique taken from Borio and Lowe (2002a). Hence, we only use information that was available to the policymakers at each point in time and we do not take advantage of the information contained in the full sample. The future value of the stress index is then regressed on the gaps. Our results confirm that macroeconomic imbalances explain a substantial part of the future stress in the Swiss banking sector. After checking the robustness of our model, we find that there is some multicollinearity and that the forecasts vary significantly with the specification of the model. One reason could be that the number of observations for the stress index in our sample is limited.

The outline of the paper is as follows. Section 2 gives a definition of the stress index. The stress index is then computed with data for Switzerland, and its validity and robustness are discussed. Section 3 briefly outlines which macroeconomic imbalances can serve as early warning signals for future stress in the banking sector of developed countries. The method for calculating the gaps is also presented. In Section 4, we forecast the stress index for Switzerland. Finally, Section 5 concludes and highlights open issues for future research.

2. Measuring stress in the Swiss banking sector

Usually, one expects the central banks' assessment of the banking sector to be a condensed judgment, ie is the situation good, bad, better/worse than in the last quarter? Ideally, the answer to this question should be a single indicator, which summarises the global condition of the banking industry. In this section, we develop such an indicator and estimate it for the case of Switzerland.

2.1 Which measure best describes the condition of the banking sector?

When we look into the literature for an indicator that describes the condition of the banking sector, we usually find binary indices, which indicate whether the banking sector is in crisis at a given point in time or not. The literature on banking crises devotes great attention to the identification and description of such crises. Unfortunately, this kind of information is not very useful for depicting the situation in most developed countries, where banking crises are rare, if not inexistent. Nevertheless, even if these countries do not experience crises, it does not mean that their banking sector remains constantly sound and stable; their banks also go through good and bad periods, where they suffer greater or lesser degrees of stress. Therefore, a measure of the banking sector's stress probably gives a better picture of the banks' condition than a simple binary crisis indicator. Furthermore, to differentiate crises from tranquil periods, critical values must be arbitrarily chosen. Stress indices, on the other hand, are continuous, which means that they do not require the definition of critical values. This eliminates part of the arbitrariness.

2.2 Definition of stress in the banking sector

The stress indicator represents a continuum of states which describe the banking sector's condition at a given point in time. The stress level is measured on a scale ranging from tranquil situations, where stress is quasi-absent, to extreme distress, where the banking sector goes through a severe crisis. It is important to distinguish the banking sector's stress from its fragility. Stress emerges from the combination of exogenous shocks and fragilities in the banking system. A fragile banking sector does not systematically suffer stress if it benefits from a quiet and stable environment. Conversely, a solid banking system can undergo stress if it experiences extreme exogenous shocks. The interaction of the shock's magnitude and the banking system's fragility determines the stress level.

2.3 How to measure stress in the banking sector

To our knowledge, there is no continuous measure in the literature which specifically focuses on estimating banking sector stress. One should, however, mention the study by Illing and Liu (2003), who build a subindex for the banking sector (by estimating the beta of a bank's stock portfolio) and use it in a general financial stress index. Another study, by Bordo et al (2000), proposes a global financial condition index without concentrating on the banking sector.

We base our stress index on the observation of crisis symptoms in the banking sector. Typically, several symptoms signal banking crises (bank run, fall in the banks' stock price, bank failures, etc). To measure the stress level, we estimate the gravity of the different crisis symptoms at a given date. If the symptoms are present and acute, the banking sector is likely to be in a crisis situation and, therefore, the stress is likely to be high.

There is an extensive literature focusing on the definition and identification of banking crises. We rely on it to define a set of variables representing crisis symptoms. We then measure their intensity and aggregate them to form our final stress index. Researchers generally agree that banking crises show up in many different ways and that identifying them implies a certain degree of subjectivity. This conclusion suggests that a single variable cannot capture the complexity of crises. To detect the many forms that a banking crisis can take, we build a stress index, which combines several types of variables (ie market prices and balance sheet data). The next section lists the variables included in our index. Most of them have been suggested by earlier studies, but some are specific to our index.

2.4 Variables included in the stress index

Since our goal is a quantitative and continuous index, we only consider quantitative variables. Each variable reflects a potential symptom of banking stress. Due to the annual frequency of some series, we compute a yearly index for the Swiss banking sector from 1987 to 2002. We use four types of information: market prices, aggregate balance sheet data, non-public information, and other structural data.[3]

Market price data

The first selected variable is the banks' stock price index. When the banking sector goes through a crisis, its intrinsic value diminishes and, subsequently, banks' stock prices fall. A period of high stress should therefore be characterised by a decreasing banking sector stock price index. This criterion has been previously used by Vila (2000) and Illing and Liu (2003) to define stress and crisis events. In order to detect falling stock prices, we look at the biggest decline in 12 months observed during the year.[4] This measure allows sharp falls to be exhibited more clearly than with the raw data (Vila (2000)).

The second market price variable used in our index is the yield spreads for bank-issued bonds. The spread reflects the risk that investors associate with the banking sector. During a crisis, a higher spread should be observed. Illing and Liu (2003) suggest this variable. We use the average spread over one year in our index.

Balance sheet data

A typical banking crisis symptom is a sudden drop in deposits, which reflects the loss of confidence on the part of depositors in the banking system (bank run). This criterion is widely used to identify banking crises (eg Kaminsky and Reinhart (1996, 1999), Demirgüç-Kunt and Detragiache (1998) and Vila (2000)). To detect this phenomenon, we incorporate the total interbank deposits in our index. We think that bank runs should be well reflected in this variable since interbank deposits are relatively liquid and very partially insured. Furthermore, we can assume the banks to be more informed on the situation of their rivals than the public.

As the fourth variable, we use the return on assets of the banking sector. Although this variable is, to our knowledge, not used in the literature, we think that it is a relevant criterion for developed countries. It seems plausible that a non-profitable banking sector is a sign of trouble and that it should be associated with high stress.

---

[3] See the appendix for the data sources.

[4] This measure is called CMAX and is computed with the following formula: $CMAX_t$ = $index_t$ / maximum index over the last 12 months.

The fifth variable is the variation in bank capital. This variable has been used by Caprio and Klingebiel (1996) and González-Hermosillo (1999) to identify systemic banking crises. If a bank is in trouble, its capital will tend to shrink and, therefore, a banking sector in crisis should experience a decrease in total capital.

Another sign of crisis is found in the banks' evaluation of their own situation. This is reflected in banks' provisions, since, if a bank thinks that its situation is deteriorating, it should accumulate provisions. To take this information into account, we integrate the banking sector's provision rate into our stress index. Unfortunately, provisions are not an unbiased signal of crisis because, in stress periods, the banks' capacity and incentive to raise provisions might be reduced. González-Hermosillo (1999) uses a similar measure, namely the loan reserve coverage of non-performing loans.

Non-public data

The Swiss Federal Banking Commission, the banking supervisory authority of Switzerland, maintains a list of banks that are under special scrutiny. A bank appears on this list if it is experiencing unusual problems. We use the total assets of the banks on this list to estimate the share of the banking sector considered to be in trouble by the banking supervisory authority. During a crisis, this share should increase. One might think that this variable represents the "true" value of banking system stress since the supervisory authority has broader access to banks' information than the public. Unfortunately, this is not the case because: (1) only banks with unusually large-scale problems, or a high degree of stress, are registered and, therefore, the list does not show minor episodes of stress; and (2) the authority might sometimes miss problems, or detect them with a delay. However, the evolution of the banks under special scrutiny is certainly correlated with stress in the banking sector.

Other structural variables

Finally, we consider the variation in the number of bank branches. This variable takes into account the possibility of bank failures or reorganisation in the banking sector. This criterion is used by Bordo et al (2000), who consider the bank failure rate, and by Kaminsky and Reinhart (1996, 1999), who look at the closures, takeovers or mergers in the banking sector, to define crises. Our hypothesis is that bank failures or, at least, bank mergers or reorganisations are more likely to occur in periods of stress.

Variables not taken into account in the stress index

Information about non-performing loans is also a variable which appears frequently in studies on banking crises. Demirgüç-Kunt and Detragiache (1998), Corsetti et al (1998) or González-Hermosillo (1999) use the ratio of non-performing loans as a sign of crisis. Bordo et al (2000) prefer the loan loss rate. Unfortunately, the data on non-performing loans for the Swiss banking sector are only available since 1999.

2.5 Construction of the index

We combined the variables described above into one single stress index. At this stage, the choice of weighting method is crucial, since it determines the impact of each variable on the final stress index. We choose the variance-equal weight method to compute our index. This technique is the most common in the literature.[5] It consists, first, in standardising the variables to express them with the same units and, second, in aggregating them using identical weights. The index formula is the following:

$$I_t = \sum_{i=1}^{k} \frac{X_{i,t} - \overline{X}_i}{\sigma_i}$$

[5] There are other techniques to construct an index. Unfortunately, we were not able to find an alternative weighting scheme which would naturally fit in the context of a banking stress index. For example, we tried to apply the factor analysis technique, but this method does not yield meaningful results in the Swiss case, since the variables do not move together.

where k is the number of variables in the index, $\overline{X}_i$ is the mean of the variable X_i and σ_i its standard deviation. We also standardise the final index to express it in terms of deviations from its mean.

2.6 How to assess the plausibility of the stress index

Assessing the plausibility of the stress index is probably the most problematic step of the process, since, by definition, the real stress sequence is not known. Illing and Liu (2003) suggest comparing the computed index with the results of experts' description and evaluation of the historical stress level. Identifying crises by using experts' assessments is relatively common in the literature. Caprio and Klingebiel (1996), Dziobek and Pazarbasioglu (1997), Bordo and Eichengreen (1999) and Lindgren et al (1996), for example, have used this technique. Unfortunately, the detected crises may vary from one expert to the other. After comparing several studies, Frydl (1999) and Eichengreen and Arteta (2000) conclude that the timing of crises differs significantly from one study to the next. Experts' opinions can obviously not be taken as the "true" value of stress, but they can still be used to assess the plausibility of our results.

For the Swiss case, it is generally agreed that, in the last 20 years, the banking sector has known two periods of high stress: at the beginning of the 1990s, when the industry went through a period of major restructuring and takeovers among regional banks, and in 2001-02, during the stock market crash. A shorter period of stress is also likely to have occurred in 1998, with the Russian and the LTCM crises.

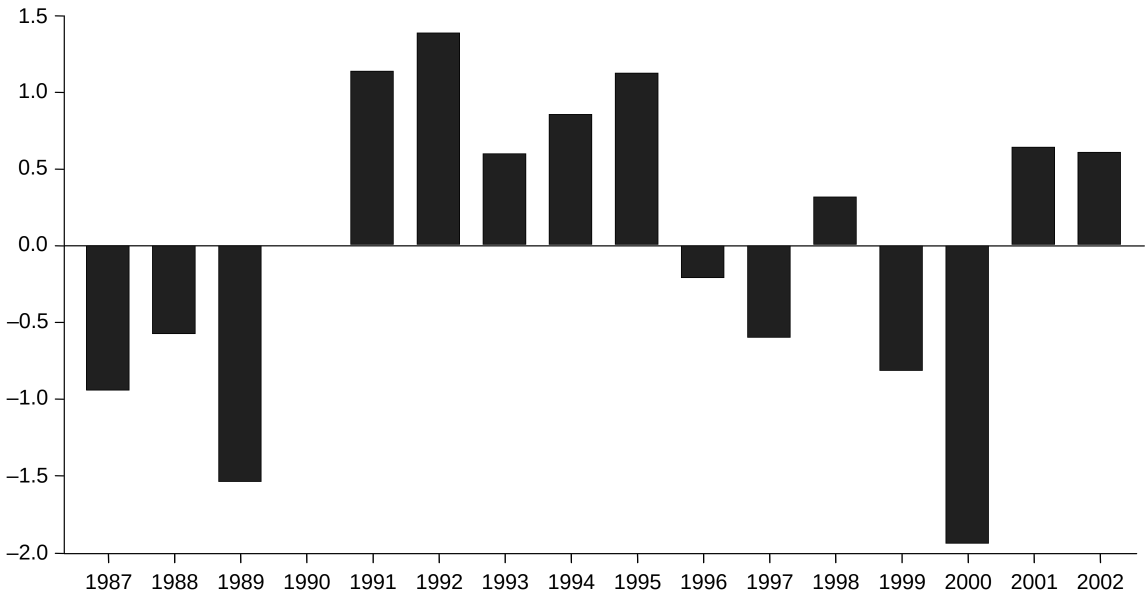
2.7 Results for the Swiss case

Estimation of the stress index

Graph 1 shows the evolution of the computed stress index for the Swiss banking sector between 1987 and 2002. A level above zero means that the stress is higher than average. The index is expressed in terms of standard deviations from its mean.

Graph 1

Stress index for the Swiss banking sector



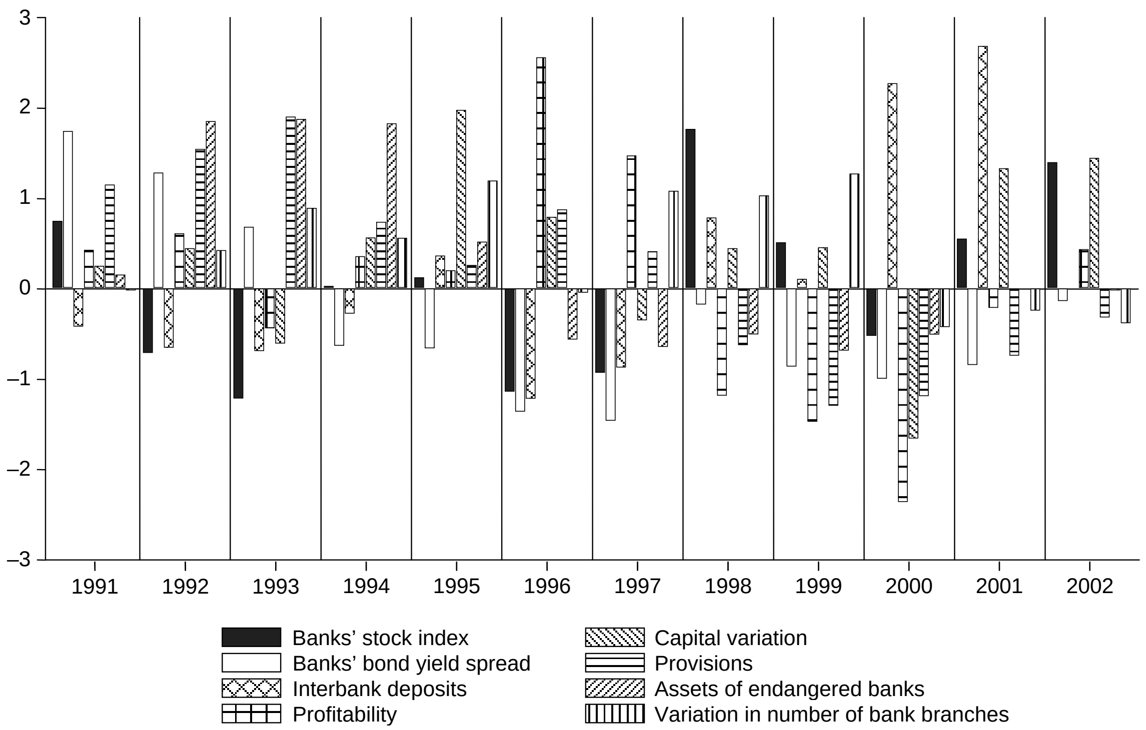
The index identifies three periods where the stress is above average: from 1991 to 1995; in 1998; and in 2001-02. This corresponds to the description of the Swiss banking sector that is commonly given by experts. The highest degree of stress is observed in 1992 and, globally, the beginning of the 1990s was the worst period for Swiss banks. The stress in 1998 was less intense than in the two other stress periods. Conversely, the year 2000 was the least stressful for the banking system in Switzerland.

Decomposition of the index

It is possible to decompose the stress index and isolate the contribution of each factor to global stress. Graph 2 presents the decomposition of our index. A positive (negative) value indicates that the variable is above (below) its sample mean and that it indicates more (less) stress to the system than it does on average.

Graph 2

Decomposition of the stress index (1991-2002)



The decomposition shows that the stress at the beginning of the 1990s is reflected in most of the variables: between 1991 and 1995 (with the exception of 1993), there are always at least six variables out of eight that indicate more stress than the average. The most recent stress period (2001-02) is the reflection of a decrease in banks' stock prices and in their capital (plus a drop in interbank deposits for 2001).

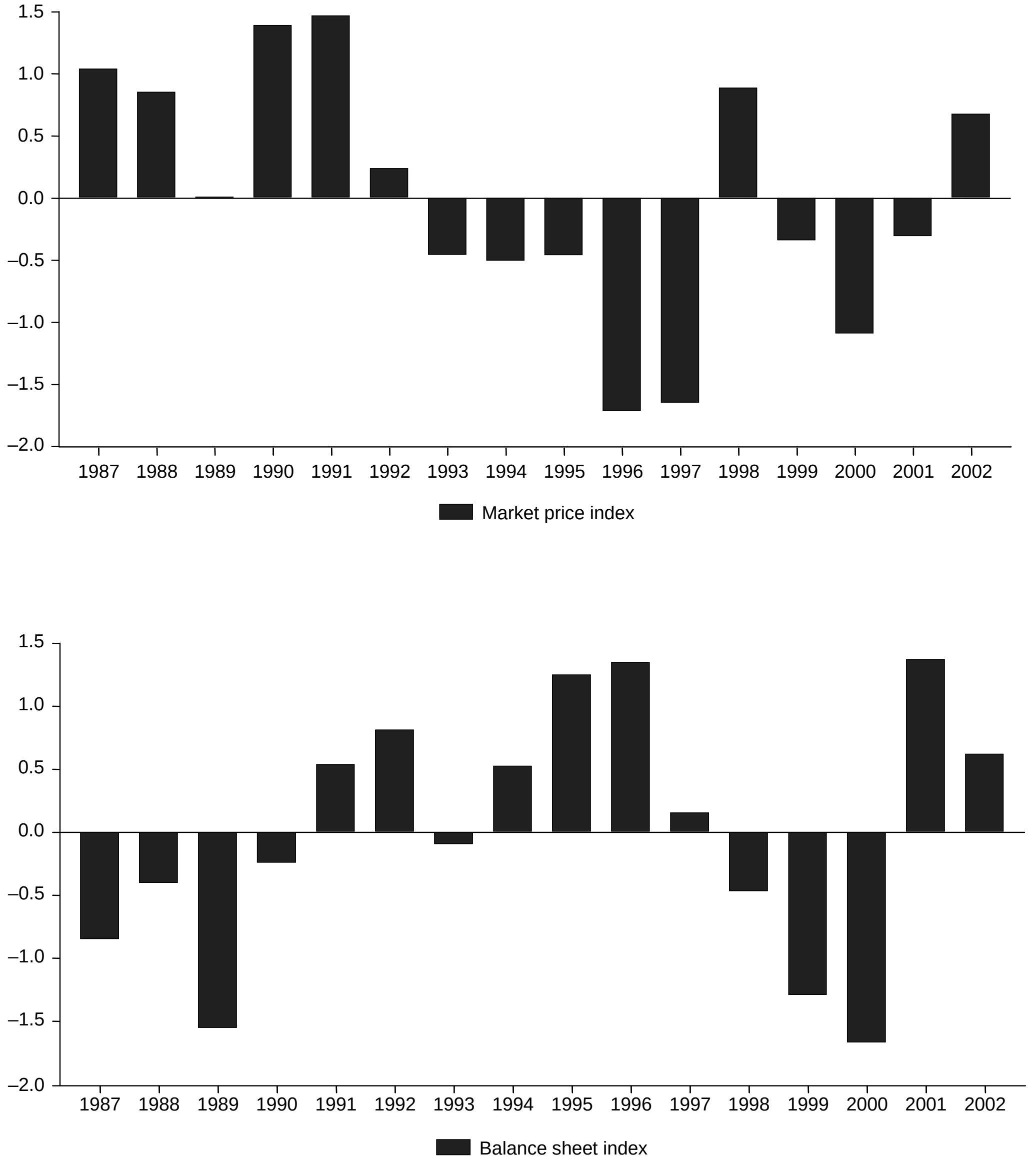
Market prices vs balance sheet data

In the literature, two types of information are commonly used to identify banking crises: market prices and balance sheet data. Indices usually rely upon either one type of variable or the other. Graph 3 presents the results for our stress index when only market prices or balance sheet data are used respectively.

Graph 3 clearly shows that the identification of stress periods largely depends on the type of variables used. The "market price index" shows four periods of stress (1987-88, 1990-91, 1998, and 2002), with the highest value for the years 1991 and before. The stress in 1998 is also important. The "balance sheet index" spots the beginning of the 1990s, with a maximum in 1996 and a break in 1993, and the years 2001-02, with an overall maximum in 2001. With the exception of 1991 and 2002, the two indices differ significantly. They also fail to give a stress pattern that fits the story related by experts. This result suggests that market prices and balance sheet data each provide only one part of the general picture. For example, the market price index, which reflects the situation of the quoted institutions, does not detect the stress at the beginning of the 1990s, which affected mostly non-quoted regional banks. Thus, combining both sources of information improves the index by allowing several symptoms of stress to be identified.

Graph 3

Market price index vs balance sheet index



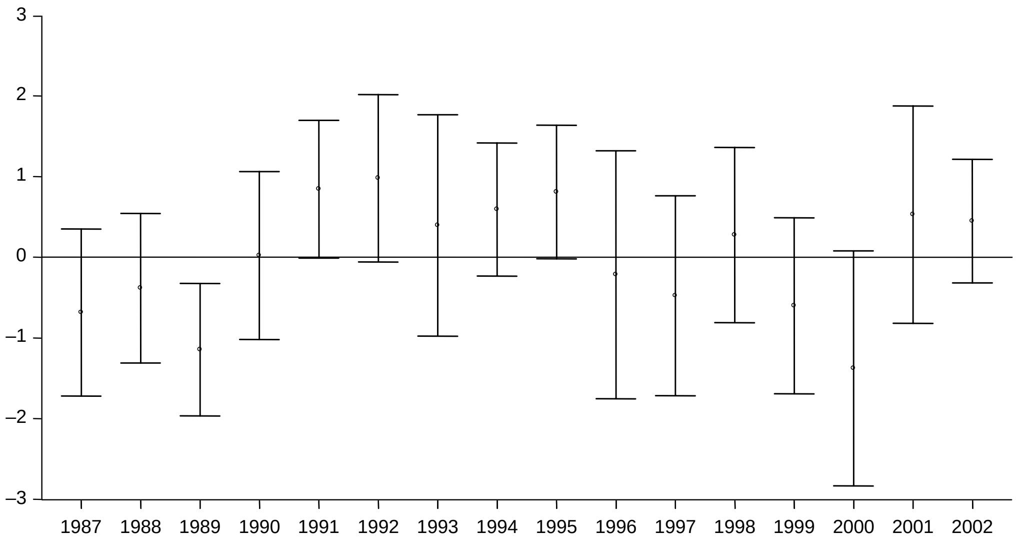
Robustness of the index

To check how sensitive our results are to the choice of variables, we computed indices using all possible variable combinations. This shows whether the results of the different combinations gather around the same value or whether they are widely spread. Graph 4 shows the truncated ranges of these indices (for each year, we exclude the highest and lowest 5% of values).

The index value seems to depend relatively strongly on the mix of the variables, as the results concerning market prices and balance sheet data in Section 2.7 have already suggested. However, globally, it is still possible to distinguish a higher level of stress at the beginning of the 1990s, in 1998 and in the last two years.

Graph 4

Robustness of the index



2.8 Limitations of the index and possible improvements

The principal limitation of our stress index is its low frequency. Unfortunately, most of the balance sheet data are collected annually and, therefore, the index can only be updated on a yearly basis. Furthermore, some of our data are only available since 1987, which makes our series relatively short. The lack of sufficient observations is problematic for the forecasts, as we will see in Section 4. However, our index is probably able to capture significant stress episodes even with an annual frequency, since, according to Frydl's (1999) survey, banking crises last on average between two and a half and four years.

The weight attributed to each variable in the final index is another controversial point. We give an equal weight to all variables, but one could argue that some variables are more relevant for stress than others and, therefore, that they should have a larger weight in the index. Unfortunately, we did not find an alternative weighting scheme that is economically more appropriate.

Another technical shortcoming of our approach is that it does not take into account the skewness of some variables for their standardisation. A potential improvement would be to use a standardisation method which incorporates this characteristic. Bordo et al (2000) propose a measure based on the median rather than the mean, and Illing and Liu (2003) mention a method based on the observed quantiles. These options should be explored in future studies.

Finally, it is also possible to refine the definition of the variables included in the stress index. As an example, Illing and Liu (2003) use a measure of banks' share price relative to the overall stock price in order to distinguish idiosyncratic shocks from economy-wide shocks. The authors also exploit GARCH techniques to take into account the serial correlation of many price series. This method is probably less useful for yearly variables, as the serial correlation tends to decrease with frequency for financial data. In any case, one should carefully consider the opportunity of pure technical improvements, as they tend to complicate the interpretation of the index.

3. Macroeconomic imbalances as early warning signals

Once the level of stress in the banking sector is assessed, another challenging question arises: is it possible to predict stress in the banking sector? A reliable estimate of future stress or at least of its variation - whether the stress will increase or decrease in the medium term - could represent a useful input for periodical assessments of the banking sector's condition. If the stress can be predicted, the supervisory authorities and the central banks might consider actions to prevent serious problems in the banking system.

In the last couple of years, a vast literature has emerged on the so-called early warning systems (EWSs). There are two distinct types of model.[6] The first type of model is based on the "micro" approach and typically projects individual bank failure. The data used for the micro approaches stem from individual banks' balance sheets: they generally include indicators for capital adequacy, asset quality, earnings and profitability, liquidity, and sensitivity to market risk. The second type of EWS model is based on a "macro" approach and aims at the early detection of systemic banking crises. The data used in the macro EWSs are essentially macroeconomic variables. More recently, financial and political indicators have also been included in the estimations of macro approaches. Because our stress index is defined to reflect the stress of the overall banking sector, and not the stress of individual institutions or certain bank categories, the macro EWS approaches seem to be more suitable in our case.

Two findings of the macro EWS models are worth mentioning. First, the models have established empirical evidence of the link between the real economy and the financial system. In other words, the economic environment and prevailing macroeconomic imbalances can, to a substantial degree, influence stress in the banking system. Demirgüç-Kunt and Detragiache (1998) conclude, in a study on both developing and developed countries, that banking crises tend to emerge when the economic environment is weak. Second, the variables reflecting a weak economic environment and the build-up of imbalances seem to have the power to predict the condition of the banking sector. These two findings motivate our choice to rely on the economic environment and imbalances to forecast the stress index.

3.1 Macroeconomic imbalances and available information

Traditionally, researchers use macroeconomic variables in levels or their growth rates to predict banking crises. We rather focus on macroeconomic imbalances, which can be considered as a common risk exposure of financial institutions, and which have the potential to create stress in the future. We use the gap approach developed by Borio and Lowe (2002a). Technically, a macroeconomic imbalance is the gap between the original series and its trend. A positive (negative) deviation means that the actual series lies above (below) its trend. We assume that the trend is the proxy for the longer-term fundamental value of a variable, around which the actual series fluctuates. Admittedly, the assumption that the fundamental value can be "correctly" determined is controversial. We believe that even with a pragmatic approach to calculate the trend, we should still be able to broadly observe the imbalances, which usually follow cycles of several years.

Using gaps puts the focus on cumulative processes, since macroeconomic imbalances can build up either through a strong above (below) trend growth one period or through a sequence of years with small above (below) trend growth. The larger and more numerous the macroeconomic imbalances are in an economy at the same time, the more likely it is that the stress in the banking sector will rise later. It is important to note that it is not the build-up of imbalances but their sudden unwinding which can cause disruptions in the economy leading to higher stress for the banking sector.

To calculate the trend and the gap at time *t*, we use only the information that was available at that particular point in time (ie the gap for 1990 is estimated with the data for 1990 and the preceding years). This means that we place ourselves in the position of the policymakers at time *t* and we do not take advantage of the information contained in the full sample. The estimation of the trend is revised every year as new information is added to the sample.

3.2 Choice of explanatory variables

The list of macroeconomic variables used in earlier studies to predict banking crises is rather long. In order to choose the "right" indicators that could best predict the stress index for Switzerland, the variables should have a significant influence on the condition of the banking sector and should have

[6] See, for example, Bell and Pain (2000) for a survey of the two groups of models.

proved to be robust across a number of other studies. Moreover, the variables should be related to typical macroeconomic imbalances in industrialised countries.[7] The variables we selected in this respect are the following: the share price index (SP), the housing price index (HP), the credit ratio (CRR) (private credits/GDP), the investment ratio (INVR) (investments/GDP), the gross domestic product of Switzerland (GDP) and euro area GDP (GDP Europe). All variables in our data set are nominal and on an annual basis. The data were collected for the period between 1970 and 2002.[8] In the following paragraphs, the main intuition for the impact of each variable on the stress is given and earlier studies that include these variables are mentioned.

Share price index and housing price index

A steep rise in asset prices (both share and housing prices) may trigger a wealth effect, which fuels consumption and subsequently leads to stronger economic growth. But when asset prices suddenly swing back, negative consequences for the banks have to be expected. The stress index could rise, especially when the banking system is already weak.

In the context of falling share prices, banks' profitability will most likely decrease, for example because commission and trading income sink and profits from the banks' own asset holdings will be lower. In the wake of falling asset prices, the balance sheets of firms and households will become weaker too. For the banks this means that more loans could end up non-performing. The banks' capital position will become weaker because of higher non-performing loans, but also because of a reduction of the value of the banks' own share holdings. Borio and Lowe (2002a,b) note that share price booms predict banking crises in a sample of 32 countries, including the G10 countries.

Households and firms typically hold a large fraction of their wealth in real estate. During a housing price boom, the debt capacity (in terms of mortgage loans) rises. Perhaps the banks' willingness to lend will also be greater. A decline in housing prices reduces the value of loans collateralised with real estate. Consequently, a higher number of defaults by mortgage-financed real estate owners will increase the number of non-performing loans for the banks. As in the case of declining share prices, an increasing amount of non-performing loans means that the profitability of the banks will be lower, and accordingly the stress will be higher.

Credit ratio

For an economy that is growing, it is normal that credits are increasing. But if credits grow much faster than GDP, this could mean lower lending standards on the part of the banks, ie loan applications are not adequately analysed. A rise in the credit ratio, which is the amount of credits to the private sector divided by GDP, could be associated with higher risk-taking by the banks in their lending business. The imbalance will start to unwind when borrowers find it more difficult to service their debt (for example because of a rise in interest rates or because the economy enters a recession). The more and the longer the credit ratio deviates from its longer-term trend - and assuming that the banks cannot fully hedge the credit risk - the more likely a rise in stress in subsequent periods. Borio and Lowe (2002a) point out that credits are a good indicator for crisis prediction. Eichengreen and Arteta (2000) conclude that rapid credit growth is among the most robust causes for banking crises, but their sample contains mostly emerging market countries.

Investment ratio

Investments have less frequently been used in macro EWS models. Nevertheless, we decided to use the investment ratio in our forecast of stress as it gives us a broader view of possible prevailing macroeconomic imbalances in the economy. Overinvestment can lead to losses for corporations and also for banks when the projects do not achieve their planned return on investment. Hardy and Pazarbasioglu (1999), in a study on leading indicators for the Asian crisis in the 1990s, include the

---

[7] For simplicity, we use the term macroeconomic imbalances in this paper, although some of the imbalances (eg in asset prices) are sometimes referred as financial imbalances.

[8] For more details, see the data sources in the appendix.

investment ratio in their data set. Unfortunately, as mentioned for example by Borio and Lowe (2002a), the variable is not always robust.

GDP and GDP Europe

The evolution of GDP reflects the general condition of an economy. If the performance of the economy is weak, the credit standing of borrowers deteriorates. The number of corporate and private defaults rises during a recession and leads to an increase in the share of non-performing loans and in risk provisioning for banks. Especially for banks that are strongly engaged in the lending business, a deep and long-lasting recession could lead to higher stress. Demirgüç-Kunt and Detragiache (1998) find in a sample of 65 and 45 countries respectively that low real GDP growth increases the probability of a banking crisis in the period between 1980 and 1994. Eichengreen and Arteta (2000) conclude that GDP growth rates generally decline before a banking crisis. The evidence is somewhat weaker when the authors limit their sample to OECD countries.

For Switzerland, being a small open economy, the business cycle is influenced to a considerable degree by the international environment. For this reason, we include euro area GDP (GDP Europe) in our data set. Swiss banks are directly affected by the European business cycle, as part of the banks' income is generated in that region. Indirectly, Swiss banks are affected by the impact of the European business cycle on domestic prospects. A recession in Europe could lead to a deterioration in economic conditions in Switzerland and thus reinforce a recession in Switzerland with the consequences described in the paragraph above.

For equity and housing prices and for credit and investment ratios, a positive gap is a priori expected to predict stress. The assumption for the GDP and the GDP Europe gap differs from the assumption for the other gaps as a recession is associated with a negative gap. One could also argue that a positive GDP gap might be an early warning signal for stress, indicating that the economy is on an unsustainable expansion path. The unsustainable expansion will sooner or later have to be corrected, which could lead to a rise in stress for the banks. From the empirical literature on banking crises, however, it is known that banking crises tend to occur when there is a recession (Borio and Lowe (2002b)).

3.3 Method for calculating gaps

For each series (GDP, GDP Europe, SP, HP, CRR and INVR), we take logs and calculate a rolling Hodrick-Prescott filter (HPF) trend. We use a "rolling" filter because we only rely on information that was actually available at each point in time. The gap is the actual series of a variable minus the trend. All gaps are then standardised in order to measure their relative size:

$$s\_gap_t = \frac{g_t}{\sigma_t}$$

The standardised gap (s_gap_t) at time t is equal to the gap (g_t) at time t divided by the standard deviation (σ_t) of g. The standard deviation corresponds to the standard deviation from the starting date (1970) of our sample up to time t. It is replaced with a more recent calculation when the sample is extended.

4. Forecasting stress in the Swiss banking sector

This section presents how we have chosen the best model to forecast stress. The main results include the estimation output, the forecasts and a discussion on the robustness. The limitations of our method and possible improvements are then given in the last part of this section.

We base our forecast of stress on a regression of our stress index on the standardised gaps described in the previous section. We focus on one-year-ahead forecasts (one period in the stress index series). Hence, at time t, the model gives a forecast for $t+1$. We estimate a model which regresses the stress index (y_t) on the k observed standardised past gaps ($x_{k,t-z_k}$). Only one lag (z_k) per gap is used in the model.

$$y_t = \beta_1 x_{1,t\text{-}z_1} + \beta_2 x_{2,t\text{-}z_2} + \ldots + \beta_k x_{k,t\text{-}z_k} + \varepsilon_t$$

4.1 How to find the best model

To pick the best model among all possible combinations of variables and lags, we use two types of criteria. First, the model has to fulfil the following plausibility criteria: (1) the regression coefficients must be significant at a 10% level and have the right theoretical sign; (2) no lag greater than four years should appear in the model;[9] and (3) the model must contain at least three explanatory variables. Second, we use the following four efficiency criteria to distinguish the best model among all those that fulfil the plausibility criteria: (1) the *R*-squared; (2) the number of correct forecasts for the direction of the stress index variation; (3) the root mean squared error (RMSE) for an out-of-sample forecast on the last seven years; and (4) the number of correct out-of-sample forecasts for the stress index variation in the last seven years.

4.2 Main results

Estimation output

According to the criteria mentioned above, one final model has emerged. This model has the best *R*-squared, it has the best RMSE, it does not make any error in forecasting the direction of the stress index variation in-sample, and it has the second best performance in the out-of-sample forecast of the stress index variation. The results are presented in Table 1.

Table 1

Determinants of the stress index

Variable	Coefficient	Standard error
GDP gap (–1)	–1.761*	0.738
GDP Europe gap (–3)	–1.297*	0.386
Share price index gap (–4)	1.452**	0.233
Housing price index gap (–3)	2.132**	0.468
Credit ratio gap (–2)	2.185**	0.617
Investment ratio gap (0)	0.893*	0.305
Constant	–0.897**	0.229
Number of observations	16	
R-squared	0.89	
Errors in-sample	0	
Root mean squared error (RMSE)	0.67	
Errors out-of-sample	1	

* Significant at the 5% level. ** Significant at the 1% level.

As shown in Table 1, the coefficients are significant at the 5% level. The *R*-squared with 0.89 is relatively high. The GDP gap and the GDP Europe gap - reflecting the general economic environment - each have a negative coefficient, which can be interpreted as a recession. The other variables, the

9 We considered a link between stress and six-year-old gaps to be theoretically implausible.

macroeconomic imbalances, have positive coefficients. The lags of the explaining variables (in brackets after the variables) are between 0 for the investment ratio gap and –4 for the share price index gap. Remember that at time *t* the stress index for the subsequent period *t*+1 is estimated. The lag of one year for Swiss GDP is relatively short and it shows that the economy enters a recession only shortly before the stress in the banking sector rises. The lag of GDP Europe with –3 is longer than the lag of Swiss GDP, indicating that the impact of the European business cycle takes longer to materialise in terms of stress.

Although the lag of –4 for the equity price index gap seems to be surprisingly long, it is consistent with the findings of Borio and Lowe (2002a). They note that the equity price gap indicates banking crises better if a lag of –5 years is used.[10] To our knowledge, many studies on macro EWS models consider a maximum time window of two years before a banking crisis. As we allow for longer lags to predict the stress in the banking sector, it is somewhat difficult to compare our results with those that use shorter time windows.[11]

The main results of our model can be summarised as follows: (1) we find that macroeconomic imbalances and the economic environment have an influence on future stress in the Swiss banking sector; (2) the gaps and the concept of the macroeconomic imbalances seem to be a useful method to detect early warning signals for stress. If we estimate our model with variables in levels or in differences, the explanatory power of the model drops; (3) the macroeconomic imbalances generally build up years before the stress rises in the banking sector; and (4) in our model, the lags for the macroeconomic imbalances are between one and five years. For the share price, the credit ratio and the housing price, the lags are longer than two years (the time horizon frequently used in other macro EWS models).[12]

Forecasts

Graph 5 shows the actual stress index value (bars), the in-sample forecasts for 1987 to 2003 given by our model (line) and its out-of-sample forecasts for 1996 to 2003 (dotted line). Our model predicts the major stress periods (beginning of the 1990s and 2001-02). It forecasts the tranquil periods at the end of 1980s and the end of the 1990s. On the other hand, it fails to foresee the stress episode in 1998. The model gives the right direction in every case for the in-sample forecast and makes one error for the out-of-sample one. The model predicts an amelioration of the Swiss banking sector's condition for 2003.

Robustness of the model

To assess the robustness of our model, we compare it with other specifications that successfully met our plausibility criteria. Three variables appear almost constantly in all these models: the equity price index, with a five-year lag; housing prices, with a four-year lag; and finally, but slightly less frequently, the credit ratio, with a lag of two or three years. The three other variables are not always significant and they do not always come out with the same lag. Their significance and their lags depend on how they are combined together.

---

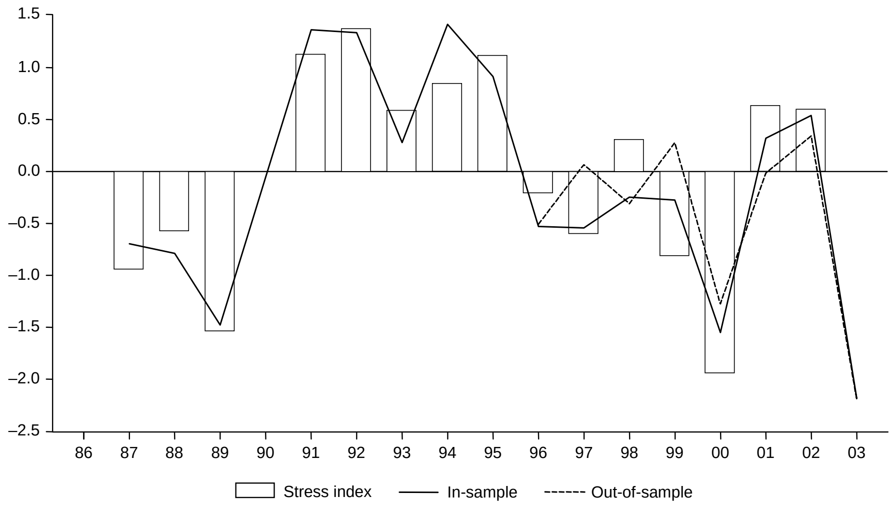
[10] Borio and Lowe (2002a) predict future banking crises with a horizon of up to three years. The equity price gap in their model has a lag of –2, so the "total" lag of the share price variable corresponds to –5 years.

[11] A summary of the results of macro EWS models and the time horizons that were used can be found in Eichengreen (2002).

[12] We also tried to include more international indicators (the Swiss franc/euro exchange rate and the MSCI global stock index) and their gaps respectively in our estimations; however, they were not significant. In a different approach, we calculated an index for the total macroeconomic imbalances by summing up the standardised gaps. The stress index was then regressed on the imbalances index. Again this did not produce satisfactory results. Finally, we took real variables instead of nominal and re-estimated our model. This deteriorates the results of our model. The reason for that might be that banks write contracts based on nominal terms rather than on real terms. Changes in the γ, the smoothing parameter of the HPF, merely alter the results as long as γ is larger than 500 for the share price index (SP) and for the housing price index (HP). The smoothing parameter reflects the average size of the gap between the actual variable and its long-term value (Hodrick and Prescott (1997)). Assuming a higher smoothing parameter for the equity price index indeed makes sense because this variable seems to deviate more from its long-term trend than other variables.

Graph 5

Stress index forecasts



The last finding is a typical symptom of multicollinearity between explanatory variables. Multicollinearity has two important consequences for us: (1) it could artificially push the *R*-squared upwards, which can induce us to select the wrong model; and (2) it might bias the value of regression coefficients, which is used as a plausibility criterion. To check if multicollinearity is a problem in our case, we computed the BKW statistic (Belsley et al (1980)) for the different models. The higher this statistic is, the more harmful the multicollinearity for the biases mentioned above. According to Belsley et al (1980), a value greater than 30 can be considered critical. All our selected models have a statistic included in a range going from two to 21, with a value of 17.0 for our final model. According to this criterion, the consequences of multicollinearity are likely to be small in our case.

Note that multicollinearity does not affect the forecast accuracy of a model, as long as the linear relation between the explanatory variables observed in the past remains identical in the future. Therefore, even if the model's coefficients are not fit for explaining the causes of stress, the forecasts are still usable. Unfortunately, due to the shortness of our stress series, we are not able to draw any conclusion on the stability of this linear relation.

Robustness of the forecasts

A crucial question, especially for policymakers, is: how robust are the forecasts? In other terms, how do changes in our basic specification affect the predictions? To check the forecast robustness, we compared the out-of-sample forecasts of the 16 models that fulfil the plausibility criteria. By looking at the dispersion of the forecasts, we obtain an idea of the degree of similarity between the models.

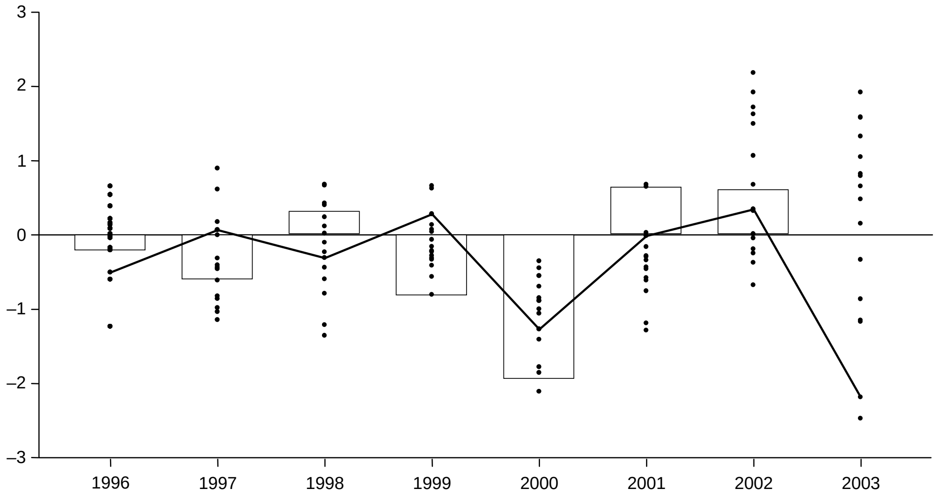
Graph 6 shows the actual value of the stress index and the out-of-sample forecast for each model (symbolised by dots). The solid line represents the forecasts of our basis model. For each model, the mean of the forecasts is not significantly different from zero, which implies that none of them systematically over- or underestimates the stress level.

One can split the results into two periods. Between 1996 and 2001, the dispersion of the forecasts is relatively constant. The standard deviation of the forecasts' distribution varies between 0.39 and 0.64, with an average of 0.54. The variance equality hypothesis between these years cannot be rejected. The dispersion is much higher in 2002 and 2003, with a standard deviation of 0.92 and 1.35 respectively. The variance equality hypothesis is clearly rejected when these two years are taken into account.

Two main conclusions can be drawn from these results: (1) globally, there is great uncertainty about the forecasts, since different models give very disparate predictions; and (2) this uncertainty seems particularly important for 2003. Indeed, the dispersion of the out-of-sample forecasts for this year is clearly greater than in the previous years.

Graph 6

Dispersion of the out-of-sample forecasts



4.3 Limitations of the method and possible improvements

The main problem of our forecast is that our stress index series is too short. Due to the frequency and the availability of the variables included in the index, we have only 17 observations for the banking sector's stress. This is not sufficient to obtain a robust estimation of our model. One way to increase the number of observations would be to drop the variables that have a low frequency or those that have been available for only a few years. The disadvantage of a reduced index is that it might be less accurate in measuring the stress (see Section 2.7). To avoid discarding some variables, it is also possible to artificially increase the frequency by simulating the evolution of a variable between two observable points. With this technique, it is possible to exploit all the information contained in the high-frequency variables without abandoning the low-frequency variables. Finally, another possibility to obtain more stress observations is to use a cross-country analysis. Unfortunately, in this case, some variables might not be available for every country, eg the Swiss Federal Banking Commission's list of banks under special scrutiny. Furthermore, a model based on a cross-country analysis would not reflect the particularities of the Swiss banking sector.

Another problem with our method lies in the estimation of trends. The HPF recognises changes in the direction of the trend only with some delay, which means that the estimated trends are less reliable at the end of each observation period. In addition, the HPF is a mechanical trend and only one possible way to identify the fundamental value of a variable. The HPF might not be the best way to do this, especially for variables such as asset prices.

Although the macroeconomic imbalances and the weak economic environment can explain a substantial part of the banks' stress, they are not the sole factors behind systemic banking sector problems. Other factors, such as deregulation or restructuring due to overbanking, could also play a role regarding banking sector stress. But they are not included in our model.

Our method also supposes a linear relation between the stress and the gaps. If this was not the case in reality, our model would rely on a spurious link between the variables and the quality of the predictions would be altered. As underlined by Eichengreen (2002), banking crises are complex phenomena which certainly involve non-linear interactions. This might explain the lack of robustness in our forecasts.

Finally, the fundamental problem of our method is that a great amount of uncertainty intrinsically exists over both the measure of the explained variable (the stress index) and the relation between the explained and the explanatory variables. In a way, it is like solving one equation with two unknowns (the stress and the model). Our solution to this problem is to implicitly assume that we had a perfect measure of the stress level and thus reduce the number of unknown variables to one (the model). Unfortunately, this is not the case, and if large discrepancies between our index and the "true" banking sector stress exist, our estimated model and forecasts could be biased.

5. Conclusion

The aims of this paper have been: (1) to develop a measure that summarises the banking sector's condition in an industrialised country such as Switzerland; and (2) to provide a framework that could help policymakers to predict the development of the banking sector's condition.

Our stress index estimates the banking sector's condition on a continuous scale ranging from tranquil periods to severe crisis. It distinguishes itself from the traditional binary indicators found in the literature because it describes a continuous range of states rather than just differentiating crises from tranquil periods. This characteristic makes it particularly appropriate for depicting banking sectors which rarely (or never) experience severe crises. To our knowledge, it is the first time that a stress index focusing on the banking sector has been constructed.

Our stress index is an aggregation of several variables, each of them being a potential symptom of banking crises. Our assumption is that the more intense these symptoms are, the higher the stress. We combine different types of variables: market prices, balance sheet data, non-public information and other structural data. The estimated index fits well the experts' evaluation of the Swiss banking sector's history and it identifies all major stressful periods. We find that the value of the index differs substantially when only one type of information (market prices or balance sheet data) is used and that, in this case, it fails to detect the entire sequence of stress episodes. This confirms the fact, widely acknowledged in the literature, that banking crises can show up in different ways. A concrete implication is that, in order to detect the different forms that a banking crisis can take, a stress index should be constructed on several variables and incorporate different types of information.

After estimating the stress index, we try to forecast it by using macroeconomic imbalances. We identify the imbalances by computing the gap between a variable and its trend. The advantage of this method is that it exhibits the accumulation of yearly imbalances rather than focusing on the variable for one year only. We then regress the banking sector's stress index on the gaps and use the estimation equation to forecast the stress index.

Our main finding is that a model incorporating Swiss and European GDP, credit and investment ratios to national GDP, the stock price index and housing prices is able to explain a large part of the Swiss banking sector's stress level. This indicates that a significant link exists between the (macro)economic environment and the banking sector's condition. Furthermore, we find that the gaps precede the evolution of the stress and that they can be used by policymakers to forecast the stress level. We observe that the lag between the gaps and the stress index could go from one up to five years (eg share price index). This confirms the result in Borio and Lowe (2002a) and, more generally, suggests that long lags are useful for early warning systems. Previous studies usually focused on shorter lags, generally of one or two years. From a technical point of view, we also observed that using gaps, instead of variables in level or in difference, significantly improved the quality of the results. Finally, we tested the robustness of our specification and our forecasts. We find that the coefficient's significance varies with combinations of the different variables, which is a typical symptom of multicollinearity. We also observe that the forecasts are clearly dependent on the model's specification and that they are dispersed around the actual value, which makes accurate predictions difficult.

The main drawback of our model is that it relies on a small number of observations for the stress index. We think that the biggest improvement to this study would be to apply this method to a larger sample. Including more observations would probably decrease the uncertainty about the coefficients' significance and improve the robustness of the forecasts. Three options are available to enlarge our data set. First, one can compute the stress index with high-frequency variables only. Second, one could extend the study to other countries. For both options, some variables would have to be dropped from the stress index's estimation (the variables with low frequency or those that are not available for countries other than Switzerland). It is possible to construct a valuable measure of the stress with fewer variables, but, as the results of this paper show, it is important to include as much information as possible in the index to identify the multiple patterns that a banking crisis can follow. The third option is to artificially increase the frequency of some variables by estimating the values that are not directly available. This would allow incorporating the information on high-frequency variables without discarding the low-frequency ones.

Another significant improvement would be to choose a more sophisticated measurement of the gaps. As mentioned in the main text, our method tends to identify imbalances with some delay. The use of an empirical macroeconomic model to compute long-term equilibrium trends would probably refine our estimation of macroeconomic imbalances and detect them earlier than the actual method does. However, we believe that the most significant improvements could be made by increasing the number of stress index observations used in our estimation.

Appendix:
Data sources

Stress variables

Swiss banks' share price index: SWITZ DS Banks (Thomson Financial Datastream) 1984-2002, weekly data. The largest 12-month price fall observed each year is computed (CMAX index).

Swiss banks' bond yield spread: spread between the banks' bond yield index and the Confederation bond yield index (1984-2000), spread between banks' and Confederation three-year bond yield (2001-02). Weekly data (Swiss National Bank). The annual average is computed.

Total amount of interbank deposits: total amount of interbank deposits, monthly data (Swiss National Bank) 1980-2002. The largest 12-month fall in deposits observed each year is computed (CMAX index).

Banking sector's rate of profitability: ratio of Swiss banks' aggregated profit to their total assets, yearly data (Swiss National Bank) 1982-2002.

Banks' total capital variation: difference in Swiss banks' aggregate capital from one year to the other, in per cent, yearly data (Swiss National Bank) 1987-2002.

Banking sector's provision rate: ratio of aggregated new provisions (and amortisation) to aggregated total assets, yearly data (Swiss National Bank) 1987-2002.

Total assets of banks under special scrutiny: aggregate total assets of the banks under the scrutiny of the Swiss Federal Banking Commission, yearly data (Swiss Federal Banking Commission and Swiss National Bank) 1987-2002.

Variation in bank branch numbers: difference in the number of bank branches from one year to the other, yearly data (Swiss National Bank) 1987-2002.

Macroeconomic variables

Share price index (for Switzerland): SWITZ DS Market (Thomson Financial Datastream) 1970-2002, quarterly data that have been converted into annual data by taking the average of the quarters for each year.

Housing price index (for Switzerland): constructed by taking the mean of the growth rates of the following subindices (Wüest & Partner): office floorspace, apartments, houses, industrial space, new and old rented flats, 1970-2002, quarterly data that have been converted into annual data by taking the average of the quarterly growth rates for each year.

Credits (for Switzerland): claims on the private sector, *International Financial Statistics* (IMF), 1970-2002, quarterly data that have been converted into annual data by taking the mean of the quarters for each year.

Investments (for Switzerland): gross fixed capital formation, *International Financial Statistics* (IMF), 1970-80, and gross domestic investment (State Secretariat for Economic Affairs), 1980-2002, quarterly data that have been converted into annual data by taking the sum of the quarters for each year.

Gross domestic product (Switzerland): Swiss GDP, current prices (Swiss Federal Statistical Office) 1970-2002, annual data.

Gross domestic product (Europe): euro area GDP, current prices, *OECD Main Economic Indicators*, 1970-2002, annual data.

References

Bell, J and D Pain (2000): "Leading indicator models of banking crises - a critical review", *Financial Stability Review*, Bank of England, December.

Belsley, D A, E Kuh and R E Welsch (1980): *Regression diagnostics: identifying influential data and sources of collinearity*, John Wiley & Sons, New York.

Bordo, M D, M J Dueker and D C Wheelock (2000): "Aggregate price shocks and financial instability: an historical analysis", *NBER Working Papers*, no 7652, National Bureau of Economic Research.

Bordo, M and B Eichengreen (1999): "Is our current international financial environment unusually crisis prone?", in D Gruen and L Gower (eds), *Capital flows and the international financial system*, Reserve Bank of Australia, Sydney, pp 18-75.

Borio, C and P Lowe (2002a): "Asset prices, financial and monetary stability: exploring the nexus", *BIS Working Papers*, no 114, Bank for International Settlements.

——— (2002b): "Assessing the risk of banking crises", *BIS Quarterly Review*, Bank for International Settlements, December, pp 43-54.

Caprio, Jr, G and D Klingebiel (1996): "Bank insolvencies: cross country experience", *Policy Research Working Paper* no 1620, World Bank.

Collyns, C and A Senhadji (2003): "Lending booms, real estate bubbles, and the Asian crisis", in W Hunter et al (eds), *Asset price bubbles: the implication for monetary, regulatory and international policies*, MIT Press, Cambridge and London, pp 101-25.

Corsetti, G, P Pesenti and N Roubini (1998): "Paper tigers? A model of the Asian crisis", *NBER Working Papers*, no 6783, National Bureau of Economic Research.

Demirgüç-Kunt, A and E Detragiache (1998): "The determinants of banking crises in developing and developed countries", *IMF Staff Papers*, 45(1), pp 81-109, International Monetary Fund.

Dziobek, C and C Pazarbasioglu (1997): "Lessons from systemic bank restructuring: a survey of 24 countries", *IMF Working Paper* no 161, International Monetary Fund.

Eichengreen, B (2002): *Predicting and preventing financial crises: where do we stand? What have we learned?*, paper prepared for the Kiel Week annual conference, Germany, 24-25 June.

Eichengreen, B and C Arteta (2000): "Banking crises in emerging markets: presumptions and evidence", *CIDER Working Paper* no 115, Center for International and Development Economics Research, University of California.

Frydl, E J (1999): "The length and cost of banking crises", *IMF Working Paper* no 30, International Monetary Fund.

Goldstein, M, G L Kaminsky and C M Reinhart (2000): *Assessing financial vulnerability: an early warning system for emerging markets*, Institute for Financial Economics, Washington DC.

González-Hermosillo, B (1999): "Determinants of ex-ante banking system stress: a macro-micro empirical exploration of some recent episodes", *IMF Working Paper* no 99/33, International Monetary Fund.

Hardy, D and C Pazarbasioglu (1999): "Determinants and leading indicators of banking crises: further evidence", *IMF Staff Papers*, 46(3), pp 247-58.

Hodrick, R J and E C Prescott (1997): "Postwar US business cycles: an empirical investigation", *Journal of Money, Credit and Banking*, 29(1), pp 1-16.

Illing, M and Y Liu (2003): "An index of financial stress for Canada", *Bank of Canada Working Papers*, no 14, Bank of Canada.

Kaminsky, G L and C M Reinhart (1996): "The twin crises: the cause of banking and balance-of-payment problems", *International Finance Discussion Paper* no 544, Board of Governors of the Federal Reserve System.

——— (1999): "The twin crises: the cause of banking and balance-of-payment problems", *American Economic Review*, 89(3), pp 473-500.

Lindgren, C-J, G Gillian and M I Saal (1996): *Bank soundness and macroeconomic policy*, International Monetary Fund, Washington DC.

Vila, A (2000): "Asset price crises and banking crises: some empirical evidence", *BIS Conference Papers*, no 8, pp 232-52.

Sectoral fragility: factors and dynamics

Ivan Alves
European Central Bank[1]

1. Introduction

Business cycles can have a major impact on the credit portfolio of banks. Using re-sampling techniques on data from US banks, Carey (2002) suggests that mean losses during a period of distress such as the 1989-91 recession are 3.5 times larger than during an expansion and about the same as the expansion distributions' 0.5% tail. In terms of the capital that would be adequate for banks, Bangia et al (2002) find that, over a one-year horizon the banks' needs increase by 25-30% in a recession relative to expansions.

This interaction between the business cycle and the quality of banks' asset portfolios motivated substantial empirical research, primarily focused, at an aggregate level, on default rates.[2] Empirical evidence so far shows a strong negative relationship between realised defaults and the economic cycle, but it also suggests that the transmission channels are complex.[3]

In parallel to these results at the aggregate level, efforts to improve the modelling of risk in investment portfolios have produced another body of literature aiming at incorporating aggregate macroeconomic effects in value-at-risk (VaR) models.[4] New developments in this approach also refine the modelling of the cyclical factors, including addressing international and cross-industry correlations, as in Pesaran et al (2003).

Two main motives fostering the development of new, or the improvement of existing, models are the need to provide realistic forecasting models and to test the resilience of a given bank's portfolio. Identifying in advance the impact on credit markets of macroeconomic developments calls for a forecasting infrastructure flexible and realistic enough to incorporate the main macroeconomic components. In the context of lending by the banking sector, advance information on credit risk developments is looked for with keen interest, as this would allow for a more flexible allocation of capital possibly better matching changes in requirements imposed by external developments. In a similar fashion, interest in testing the resilience of the banking sector has increasingly fuelled the incorporation of macroeconomic elements. Stress-testing models analyse the exposure of banks to lending and to credit risks associated with the lending, also including macroeconomic developments. Typically, initial efforts have looked at the behaviour of loan loss provisions or non-performing loans (for a given bank or

---

1 The views expressed in this paper are those of the author and do not necessarily represent the views of the European Central Bank (ECB). The paper benefited from the useful discussions in the Task Force on Stress Testing of the Banking Supervision Committee, and the author is grateful for the insights provided by its members. Special thanks go to John Fell, Darren Pain, Jukka Vesala and Nico Valckx for useful discussions and comments. All errors remain the responsibility of the author.

2 For example, Duffie and Singleton (2003, Section 3) provide an overview of the issues as well as references to related literature. In particular, taking as a starting point the seminal work of Fons (1991) (who considered the impact of GDP growth on the default rate of a number of rating pools of bonds), Jonsson and Fridson (1996), among others, extended the analysis by considering a larger number of macroeconomic variables and the ageing effect in the pools of bonds. Helwege and Kleiman (1996) refine the model further by introducing a new method of gauging macroeconomic effects on default behaviour. Also Wilson (1997a,b) models the impact of macroeconomic variables on the probability of defaults at the industry level.

3 Friedson et al (1997), to cite one example, find a relation between macroeconomic conditions and probability of default. In particular, they find that as real interest rates increase, asset values decrease, thereby increasing the estimate of the probability of default. Furthermore, interdependence may run through financial market shocks, as can be observed in periods of high instability.

4 Recent surveys on this literature include Saunders and Allen (2002), Crouhy et al (2000), Allen and Saunders (2002), Gordy (2000), and Koyluoglu and Hickman (1998).

groups of banks) in response to macroeconomic shocks.[5] The main obstacle of this aggregated approach is the limited granularity and time availability of information on loan loss provisions.

Our objective is to contribute to developing a framework for stress-testing the resilience of the banking sector by exploiting information contained in available credit risk measures. In doing so, we construct an alternative forecasting engine that increases the granularity of past exercises by further breaking down exposure along sectoral lines. More specifically, we exploit the sensitivity of industry-wide expected default frequencies, or EDFs, from Moody's KMV to macroeconomic developments in order to model the evolution of bank portfolios' fragility. Imbalances in the performance of different industries suggest that greater granularity in the treatment of banks' corporate counterparties could increase the measurement accuracy of their fragility.[6]

When pursuing the analysis at a more granular level, it is also important to characterise the linkages interrelating risk across industries.[7] This issue has recently been illustrated by the high degree of correlation and default transmission observed between industries. For example, default rates in the technology and telecommunications sectors often go hand in hand and have generally preceded problems in other industries. Also, albeit in a more subtle manner, other sectors such as insurance and banking can transmit problems to the rest of the economy owing to their important role in financial intermediation. Indeed, one can think of a number of micro- and macroeconomic relations tying the default performance across industries.

The capital allocation decision across sectors can also be seen from the perspective of an "aggregate investor". In the portfolio composed of sectoral assets, the latter are subject to both idiosyncratic (sector-specific) and common risks (such as macroeconomic or geopolitical factors or one-time events). In order to assess the risk for this portfolio, the aggregate investor would optimally account for the risk correlations between the different "assets". Such distinction is also characteristic of VaR modelling of bank portfolios, where very granular information on loan characteristics, including the corporate client sector, is used in specifying the risk correlation within a given portfolio.

Developing a minimum understanding of sectoral risk relationships and their dynamics appears high on the research agenda, but progress is marred by the absence of data and a workable multivariate framework for assessing the nature and extent of the interaction. In principle, the firm-level exercise could be mimicked at the aggregate level, by replicating cross-portfolio channels and thus incorporating risk dynamics between different (aggregate) portfolio elements. However, the lack of sufficiently detailed aggregate and consolidated information on exposures and default rates precludes progress in this direction. Alternatively, information on sectoral lending volumes and asset quality could be analysed in the light of industry-specific developments and wider economic and idiosyncratic shocks, possibly exploiting information on cross-industry channels. Alas, this would also require a substantial amount of information at the sectoral level, currently available only to some bank supervisors, on the sectoral exposures and non-performing assets of financial institutions.

The use of market-based information for assessing sectoral risk is a feasible way forward out of the data bottleneck. Sectoral EDFs provide a measure of sectoral risk similar to that of default rates. In contrast to default rates or information on non-performing assets, EDFs contain information on the ex ante expectation of default as embodied in the asset valuation of a firm, thus making a distinction between latent and realised risk. In addition to being a micro-founded indicator of fragility, EDFs have been observed to have good early warning properties.[8] Being widely used in the assessment of asset

---

[5] Many national banking supervisors have carried out this exercise in one form or another. For instance, Morin (2003) and Trucharte and Saurina (2002) follow this procedure, the latter also incorporating some of the elements considered in this paper.

[6] Moody's, for example, reports insolvencies in the telecommunications and technology sectors during 2001 and 2002 representing about 70% of the total dollar-weighted defaults in Europe. In contrast, the financial sectors (excluding the insurance sector) have experienced a very small number and volume of defaults. In addition, one could expect that non-cyclical sectors are slower to respond to changes in the macroeconomic environment, thus making them less sensitive to aggregate fluctuations.

[7] The theory underlying one such possible interlinkage of default processes is presented in Jarrow and Yu (2001). They model the default intensity as depending on macroeconomic factors and an interdependence term linking firms across industries and sectors.

[8] See Delianedis and Geske (1998) and Kurbat and Korablev (2002) and references in the latter for recent studies of the relationship between EDFs and actual default rates, or Gropp et al (2002) for a comparison applied to the European banking sector.

quality and available for a large number of firms, they can be easily incorporated into a simple econometric model to obtain a sensible measure of the nature, direction and magnitude of risk cross-dynamics, and assist in modelling the future evolution of exposures.[9]

This paper explores both the interaction between risk factors across different economic sectors through time and their joint and idiosyncratic sensitivity to macroeconomic and systemic developments, and therefore represents the first step in setting up a macro VaR model (derivation of probabilities of default and cross-correlations). In particular, the modelling exercise aims to better depict sectoral risk interactions, the duration of shocks, and the relationship between industry-specific risk and macroeconomic activities.

Section 2 presents the cointegrated autoregressive (CVAR) framework used to model the risk interactions as well as the data used. In Section 3 the procedures are explained and the results are derived. The model's forecasting abilities are discussed in Section C. Section 4 concludes.

2. Elements of the model

2.1 A reduced-form dynamic model

The characteristics of the industry risk indicators described above point to the need to account for the strong cross-sectoral risk correlations. One relatively simple model capturing such strong interaction is the so-called *cointegrated vector autoregression* (CVAR) model. This linear representation of a system of interrelated variables is widely used for modelling problems with similar characteristics.[10] Indeed, this model accounts for properties characterising sectoral risk, thus offering a richer representation of the system-wide counterparty risk facing the banking sector.

We consider a set of n industries $i \in I = \{1, \ldots, n\}$ to which the EU banking sector has an exposure $x_i, i \in I$. The risk in industry i is denoted by the random variable r_i. Accordingly, the vector of industry risks is denoted by $\mathbf{r}$ and has dimension n. The various elements of $\mathbf{r}$ interact contemporaneously (the risk level in some industries may serve as a factor to that of others) and through time (difficulties in an industry may affect the originating and other industries with a delay). In addition, sectoral risk clearly depends on the overall macroeconomic environment. The macroeconomic and/or systemic factors are represented in our framework by a vector $\mathbf{y}$ of exogenous processes. Finally, extraordinary events affecting sectoral risk profiles, such as the events of 11 September 2001, can be described by a vector of shock dummies $\mathbf{d}$.[11]

In order to illustrate the elements of the model, consider two industries with a cross-impact taking place over one period only, and whose risk is only affected by one current (not past) macroeconomic variable y_t and one deterministic shock d_t. The risk processes for these two industries can be represented by the following system of equations:

$$r_{1,t} = b_{10} - b_{12}r_{2,t} + \gamma_{11}r_{1,t-1} + \gamma_{12}r_{2,t-1} + c_1 y_t + \psi_1 d_t + \varepsilon_{1,t}$$

$$r_{2,t} = b_{20} - b_{21}r_{1,t} + \gamma_{21}r_{1,t-1} + \gamma_{22}r_{2,t-1} + c_2 y_t + \psi_2 d_t + \varepsilon_{2,t},$$

which, in a more compact matrix form, can be written as

---

[9] See also Appendix A for a summary of the construction of EDFs. Combined with sectoral exposure information (for example from countries with credit registries), these EDF-based measures can provide a first approximation of the nature and magnitude of lending exposure (to default). Depending on the availability of recovery rates by sector, especially if these are modelled dynamically, the notion of VaR would be exact.

[10] Johansen (2000, p 361), for example, suggests that

> [if] we want to find relations between simultaneous values of the variables in order to understand interactions of the economy one would get a lot more information by relating the value of the variable to the value of other variables at the same time point rather than relating it to its own past. One can say that if we want to discuss relations between variables, then one should take combinations of simultaneous values and if we want to discuss dynamic development of the variables we should investigate the dependence on the past.

[11] This approach has much in common with autoregressive, distributed-lag (ARDL) models, a survey of which can be found in Pesaran and Smith (1998).

$$B\mathbf{r}_t = \Gamma_0 + \Gamma_1\mathbf{r}_{t-1} + Cy_t + \Psi d_t + \varepsilon_t, \tag{1}$$

where

$$B = \begin{bmatrix} 1 & b_{12} \\ b_{21} & 1 \end{bmatrix}; \quad \Gamma_0 = \begin{bmatrix} b_{10} \\ b_{20} \end{bmatrix}; \quad \Gamma_1 = \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix}; \quad C = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}; \quad \Psi = \begin{bmatrix} \psi_1 \\ \psi_2 \end{bmatrix}; \quad \varepsilon_t = \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}.$$

Equation (1) is the primitive form of the vector autoregression (VAR) process. The standard form can be obtained by premultiplication by the inverse of matrix B, resulting in:

$$\mathbf{r}_t = \pi + \Pi_1\mathbf{r}_{t-1} + Zy_t + \varphi d_t + \mathbf{e}_t, \tag{2}$$

where

$$\pi = B^{-1}\Gamma_0; \quad \Pi_1 = B^{-1}\Gamma_1; \quad Z = B^{-1}C; \quad \varphi = B^{-1}\Psi; \quad \mathbf{e}_t = B^{-1}\varepsilon_t.$$

In general, the interaction across risk processes and with the exogenous macroeconomic factors may take place over p (monthly) periods. In the presence of a vector of exogenous processes $\mathbf{y}_t$ (with an impact over k periods) and a dummy vector of exogenous shocks $\mathbf{d}_t$ (effect over l periods), equation (2) can be written as:

$$\begin{aligned}\mathbf{r}_t = {} & \pi + \Pi_1\mathbf{r}_{t-1} + \Pi_2\mathbf{r}_{t-2} + \cdots + \Pi_p\mathbf{r}_{t-p} \\ & + Z_0\mathbf{y}_t + Z_1\mathbf{y}_{t-1} + \cdots + Z_k\mathbf{y}_{t-k} \\ & + \varphi_0\mathbf{d}_t + \cdots + \varphi_l\mathbf{d}_{t-l} + \mathbf{e}_t.\end{aligned} \tag{3}$$

If the equilibrium is to be meaningful, the risk series need to be stationary, ie should not be characterised by a unit root. If any of the series $\mathbf{r}_t$ is integrated, denoted by $I(1)$, the regression results are not valid. In particular, whereas the estimated coefficients are still unbiased, their t-values are overrepresented.[12]

Integrated series could be brought back to stationarity by (the linear transformation of) differencing, $\mathbf{r}_t - \mathbf{r}_{t-1} = \Delta\mathbf{r}_t$. However, there may exist a non-zero linear combination of the integrated risk series, $\beta\mathbf{r}_t$, that is stationary. If this is the case, differencing the integrated series would ignore valuable information about long-term relationships that may exist between the series, such as both real and financial microeconomic linkages tying the different sectors. If such linear combinations exist, then the system is said to be cointegrated (CVAR) and the dynamic restrictions they impose on the system are testable. The linear relations between the sectoral risk measures are often called long-run equilibria.[13]

Cointegration is more easily visualised through an error correction representation of equation (3) above. For example, with restrictions of only one exogenous process and shock respectively, and $p = 2$, $k = 0$ and $l = 0$, equation (3) simplifies to:

$$\Delta\mathbf{r}_t = \Phi_1\Delta\mathbf{r}_{t-1} - \Pi\mathbf{r}_{t-1} + \pi + Zy_t + \varphi d_t + \mathbf{e}_t, \tag{4}$$

where $\Pi = I_p - \Pi_1 - \Pi_2$ and $\Phi_1 = -\Pi_2$. The term Π embodies the long-term effects in levels (adjustment to previous disequilibria in the risk profile across sectors), whereas Φ_1 represents the short-term or transitory shocks (adjustment to previous changes in risk).[14]

The hypothesis of cointegration is formulated as a reduced rank of the Π matrix:

$$H_1(r): \Pi = \alpha\beta', \tag{5}$$

where α and β are $p \times r$ matrices of full rank. The cointegration hypothesis implies that the process $\mathbf{r}_t$ is non-stationary, but that $\Delta\mathbf{r}_t$ and $\beta'\mathbf{r}_t$ are stationary.

---

12 For details on the nature of the problem, see for instance Hendry and Juselius (2000a).

13 See Hendry and Juselius (2000b) and Doornik et al (1998) for a comprehensive and clear exposition of cointegration analysis of VARs.

14 A number of alternative and equivalent error correction representations are possible, each emphasising a different aspect of the dynamic relationship. While they have all equivalent explanatory power and can be estimated by ordinary least squares, inference on some parameters will not be standard when the risk processes are integrated. See Hendry and Juselius (2000b, Section 4) for details on the different representations.

In the context of risk management, the cointegrating vectors in β can be thought of as *optimal portfolios*, as the risk profile of $\beta'\mathbf{r}_t$ is, for each vector in β, constant in expectation. That is, holding assets of the different sectors in proportions given by the inverse of the coefficients in the columns of β creates portfolios with stationary risk. The presence of several cointegrating relationships does not necessarily mean that there is more than one long-run equilibrium position. More likely, it may hint at the existence of a long-run equilibrium with embedded sectoral equilibria or cointegrated subsets of variables.

2.2 Data elements

As described above, our model consists of two main components, one endogenous and the other exogenous, and some additional elements that allow modelling one-time events. The first element is a *set of sectoral risk indicators* constituting a closed system, ie possibly interdependent, which we construct on the basis of firm-level EDFs. The second element is a *set of exogenous variables* that are orthogonal to the space of sector-specific shocks. These macro variables can also be thought of as a toolbox to be used in forecasting, scenario building and stress testing the system. We first specify the properties of the closed system before incorporating the exogenous elements and running the scenarios and stress tests.

2.2.1 Sectoral measures of risk

The chosen sectoral aggregation relating firm-specific EDF information to industry-specific risk measures $\mathbf{r}_t$ needs to weigh the positive information content of a possibly large set of indicators and their cost in terms of modelling requirements (allowing distinct characteristics across sectors). We define seven broad industries. EDFs for firms are available from KMV on a common methodology from January 1992 until May 2003 (137 monthly observations). Using as a basis the EU classification of economic activities (NACE Rev. 1), the over 1,500 SIC codes were mapped to a simpler classification of *seven broad industries* characterising the largest distinct economic sectors of interest (see Table 1).

Table 1

Economic industries

BaC	Basic goods and construction
EnU	Energy and utilities
Cap	Capital goods
CCy	Consumer cyclicals
CNC	Consumer non-cyclicals
Fin	Financial
TMT	Technology, media and telecommunications

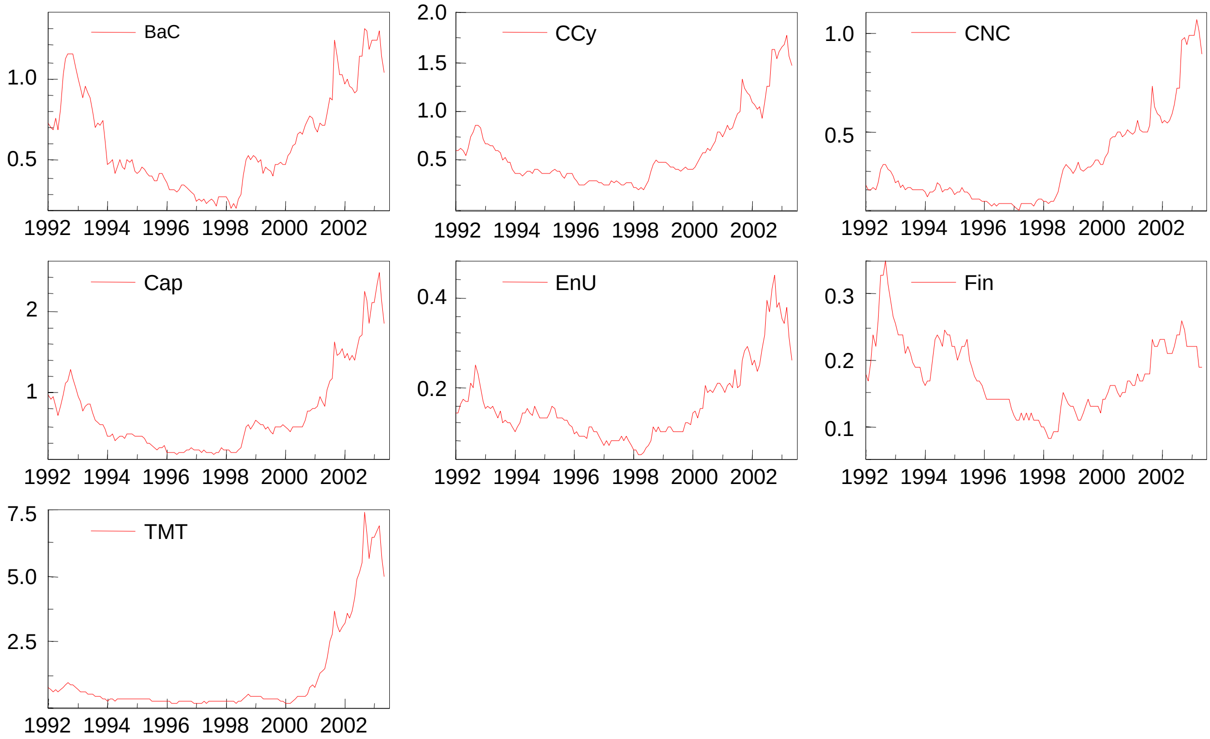
Once the industries (also referred to as sectors in what follows) have been defined, there are a number of ways of aggregating the firm-level EDF information into measures of sectoral default probability.[15] Of these, the simplest to implement is the sector's sample median (Graph 1).[16]

[15] See Appendix A for a discussion of the construction of the sectoral fragility indices.

[16] Although the sample median is not a sufficient statistic for the population mean, it converges (eg in mean) to the population mean when the population's distribution is symmetric, as described by Rose and Smith (2002). The median is robust to outliers in the sample, but has two weaknesses: it does not account for the potential risky tail of the distribution (does not weigh in information on the very risky firms) and ignores the size of the exposure to individual names. These issues remain to be addressed in future work.

Graph 1

Industry median EDF measures



Sectoral (median) EDF series are highly correlated across industries, denoting the close interaction of their measures of risk, and their possible sensitivity to common systemic or macroeconomic effects. The financial (Fin) and consumer non-cyclical (CNC) industries show the lowest correlation coefficients (Table 2), whereas those with the stronger contemporaneous correlations are the basic goods and construction (BaC), capital goods (Cap), consumer cyclical (CCy) and energy and utilities (EnU) industries.

Table 2

Industry EDF correlations

BaC						
0.9068	**CCy**					
0.7039	0.8875	**CNC**				
0.9223	0.9739	0.8196	**Cap**			
0.8596	0.9448	0.8483	0.9040	**EnU**		
0.7576	0.5726	0.2850	0.6085	0.6685	**Fin**	
0.6721	0.8800	0.7853	0.8738	0.8246	0.3789	**TMT**

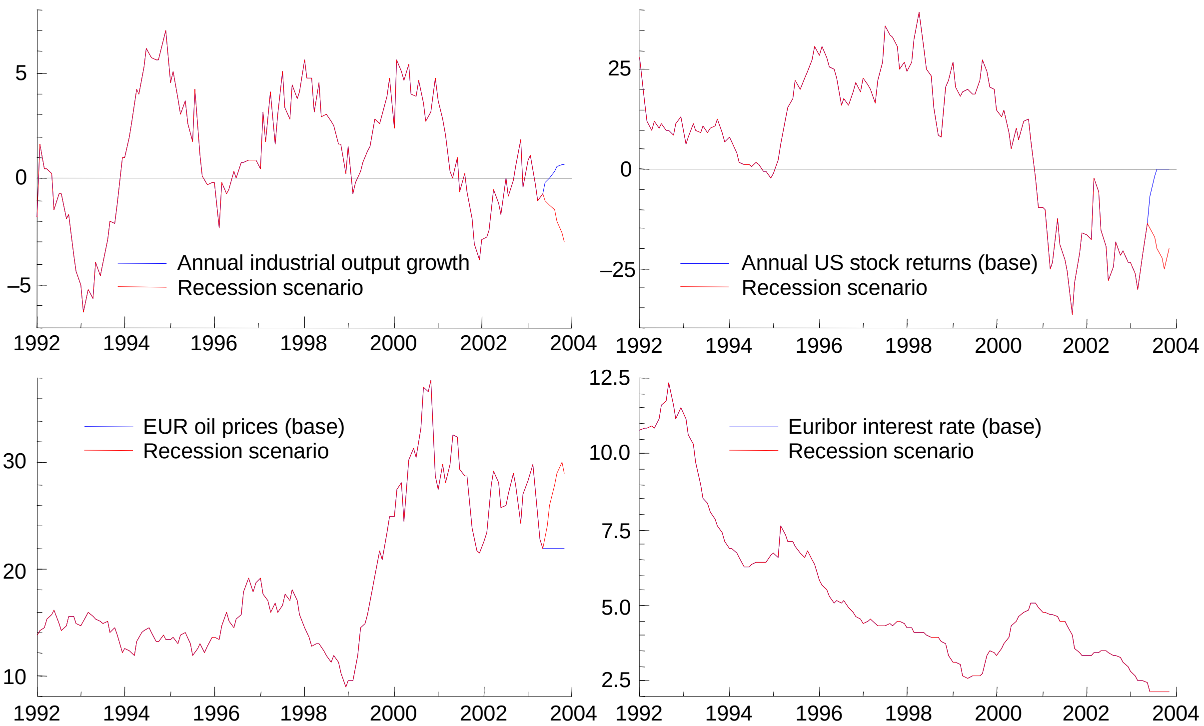
2.2.2 *Exogenous variables*

Three broad macroeconomic measures and a common stock market element are considered. The impact of *aggregate demand* changes is proxied by industrial output innovations.[17] Shocks emanating from *aggregate supply* are captured by innovations in oil prices, denominated in euros. Finally, *monetary shocks* are described by changes in a benchmark interest rate, which we have chosen to be the three-month Euribor rate. By construction, firm-level EDF information is negatively correlated to the firm's stock valuation. Movements in the stock markets that are widespread are therefore likely to have an impact on the sectoral measure of risk, suggesting the need to identify distinctly this type of *common equity shock*. In order to avoid endogeneity problems, we rely on recent results in the contagion literature suggesting that large shocks to equity markets are transmitted internationally, therefore hinting at the use of innovations in a foreign benchmark to proxy for economy-wide stock effects. We selected innovations in the DATASTREAM benchmark US stock index for this purpose.

The four exogenous variables are displayed, together with some assumed scenarios (see Appendix C below), in Graph 2.

Graph 2

Exogenous macroeconomic variables and scenarios



2.2.3 *Additional elements*

In addition to the endogenous and systemic exogenous indicators, we specify purely exogenous idiosyncratic shocks, such as the 11 September or ERM crisis shocks. Such events clearly affect the risk profile of firms and industries in quite a fundamental way and need to be accounted for in the exercise. We model these shocks by resorting to dummies, making careful note of the significant events underlying periods of "unusual" activity.[18]

[17] The choice of industrial production instead of GDP enables the analysis to be carried out on a monthly basis. We consider the 12-month growth rate, so as to avoid seasonal effects.

[18] These elements are not only desirable from a model-specification point of view, but they could also potentially serve to replicate similar shocks in the context of a historical stress test.

3. Econometric estimation

The risk series $\mathbf{r}_t$ do not show a trend but exhibit a high degree of persistence, with unit root Dickey-Fuller tests failing to reject the null hypothesis of an integrated process. The integrated nature of the risk processes requires differencing the risk series or incorporating cointegration analysis. On the basis of the high degree of covariance in the series (Graph 1), we look for cointegrating relationships across sectors, estimate the full cointegrated model, and analyse the cointegrating relationships.

It is evident from the preliminary model selection exercise in Appendix B that errors are significantly larger in some episodes of weak economic activity, active monetary policy, supply shocks and systemic stock market instability. For example, both periods of economic underperformance in the early 1990s and 2002 are marked by clearly larger levels of risk across sectors. Conditioning on external macroeconomic and financial market factors, as in equation (3), provides a set of "instruments" for carrying out stress-testing exercises on sectoral risk.[19]

3.1 Basic setup with exogenous factors $\mathbf{y}_t$

We consider the impact of the four exogenous variables described in Section 2.2.2, namely the 12-month change in the log of industrial output (proxying demand shocks), the Euribor three-month interest rate (proxying monetary policy changes), the price of oil in euros (identifying supply shocks), and the 12-month change in the log of the DATASTREAM benchmark US stock index (capturing large system-wide stock-related factors that are orthogonal to the industry-specific shocks). In equation (2), the lag specification of $\mathbf{y}_t$ is that of the endogenous variables, ie $p = k = 2$.

The diagnostic statistics on the individual equations improve relative to the closed model of Appendix B following the incorporation of systemic and macroeconomic variables (Table 3). The VAR(1) specification appears overall congruent with the data. Some problems remain with the EnU and, especially, the TMT sector.

Table 3

Model diagnostics for individual equations in the full model

Test	BaC	CCy	CNC	Cap	EnU	Fin	TMT
AR 1-12	0.82283	1.2097	1.6449	1.0137	0.70774	1.3571	4.5086
F(12,109)	[0.6267]	[0.2857]	[0.0897]	[0.4416]	[0.7410]	[0.1978]	[0.0000]**
ARCH 1-12	1.3075	0.94573	1.1426	1.6383	2.9640	0.75391	5.3021
F(12,97)	[0.2267]	[0.5056]	[0.3358]	[0.0935]	[0.0015]**	[0.6955]	[0.0000]**
Normality	3.9161	3.0566	8.0917	7.2450	14.709	2.3625	71.637
$\chi^2(2)$	[0.1411]	[0.2169]	[0.0175]*	[0.0267]*	[0.0006]**	[0.3069]	[0.0000]**
Hetero	1.9279	2.7686	3.4031	2.9249	3.3417	2.3137	5.1990
F(22,98)	[0.0154]*	[0.0003]**	[0.0000]**	[0.0002]**	[0.0000]**	[0.0027]**	[0.0000]**
Hetero X	1.5412	1.6428	1.3300	1.9305	4.2282	1.5192	2.7440
F(77,43)	[0.0621]	[0.0392]*	[0.1554]	[0.0103]*	[0.0000]**	[0.0686]	[0.0003]**

19 A framework with emphasis on the interdependence between sectoral risk dynamics and macroeconomic variables would consider these factors as additional endogenous variables in the VAR, possibly testing their (weak) exogeneity. We assume full exogeneity from the outset, because we are interested in the impact of systemic events on the sectoral risk profile of a chosen set of scenarios. An enhanced general equilibrium specification could be the focus of future work.

3.2 Characteristics of the system

Overall, the model is quite satisfactory in terms of its econometric properties. But how does it square with the intuition in terms of how macroeconomic events affect the risk profile of the different sectors? The significance of the retained regressors and the estimated coefficients is reported in Table 4.

First, the risk processes show strong and significant persistence (underpinned by the significant and large coefficients on their own lags), indicating the non-stationary nature of sectoral risk. In addition, all sectors show some degree of interconnection, even after conditioning on the macroeconomic processes. The significant and persistent impact of the CCy and TMT sectors on the remaining sectors also stands out. Whereas the former is significant at the 1% level in all equations in both lags, the first lag of the latter is significant in four other equations and mostly at the 5% level. Second in importance are the first lags of the CNC and EnU sectors, which show a significant impact on the Cap and TMT (at the 1% level) and Fin and TMT sectors (at the 5% level) respectively. Somewhat weaker is the effect of risk profile changes in the BaC and Cap sectors, which only appear significant at the 5% level in the CNC and TMT equations respectively. The risk profile of the Fin sector, while sensitive to the CCy and EnU sectors, does not appear to affect any of the other sectors.

The model is also congruent with some stylised facts regarding the systemic variables. First, the sign of the coefficients of the exogenous variables is, where significant, as expected: higher money market interest rates increase the risk of the given industry, whereas higher output and stock exchange growth rates decrease it. Positive shocks in the US stock market decrease the contemporaneous risk in any one industry (positive stock exchange impact), but appear to have some ripple effect (the lag of the opposite sign which is then reverted again in the second lag). The somewhat surprising result is that risk does not seem to be affected by oil prices (output shocks), except in the TMT sector, and then with an unexpected sign (higher oil prices lower the risk in the TMT sector). The second notable result is that deterministic shocks affect all the different industries, as denoted by the significance of the coefficients of the deterministic dummies in each equation of Table 4.

The results of Table 4 also suggest, however, that the estimation can be improved, in particular regarding the inference about the degree of interaction between sectors and the overall congruence of the model. This is illustrated in particular by the strongly significant one-lag coefficients in all equations. In order to address this potential shortcoming and so as to extend the cointegration analysis of the previous section, we explore the cointegrated relationships of the expanded open model.

3.3 Cointegrating relationships

Testing the cointegration rank r of the seven industry risk systems suggests the presence of a number of cointegration relationships ($r \neq 0$). The estimated eigenvalues, the log likelihood and the trace and maximum likelihood tests are tabulated in Table 5 below.

The trace test suggests that all λ_i are different from zero, indicating that all risk series are stationary. The variance in the value of the λ points out that the adjustment to the cointegration relationships varies substantially across cointegrating vectors. Since the λ_i can be interpreted as a squared canonical correlation coefficient, it provides a measure of the correlation with the stationary part. Accordingly, the drop in the magnitude of λ_i relative to λ_{i-1} would suggest a stronger correlation in the first $i-1$ relationships. However, the estimated λ_i for i sufficiently large are still showing substantial correlation with the stationary components. Notwithstanding this, some notable drops can be observed between the third and the fourth and between the fifth and the sixth λ. The maximum likelihood tests also pick up these drops, but still fail to reject the hypothesis of zero λ_i for i sufficiently large. We remain, therefore, suspicious about the significance of the fourth and higher cointegrating relationships.

Table 4

Full preliminary VAR(1) model

		Equation													
		BaC		**CCy**		**CNC**		**Cap**		**EnU**		**Fin**		**TMT**	
Regressor	***t***	**coeff**	**p-val**	**coeff**	**p-val**	**coeff**	**p-val**	**coeff**	**p-val**	**coeff**	**p-val**	**coeff**	**p-val**	**coeff**	**p-val**
BaC	–1	0.817	0.000	0.070	0.503	0.002	0.976	0.081	0.685	–0.039	0.474	–0.042	0.279	–0.159	0.620
	–2	0.049	0.740	0.042	0.672	0.153	0.038	0.176	0.347	0.098	0.059	0.030	0.407	0.369	0.221
CCy	–1	0.621	0.002	1.024	0.000	0.232	0.020	1.143	0.000	0.256	0.000	0.072	0.150	1.924	0.000
	–2	–0.700	0.001	–0.406	0.003	–0.460	0.000	–0.686	0.008	–0.284	0.000	–0.147	0.004	–1.426	0.001
CNC	–1	–0.312	0.155	–0.081	0.579	0.682	0.000	–0.742	0.008	–0.100	0.188	–0.053	0.330	–1.779	0.000
	–2	0.438	0.071	0.263	0.104	0.212	0.077	0.690	0.025	0.097	0.247	0.047	0.431	0.581	0.238
Cap	–1	0.064	0.543	0.054	0.444	0.016	0.765	0.574	0.000	–0.047	0.202	0.039	0.132	–0.505	0.020
	–2	–0.019	0.849	–0.052	0.446	0.032	0.532	0.001	0.996	0.040	0.263	0.015	0.562	0.381	0.069
EnU	–1	0.126	0.652	0.265	0.158	0.267	0.056	–0.076	0.829	0.604	0.000	0.138	0.049	1.132	0.049
	–2	–0.065	0.839	–0.083	0.699	0.064	0.689	–0.580	0.154	0.069	0.536	0.060	0.450	0.753	0.251
Fin	–1	0.246	0.537	–0.004	0.987	0.024	0.904	0.348	0.489	0.112	0.419	0.812	0.000	–0.487	0.549
	–2	–0.367	0.332	–0.259	0.306	–0.322	0.087	–0.376	0.432	–0.095	0.468	–0.121	0.197	–0.518	0.502
TMT	–1	0.007	0.815	0.049	0.018	0.030	0.046	0.083	0.032	0.028	0.010	–0.007	0.351	1.188	0.000
	–2	–0.004	0.891	–0.017	0.380	–0.014	0.355	–0.042	0.260	–0.017	0.094	0.003	0.693	–0.227	0.000

Table 4 (cont)

Full preliminary VAR(1) model

		Equation													
		BaC		CCy		CNC		Cap		EnU		Fin		TMT	
Regressor	***t***	**coeff**	**p-val**	**coeff**	**p-val**	**coeff**	**p-val**	**coeff**	**p-val**	**coeff**	**p-val**	**coeff**	**p-val**	**coeff**	**p-val**
i		0.034	0.108	0.013	0.359	0.014	0.191	0.018	0.506	0.011	0.130	0.012	0.023	0.076	0.077
	–1	–0.021	0.450	–0.003	0.861	–0.022	0.119	–0.002	0.964	–0.007	0.474	–0.010	0.168	–0.106	0.067
	–2	–0.004	0.843	–0.004	0.756	0.005	0.630	–0.018	0.490	–0.004	0.566	0.001	0.798	0.002	0.958
y		–0.008	0.046	–0.008	0.003	–0.003	0.174	–0.002	0.693	–0.002	0.133	–0.001	0.306	–0.009	0.266
	–1	0.000	0.968	0.001	0.665	0.001	0.778	0.008	0.119	0.002	0.127	0.001	0.545	0.014	0.090
	–2	0.003	0.377	0.004	0.082	0.004	0.027	0.000	0.918	0.001	0.685	0.000	0.916	–0.001	0.897
us		–0.002	0.045	–0.003	0.000	–0.001	0.021	–0.003	0.004	–0.001	0.030	–0.001	0.000	–0.009	0.000
	–1	0.003	0.011	0.003	0.001	0.002	0.019	0.006	0.000	0.001	0.042	0.001	0.002	0.010	0.000
	–2	–0.002	0.043	–0.002	0.013	–0.001	0.274	–0.002	0.088	–0.001	0.109	–0.001	0.001	–0.003	0.127
oil		0.002	0.534	0.000	0.813	0.000	0.847	0.004	0.219	0.002	0.068	0.000	0.650	0.010	0.062
	–1	0.000	0.963	0.001	0.788	0.000	0.915	–0.001	0.761	–0.001	0.686	0.000	0.690	0.001	0.840
	–2	–0.002	0.519	–0.001	0.745	0.000	0.756	–0.005	0.178	–0.001	0.417	–0.001	0.372	–0.012	0.031
c		0.026	0.534	0.058	0.041	0.046	0.028	–0.001	0.986	0.018	0.220	0.039	0.000	0.161	0.064
FinT		0.190	0.000	0.242	0.000	0.145	0.000	0.426	0.000	0.052	0.000	0.032	0.000	1.557	0.000
S11		0.301	0.000	0.247	0.000	0.187	0.000	0.332	0.000	0.008	0.535	0.026	0.006	0.784	0.000
ERM		0.084	0.005	0.071	0.000	0.042	0.005	0.076	0.041	0.032	0.002	0.042	0.000	0.099	0.097

Table 5

Cointegration tests of the open model

rank (i)	λ_i	loglik	H_o: $r \leq i$			
			Trace test		Max test	
0		2084.185	324.93	[0.000]**	101.05	[0.000]**
1	0.56619	2140.558	223.88	[0.000]**	61.94	[0.000]**
2	0.40066	2175.113	161.93	[0.000]**	54.38	[0.000]**
3	0.36199	2205.448	107.55	[0.000]**	36.31	[0.002]**
4	0.25927	2225.706	71.24	[0.000]**	35.02	[0.000]**
5	0.25132	2245.243	36.22	[0.000]**	22.33	[0.002]**
6	0.16853	2257.701	13.89	[0.000]**	13.89	[0.000]**
7	0.10842	2265.447				

In addition to the trace and maximum likelihood tests of Table 5, other criteria for specifying the cointegration rank include (Hendry and Juselius (2000b, p 22)):

1. the t-values of the α coefficients;
2. the graph of the cointegrating relation;
3. the recursive graph of the trace statistic for $\tilde{r} = 1, 2, \ldots, p$;
4. the characteristic roots of the model;
5. economic interpretability of the results.

The rank estimation of the unrestricted cointegrated model yields the cointegration matrices reported in Table 6. The t-values for the α coefficients would support ignoring the last two cointegrating vectors (no coefficients above the benchmark value of 3).

The graphs of the cointegrating vectors $\beta_i \mathbf{r}_t$ (Graph 3) indicate that the last three vectors appear to be non-stationary. This is also the case, but to a lesser degree, with the fourth cointegrating vector towards the end of the period, and, sporadically, also for the third vector. As the cointegrating relations are supposed to be stationary, this suggests that the cointegration rank is unlikely to exceed 4.

Since the variable $T_j \ln(1-\lambda_i)$ for $j = T_1, \ldots, T$ grows linearly over time when $\lambda_i \neq 0$, the recursively calculated components of the trace statistic should increase linearly for the first r components, but stay constant for the remainder. The recursive components of the unrestricted model's trace statistics shown in Graph 4 would support the correct specification to be $r = 1$, as the trace statistics for $r \geq 2$ all seem to be constant.

The information derived from the characteristic roots (Table 7) does not provide any conclusive information about the correct rank specification. In principle, if the $r\text{th} + 1$ cointegration vector is non-stationary and is wrongly included in the model, then the largest characteristic root will be close to the unit circle. The largest characteristic root when $r = 1$ has the smallest value among the largest common roots for different restrictions on the cointegration rank, supporting the possibility of a single cointegration relationship. Other troughs are found at $\tilde{r} = 4$ and 6.

Table 6

Cointegration matrices of the open model

	β							α						
	1	**2**	**3**	**4**	**5**	**6**	**7**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
BaC	1.00	0.54	1.12	1.99	–0.31	0.05	–48.50	0.019	–0.044	0.016	–0.028	0.103	0.038	0.001
CCy	–3.83	1.00	–2.50	0.27	–0.15	–0.15	86.41	0.021	0.058	0.144	–0.024	0.129	0.008	0.000
CNC	–5.39	–0.45	1.00	–6.22	0.23	–0.26	3.51	0.017	0.015	0.089	0.008	–0.110	0.087	0.000
Cap	2.82	–0.76	0.08	1.00	0.04	0.10	–19.98	0.011	0.515	0.061	–0.020	–0.031	–0.216	0.001
EnU	18.47	–1.26	0.79	15.14	1.00	–0.47	–16.18	0.004	0.014	0.012	–0.012	–0.166	0.095	0.000
Fin	–15.10	0.59	–0.91	–3.57	0.33	1.00	31.02	0.014	–0.044	0.002	–0.006	–0.055	–0.081	0.000
TMT	–0.26	0.05	0.21	–0.26	–0.02	0.02	1.00	0.131	0.711	–0.171	0.007	0.503	0.271	0.000
i	0.65	–0.01	–0.05	–0.31	–0.02	0.02	11.95							
–1	–0.83	0.01	0.05	–0.23	0.02	–0.03	–10.72							
–2	0.15	–0.03	0.01	0.31	0.00	0.00	0.16							
y	–0.08	0.00	–0.02	0.13	–0.01	–0.01	–1.22							
–1	0.05	0.01	–0.01	–0.05	–0.01	0.00	–0.67							
–2	0.03	0.00	0.03	0.02	0.00	0.01	1.88							
us	–0.06	0.00	0.00	0.03	0.00	0.00	–0.06							
–1	0.06	0.01	0.00	–0.04	0.00	0.00	1.11							
–2	–0.03	0.00	0.00	0.04	0.00	0.00	–0.46							
oil	0.03	0.01	–0.01	–0.05	0.00	0.00	0.28							
–1	0.00	0.00	0.00	0.00	0.00	0.00	–0.26							
–2	–0.05	–0.01	0.01	0.02	0.00	0.00	–0.55							

The greyness of the alpha value is determined by its t-value: if in excess of five, it is dark grey; if in excess of four, it is grey; and if in excess of three, it is light grey.

Graph 3

Unrestricted cointegration relations $\beta_i r_t$

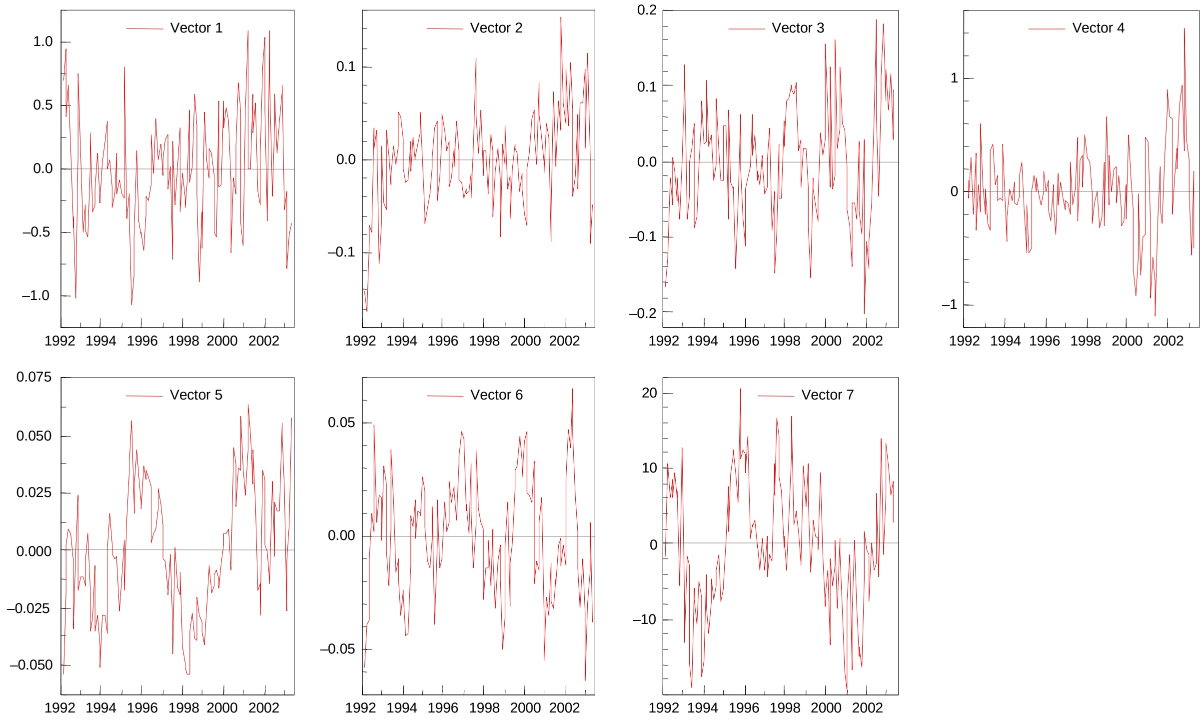


Table 7

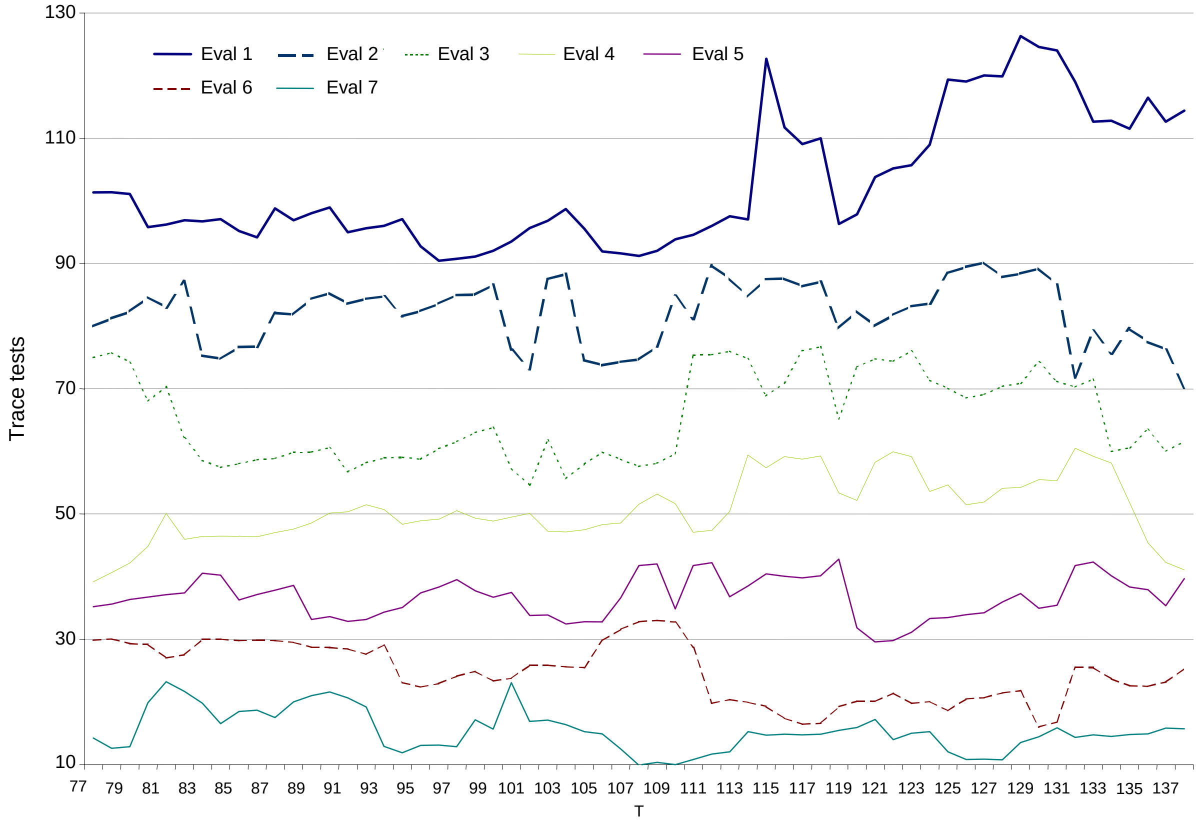
Largest characteristic roots of model

H_o: r =						
1	**2**	**3**	**4**	**5**	**6**	**7**
1.00	1.00	1.00	1.00	1.00	1.00	0.96
1.00	1.00	1.00	1.00	1.00	0.86	0.96
1.00	1.00	1.00	1.00	0.88	0.86	0.79
1.00	1.00	1.00	0.86	0.77	0.77	0.75
1.00	1.00	0.91	0.86	0.77	0.77	0.75
1.00	0.92	0.75	0.61	0.65	0.64	0.65
0.76	0.43	0.44	0.61	0.65	0.64	0.65

In order to assess the correct rank, we check the economic interpretability of the higher-order ranks. In this task, the significance of the adjustment parameter α (Table 6) is useful, as it identifies the sector whose risk process is correlated with the stationary part. Only the EnU sector adjusts to the fifth cointegration vector. Indeed, looking at the coefficients of β_5 suggests that it is essentially a unit vector describing risk in the EnU sector. Looking at the evolution of this vector in Graph 3 and the evolution of r_{EnU} in Graph 1 supports this view, especially until 2001 (where the series have not been filtered for deterministic shocks). Similar reasoning would also apply to β_4 (tying r_{CCy} and r_{TMT}, possibly through some consumption goods channel), β_3 (tying r_{Cap} and r_{TMT} possibly through an investment channel), and β_2 (a unit vector for r_{TMT}, with a strong impact on the Cap sector).

Graph 4

Recursive trace statistic $T\ \ln(1-\lambda_i)$



With the prerogative of treating near unit roots as unit roots (thereby increasing the accuracy of statistical inference),[20] we adopt a bias to reduce the cointegration rank wherever the criteria do not strongly support the advantage of increasing the rank. The strict implementation of these criteria would indicate $r = 1$, with a somewhat more lenient alternative of $r = 4$. So as to facilitate discussion, we take the strict route and consider a unitary cointegration rank.

3.3.1 *Identification, hypothesis testing and weak exogeneity*

A simple rotation of the cointegration space (leaving the estimated long-run matrix Π in equation (5)) is sufficient to uniquely determine the cointegrating space (in this case, vector).[21] We set $\beta_1 = 1$, thus obtaining the cointegrating vector[22]

$$\hat{\beta}' = \left[\begin{array}{ccccccc} BaC & CCy & CNC & Cap & EnU & Fin & TMT \\ 1 & -3.83 & -5.39 & 2.82 & 18.50 & -15.10 & -0.26 \\ (0.00) & (1.76) & (1.31) & (0.76) & (3.29) & (2.30) & (0.11) \end{array}\right.$$

$$\left.\begin{array}{cccccccccccc} i_t & i_{t-1} & i_{t-2} & y_t & y_{t-1} & y_{t-2} & us_t & us_{t-1} & us_{t-2} & oil_t & oil_{t-1} & oil_{t-2} \\ 0.65 & -0.83 & 0.15 & -0.08 & 0.05 & 0.03 & -0.06 & 0.06 & 0.03 & 0.03 & 0.00 & 0.05 \\ (0.20) & (0.28) & (0.20) & (0.04) & (0.04) & (0.04) & (0.01) & (0.01) & (0.01) & (0.03) & (0.03) & (0.03) \end{array}\right].$$

[20] See Hendry and Juselius (2000b, pp 22-24) for an elaboration on the preferability of treating near unit roots as unit roots.

[21] In general, just identification can be achieved by imposing one appropriate normalisation (ensuring that this coefficient is non-zero) and *r – 1* restrictions on each β_i.

[22] Standard errors in parenthesis.

Testing the significance of the β coefficients on i_{t-2}, y_{t-1}, y_{t-2} and all lags of *oil*, we cannot reject the null that excluding them from β does not remove any valuable information ($\chi^2(6) = 7.89[0.25]$). The oil sector, therefore, does not appear to have a permanent impact on the risk profile of the industrial sectors. This also applies to the lagged values of the industrial output and the second lag of the interest rate. The fact that some lags enter the cointegrating relationship would suggest that the equilibrium has an important backward-looking element. Thus, the level of return in the US stock markets and the previous month's level of interest rates in Europe form part of *today's* equilibrium risk profile.

The significance of the α coefficients of the cointegrating relations points to further ways to restrict Π so as to more accurately pin down the structure of cointegration. The hypothesis that a variable influences the long-run development of other variables, *but is not influenced by them*, is called the weak exogeneity or *no-levels feedback*. In our model, the α coefficients for the Cap and EnU sectors do not appear significant, indicating that risk in these sectors is weakly exogenous. As their risk does not adjust to the long-term relations implied by the cointegrating vectors, it can be considered as a driving trend in the system (common stochastic trend). Curiously, the two sectors that do not adjust to the cointegrating relation, the Cap and EnU, are the main "infrastructure" sectors. This suggests that risk in these two sectors is in disequilibrium vis-à-vis the remaining sectors (is not integrated).

After imposing these restrictions, the coefficient adjustment and cointegrating vectors of the model are the following:

$$\hat{\alpha}' = \begin{bmatrix} BaC & CCy & CNC & Cap & EnU & Fin & TMT \\ -0.009 & -0.010 & -0.008 & & & -0.007 & -0.067 \\ (0.004) & (0.003) & (0.002) & & & (0.001) & (0.009) \end{bmatrix}, \tag{6}$$

$$\hat{\beta}' = \begin{bmatrix} BaC & CCy & CNC & Cap & EnU & Fin & TMT & i_t & i_{t-1} & y_t & us_t & us_{t-1} & us_{t-2} \\ 1 & 10.43 & 7.68 & -7.50 & -38.69 & 25.64 & 0.66 & -1.11 & 1.10 & 0.11 & 0.09 & -0.10 & 0.06 \\ (0.00) & (3.06) & (2.15) & (1.35) & (5.52) & (3.79) & (0.20) & (0.37) & (0.36) & (0.05) & (0.02) & (0.02) & (0.02) \end{bmatrix}. \tag{7}$$

The five sectors adjusting to deviations from the long-run equilibrium do so in a similar fashion: positive deviations reduce risk in the sectors. In the long run, the same sectors appear as "substitutes in risk", as a long-run increase in any one of them would require a long-run decrease in the other. The five as a whole are "complements in risk" to the two sectors that do not adjust to the long-term equilibrium, ie the infrastructure sectors Cap and EnU. The long-run coefficients of the exogenous factors are intuitive, as a concurrent increase in the interest rate, decrease in output or decreased returns in US stock markets would increase the level of the endogenous industry risk levels. The one-lag effect of interest rates and returns in the US stock markets has the opposite impact, suggesting that they serve as counteracting elements to the concurrent effects, thus "slowing down" the long-term adjustment. This cyclical characteristic is again observed in the long-run coefficient of the second lag of US stock returns, which again reverts to lowering risk levels across the adjusting industries.

3.4 Reduction to the non-integrated model

The final step in constructing the model involves mapping the data to the non-integrated differences of the original series having accounted for the cointegration factor. In its unrestricted form, the $I(0)$ model obtained accounting for the single cointegration relationship is given in Table 8.

It is clear from Table 8 that some differences remain between the integrated and non-integrated models. Innovations in the EnU and Fin sectors do not have a significant impact on innovations in any other sector, while innovations in the BaC and Cap sectors affect only one other sector at the 5% level of significance. The results underline the key role played by the CCy sector, whose fragility positively and significantly affects that of all other sectors. In particular, the Cap and TMT sectors show a strong response, magnified in the case of the latter to almost double.

Table 8

Model in its non-integrated form

		Equation													
		ΔBaC		ΔCCy		ΔCNC		ΔCap		ΔEnU		ΔFin		ΔTMT	
Regressor	***t***	**coeff**	**p-val**	**coeff**	**p-val**	**coeff**	**p-val**	**coeff**	**p-val**	**coeff**	**p-val**	**coeff**	**p-val**	**coeff**	**p-val**
ΔBaC	–1	–0.055	0.682	–0.065	0.511	–0.144	0.044	–0.167	0.343	–0.057	0.238	–0.050	0.181	–0.334	0.295
ΔCCy	–1	0.573	0.001	0.325	0.010	0.337	0.000	0.937	0.000	0.213	0.001	0.108	0.023	1.909	0.000
ΔCNC	–1	–0.442	0.025	–0.061	0.672	–0.068	0.511	–0.735	0.005	–0.090	0.204	–0.033	0.548	–1.417	0.003
ΔCap	–1	0.021	0.807	0.046	0.473	0.017	0.709	–0.084	0.463	–0.042	0.182	–0.002	0.931	–0.422	0.045
ΔEnU	–1	0.080	0.768	0.224	0.265	0.080	0.579	0.574	0.110	–0.177	0.072	–0.043	0.572	–0.693	0.287
ΔFin	–1	0.308	0.376	–0.021	0.936	0.063	0.733	–0.013	0.978	0.092	0.461	0.078	0.424	–0.039	0.962
ΔTMT	–1	0.020	0.437	0.031	0.107	0.014	0.301	0.071	0.041	0.030	0.002	–0.003	0.667	0.312	0.000
CI	–1	0.002	0.688	0.000	0.900	–0.003	0.210	0.016	0.015	0.002	0.240	–0.003	0.014	–0.043	0.000
c		–0.007	0.703	0.004	0.733	0.014	0.127	–0.052	0.023	–0.008	0.225	0.012	0.017	0.170	0.000
FinT		0.184	0.000	0.257	0.000	0.159	0.000	0.479	0.000	0.056	0.000	0.031	0.000	1.689	0.000
S11		0.329	0.000	0.281	0.000	0.193	0.000	0.388	0.000	0.019	0.137	0.035	0.001	0.895	0.000
ERM		0.114	0.000	0.092	0.000	0.038	0.005	0.098	0.004	0.039	0.000	0.050	0.000	0.082	0.180

Also a factor important to the fragility of a number of sectors, changes in the risk profile of the CNC sector show a *negative* impact on the risk profile of the sectors that are statistically significant to them, notably BaC, Cap and TMT. The negative and generally large coefficient in these relations may capture some sort of impact of investment strategies seeking more cyclical targets once the non-cyclical sectors show signs of fragility. This stylised view of seeking "faster-rebounding" investments is reinforced by the large coefficient in the TMT sector, usually a strong performer in economic rebounds. The other sector whose fragility shows correlation with that of other industries, the TMT industry, has positive coefficients, denoting the standard positive sign of "contagion", albeit the scale is smaller, except for the autoregressive term, which shows how strong the persistence is in the risk process of this sector.

The adjustment coefficients do not point to any peculiarity other than long-term adjustment not significantly affecting more than three sectors, of which the negative impact on the TMT sector stands out. The deterministic shocks show a significant impact on risk across sectors. Only the EnU and Fin industries show a small response to the deterministic components.

Overall, the model performs reasonably well, as portrayed by the goodness of fit measures: the R^2 based on the likelihood ratio is 0.83, whereas that based on the Lagrange multiplier reports a lower 0.19. Model diagnostics also suggest that some problems are still present in the TMT equation, where the errors show some degree of kurtosis, and serial correlation, thereby suggesting that the results from this equation be viewed with caution. Notwithstanding this, the remaining equations perform well.

Two main points stand out in the analysis so far. The first underlines the fact that the very simple structure we imposed on the system already reveals great complexity in the cross-industry risk linkages. The interaction between the median EDF measures across sectors depicted by the VAR model has two interesting components: across industries the correlations may be negative, suggesting some "complementarity" property probably driven by large capital movements across sectors, and the time dimension suggests that some sectors, notably consumer cyclicals, serve as shock transmission channels, also potentially signalling early warning properties.

The second property supported by the model is the perception that systemic variables, in particular those related to the macroeconomy, do not appear to influence the behaviour of the series at the monthly frequency. This result is to be taken with caution, as there are a number of simplifications assumed in constructing the systemic variables. Nonetheless, the results are at least suggestive of other factors driving the behaviour of a significant proportion of risk in the sectors. Also supporting this perception is the important role of deterministic components (dummies), which capture the sensitivity of risk to events outside the economic framework.

4. Concluding remarks

This exercise is an example of the use of market-based information in the assessment of fairly aggregated sectoral fragility. While the preliminary results are encouraging in terms of both statistical fit and modelling usefulness, the model could benefit from having longer time series covering a full macroeconomic cycle, currently not fully available.

The results from the model underline three factors defining risk across economic sectors. The first one refers to the important observation that risk modelling ought to consider the important cross-dynamics transmitting risk across industries and time. It appears that some progress can be made in modelling the structure through which risk is propagated across sectors and time, and that imposing further restrictions on the reduced form of the VAR may well provide further insight about the structure of risk transmission. The second element is the notion that risk exhibits evidence of evolving to a long-run equilibrium. Systemic and macroeconomic factors affect the steady state levels and thus represent important determinants of the steady state risk profile of most of the sectors. Ignoring this interaction weakens the strength of the forecasts, which benefit from incorporating this significant adjustment factor. The final element is a word of caution. The model fails to detect a large significant impact stemming from macroeconomic and systemic elements, and a substantial share of the variation in risk across industries remains unaccounted for. This outcome may be uncovering some uncomfortable results, namely that much of the change in risk profiles is driven by elements that are independent of the economic performance, possibly owing to some herding factor.

A number of potential applications for the framework are possible. For example, and in combination with aggregate sectoral exposure data, one could assess the exposure at risk of the banking sector's lending portfolio, as well as its variation in selected scenarios for the systemic variables (scenario or stress testing). In addition, one may want to use the model to assess future levels of risk by sector, or in the aggregate, benefiting from information on the interaction of risk across industries and the forces driving these to the long-run equilibrium. Furthermore, the sensitivity of sectoral risk to other factors, including risk in other sectors, can be further refined, thus allowing for a more refined assessment of optimal investment strategies across sectors.

The exercise lends itself to a number of improvements and variations. A key characteristic of the model is the substantial amount of volatility experienced in periods of economic distress. We have accounted for this by carefully constructing deterministic variables characterising the main aspects of these periods. Some preliminary analysis with Markov-switching VARs suggests that this may indeed lead to a potential improvement in the model by pinning down the factors affecting the transition probabilities between states of high and low volatility. This strategy would require careful consideration of the type of switching mechanism that would best suit the properties of the median EDF. A second constraint on this exercise is the high number of parameters needing estimation. A possible way forward is to implement Bayesian estimation techniques reducing the number of parameters by imposing prior restrictions based on the experience gained so far. It is difficult to assess how much mileage to extract from this, however, given that the persistence of the estimated VAR appears to be quite low (one lag sufficed to remove the serial correlation of the errors, except perhaps in the TMT sector) and the substantial differences in the behaviour of the probability of default across sectors. Finally, and in the light of the weak explanatory power of the systemic indicators (exogenous variables), further investigation would be appropriate in determining the nature of the common factors driving the movement in the median EDF measures. Two aspects affecting risk developments are (i) performance announcements and (ii) estimated future outlooks. These elements suggest that one looks at either contemporaneous *expected* values, or future values of the exogenous variables (assuming that forecasts are correct). It is difficult to argue, however, that relatively high-frequency information will be driven by lower-frequency announcements, so it may be desirable to look at the behaviour of trends (smoothed data).

Each of these issues may deserve some attention in the future.

Appendix A:
Financial fragility

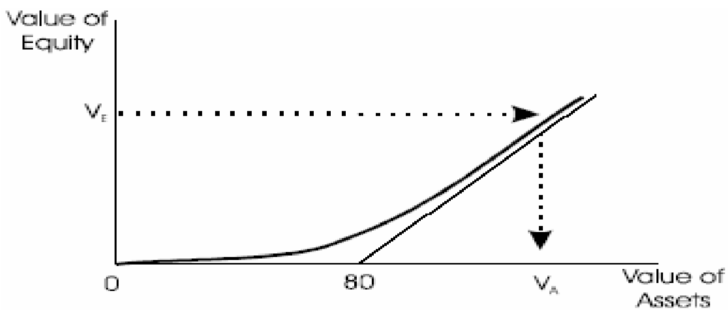
Among the measures that have been proposed to gauge corporate fragility, option-based indicators, such as KMV's expected default frequency (EDF), have shown to have desirable leading-indicator properties. Theoretically founded on the well known Black-Scholes option-pricing equation, they combine three elementary components (assets' value, their risk and the firm's leverage) into a unique measure of default risk.[23] Their ability both to predict overall levels of defaults and to discriminate between defaults and non-defaults is known and valued among practitioners.[24] EDFs can be directly used in calculating exposures at risk whereas alternative forward-looking indicators (subordinated debt spreads or equity price implied volatility) have to first be converted into a meaningful measure of probability of default. Together with proper recovery rates, they can also serve to assess loss-given-default. In this section we first derive the firm's EDF from high-frequency market and financial information on the firm. For further details, see Crosbie and Bohn (2002).

A.1 Corporate fragility

The practical approach implemented by KMV rests on three basic steps: the estimation of asset value and volatility, the calculation of the distance-to-default measure, and the derivation of the EDF. The first step in this derivation is based on the observation that *equity is essentially the same as a call option on the firm's assets with a strike price equal to the book value of the firm's liabilities* (at liquidation). The option nature of equity serves to derive the underlying market value and volatility of the firm's assets, the volatility of equity, and the book value of liabilities.[25] This process is similar in spirit to the procedure used by option traders in the determination of the implied volatility of an option from the observed option price and exploits the close relationship between the value of debt and that of equity as they are both really derivative securities on the underlying assets of the firm. The option nature of equity can be thus exploited to relate the market value of equity and the book value of debt to determine the implied market value of the underlying assets. Graph 5 illustrates the derivation of the market value of assets V_A from the value of equity V_E and an option-pricing relationship (thick line) for a simple leveraged mutual fund.

Graph 5

Derivation of the value of assets V_A from the value of equity V_E



[23] See Crosbie and Bohn (2002) for further details on the construction of the distances to default on which EDFs are based.

[24] KMV has produced a number of technical documents on the subject. See, for instance, Kurbat and Korablev (2002) for further references on the subject.

[25] The model was developed at KMV by Oldrich Vasicek and Stephen Kealhofer as an extension of the Black-Scholes-Merton framework and is known as the Vasicek-Kealhofer (VK) model. This model assumes that the firm's equity is a perpetual option with the default point acting as the absorbing barrier for the firm's asset value. When the asset value hits the default point, the firm is assumed to default. Multiple classes of liabilities are modelled: short-term liabilities, long-term liabilities, convertible debt, preferred equity, and common equity. When the firm's asset value becomes very large, the convertible securities are assumed to convert and dilute the existing equity. In addition, cash payouts such as dividends are explicitly used in the VK model. See Crosbie and Bohn (2002) for further details.

In fact, accounting for more complicated examples, the value of assets V_A and its volatility σ_A are derived from the following simultaneous relationships:

$$\mathbf{V}_E = \text{Option Function}\left(V_A,\quad \sigma_A,\quad \begin{bmatrix}\text{Captial}\\ \text{Structure}\end{bmatrix},\quad \begin{bmatrix}\text{Interest}\\ \text{rate}\end{bmatrix}\right)$$

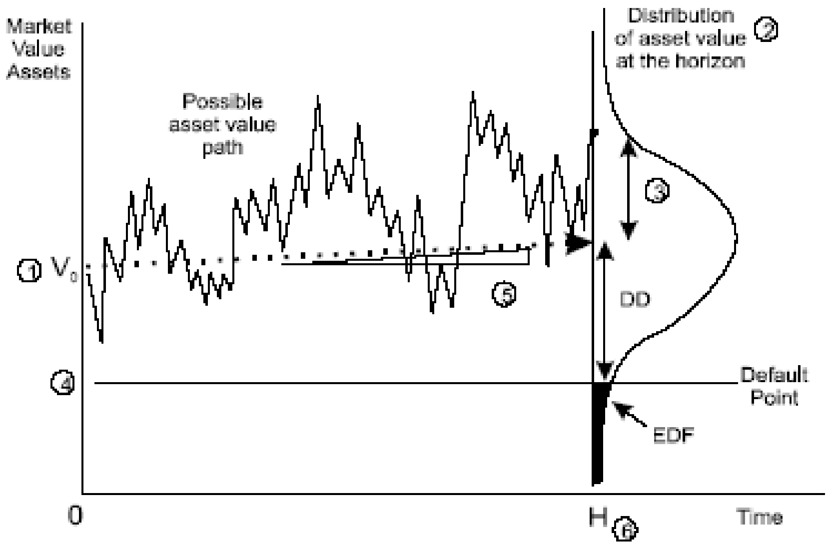
$$\sigma_E = \text{Option Function}\left(V_A,\quad \sigma_A,\quad \begin{bmatrix}\text{Captial}\\ \text{Structure}\end{bmatrix},\quad \begin{bmatrix}\text{Interest}\\ \text{rate}\end{bmatrix}\right)$$

from where it is clear that only the market value of assets V_A and its volatility σ_A are unknown. Both are derived by solving the relationships from the other known variables.

The second step involves the calculation of the distances to default and requires six measures. Considering the horizon from now until H, the variables required are: (1) the current asset value, (2) the distribution of the asset value at time H, (3) the volatility of the future assets at time H, (4) the level of the default point (book value of liabilities), (5) the expected rate of growth in the asset value over the horizon, and (6) the length of the horizon H. These elements are illustrated in Graph 6.

Graph 6

Calculation of the distance to default



The first four (asset value, future asset distribution, asset volatility and the level of the default point) are the main variables, as the expected growth in the asset value has little default discriminating power and the analyst defines the length of the horizon. If the future distribution of the distance to default were known, the default probability (EDF value) would simply be the likelihood that the final asset value was below the default point (the shaded area in Graph 6). In practice, however, the distribution of the distance to default is difficult to access, as the usual normal or log-normal distributional assumptions cannot be used. The likelihood of large adverse changes in the relationship of asset value to the firm's default point is critical to the accurate determination of the default probability. These changes may come about from changes in asset value or changes in the firm's leverage. In fact, changes in asset value and changes in firm leverage may be highly correlated. Consequently, the distance to default is first measured as the number of standard deviations the asset value is away from default:

$$DD = \frac{V_A - \begin{bmatrix}\text{Default}\\ \text{Point}\end{bmatrix}}{V_A \sigma_A}.$$

Empirical data are then used in the third and final stage to determine the corresponding default probability. KMV obtains the relationship between distance to historical default and bankruptcy frequencies from a database including over 250,000 company years of data and over 4,700 incidents of default or bankruptcy. From these data, a look up or frequency table can be generated which relates the likelihood of default to various levels of distance to default.[26]

A.2 Sectoral fragility

Forward-looking indicators assist in predicting the trend of the expected bank losses in the near future. As we have seen, EDFs provide an approximation of the expected probability of default for individual firms. Each firm is associated with an industry and thus *industry risk* measures can be constructed in a number of ways. For example, we could resort to the industry's median EDF, a weighted average (by market asset value or liabilities, for instance) EDF, or other kernel measures aggregating firms' EDFs. The weighted average effectively incorporates information on the *large players* affecting the sector's riskiness, but is subject to spurious variation due to classification changes, especially of large players. Because the problem with weighted averages may be significant in our data sample, we instead obtain sectoral measures of risk by grouping firms into sectors and taking the median of those.[27] Likewise, an aggregate measure of risk can be derived by considering the whole population in the sample from where the median is drawn. Denoting a group of firms in our data set by J and a firm j's EDF by p_j, our measure of risk for group J, r_J, is therefore the median p_j of the group.

[26] The relationship between distance to default and default frequency for industry, size, time and other effects was tested by KMV and was found constant. This is not to say that there are no differences in default rates across industry, time and size but only that it appears that these differences are captured by the distance-to-default measure.

[27] Other measures have been suggested that are less subject to the spurious fluctuation due to classification changes while at the same time providing a greater significance to the large players. One measure that could be implemented is the median of the n largest (by liabilities) corporations. Some sensitivity analysis would be required for establishing the optimal n and this option could be considered in future work.

Appendix B:
Preliminary model specification

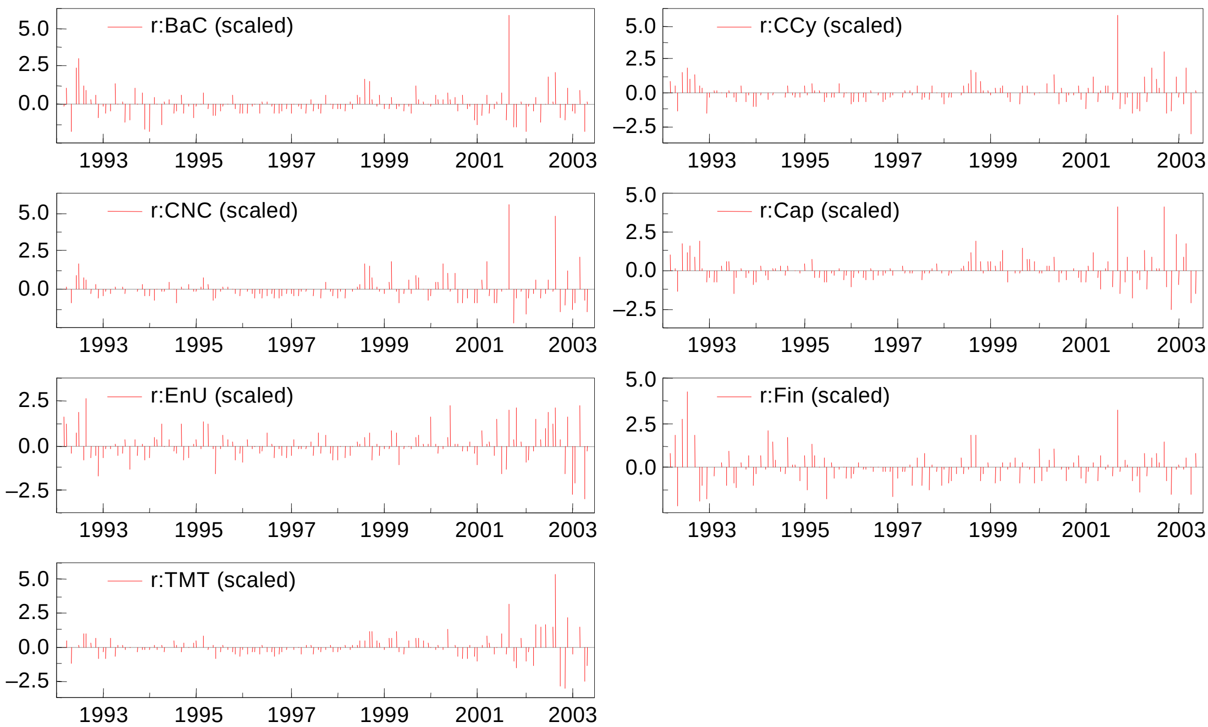
In the spirit of Hendry and Juselius (2000b), we begin by first tentatively estimating a VAR system under the presumption that the risk processes are not integrated, as in equation (3), excluding the dependency of systemic and macroeconomic effects.[28]

In this appendix, we verify the model's specification congruency with the data on the basis of the three core criteria for statistical inference identified by Hendry and Juselius (2000b): parameter constancy, serially uncorrelated residuals, and residual skewness.[29]

Two monthly lags suffice to account for the 12-month autocorrelation of four of the sectoral equations.[30] The remaining large errors (Graph 7) are concentrated in summer 1992 (ERM crisis), early autumn 2001 (events of 11 September), and autumn 2002 (financial market turbulence associated with uncertainty over the impact of the bursting of the equity market bubble). In addition, the volatility of the residuals appears higher towards the end of the sample, suggesting some form of non-linearity in the system.

Graph 7

Residuals in the closed VAR with no deterministic or exogenous variables



[28] We thereby first focus on selecting the most parsimonious specification that does not show residual serial correlation. Selecting a parsimonious lag order is also recommended for testing the cointegration rank, as elaborated by Ho and Sorensen (1996). One could alternatively resort to standard tests for the optimal lag. See Doornik and Hendry (2001), Hansen and Juselius (2002) or Enders (1996) for further material on this subject and references to original theoretical work.

[29] The software package used for estimation is GiveWin version 2.2.

[30] The tests for 12-lag serial correlation do not reject, at the 1% level of significance, the residuals for the consumer cyclical (CCy) and non-cyclical (CNC) goods and technology and telecommunications (TMT) sectors, which are still subject to very large shocks affecting the tests for serial correlation. Because of these shocks, incorporating additional lags does not correct for the pattern captured by the serial correlation tests.

We therefore correct for the presence of deterministic exogenous shocks by adding shock dummies for the periods of exogenous instability (summer 1992, autumn 2001, and the turbulent period in financial markets in autumn 2002).[31] The dummies used in the exercise are "neutral" over time or mean-zero, as for example the Fall01 dummy, which equals 1 in September and –1 in October 2001 (and zero otherwise). In addition, and in agreement with the discontinuities observed in Graph 1, the FinT dummy has scaled values in the period from June 2002 to May 2003 which add up to zero, similar to the ERM dummy in the period June-December 1992.[32]

Accounting for deterministic shocks noticeably corrects the congruence of the model, with the three shocks being significant at the 1% level. Serial correlation (12 lags) remains insignificant at the 1% level of significance in all industries except for the technology and telecommunications (TMT) sector, where large shocks in April and May 2002 continue to account for errors larger than 3.5 standard deviations, suggesting that some sector-specific factors have still not been accounted for.[33] The persistent presence of large shocks also affects multivariate tests on serial correlation. As pointed out by Hendry and Juselius (2000b, p 6), however, in economic applications the multivariate normality and serial correlation are seldom satisfied, and accurate inference must rely on the careful interpretation of remaining problems.

Most one-step residuals of a recursive estimation are within two standard deviations, indicating parameter constancy.[34] Owing to the more volatile environment surrounding the bursting of the equity bubble, errors also appear heteroskedastic, except for the energy and utilities (EnU) and TMT sectors, where the assumption of homoskedastic errors cannot be rejected at the 5% level of significance. The assumption of normality can be rejected at the 5% level, with the exception again of the EnU and TMT sectors. A closer inspection of Graph 8, displaying the errors' residual density, suggests that some unusual spikes (long tails in the distributions) are at the source in the case of the TMT industry. For the EnU sector, and less markedly for the consumer non-cyclical goods (CNC) sector, the distribution of the errors is clearly skewed to the left, suggesting that a number of "negative" shocks have not been accounted for yet.

It is worth noting that, overall, error distributions are not skewed. Lack of residual skewness, in contrast to lack of kurtosis, is identified by Hendry and Juselius (2000b, p 7) as an important requirement for correct model specification. Statistical inference is moderately more robust to the validity of the latter.

The eigenvalues of the companion matrix suggest that the system is stable (see, for example, Hendry and Juselius (2000b, Section 3.4)), as all of the 14 ($2 * p = 14$) moduli of the eigenvalues of the companion matrix are inside the unit circle (two moduli are close to 0.98, and four above 0.92). The fact that a number of eigenvalues are close to the unit circle also indicates the possibility of a stochastic trend, and suggests that the processes may be cointegrated.

---

31 The econometric implications of using indicators (dummy) variables are discussed in Doornik et al (1998, Section 2.2).

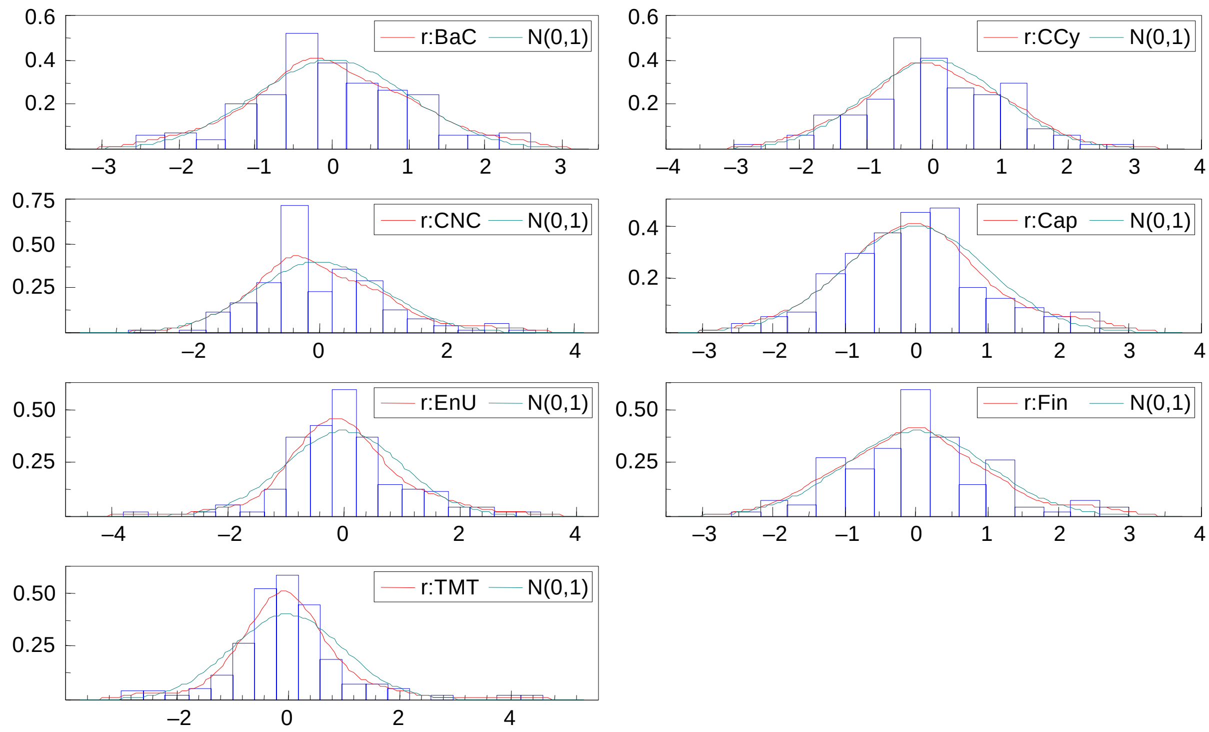
32 For example, within FinT June has weight 0.25, September 2002 1, October 2002 –0.74, November 2002 –0.5, December 2002 0.25, March 2003 0.25, and April 2003 –0.75.

33 The very peculiar behaviour of the TMT sector requires some special attention. Clearly, developments in this sector have influenced to a significant degree risk in financial markets after March 2000.

34 The recursive estimation was carried out over 50 periods. Only the TMT sector has a number of spikes outside the two-standard deviation benchmark for the one-step residuals. These are only towards the end of the sample, denoting the profound change that has taken place in this sector since late 2000.

Graph 8

Residual error density in the closed model with deterministic components



By way of a prelude to the cointegration analysis of the full model, we briefly look at the integrated nature of the risk series. Testing the cointegration rank *r* of the closed system suggests the presence of two cointegration relationships between the risk processes ($r = 2$).[35] The results are tabulated in Table 9 below.

Table 9

Cointegration analysis of the closed model

rank (*i*)	λ_i	**loglik**	H_o: $r \leq i$			
			Trace test		**Max test**	
0		2019.149	173.01	[0.000]**	61.70	[0.000]**
1	0.38017	2051.675	111.30	[0.002]**	48.00	[0.003]**
2	0.31070	2076.976	63.31	[0.147]	33.41	[0.053]
3	0.22815	2094.585	29.90	[0.725]	14.27	[0.800]
4	0.10475	2102.109	15.63	[0.743]	7.39	[0.928]
5	0.055647	2106.003	8.24	[0.447]	4.40	[0.811]
6	0.033546	2108.323	3.84	[0.050]	3.84	[0.050]
7	0.029321	2110.347				

[35] The sequence of trace tests used in the determination of the cointegration rank represents a consistent procedure, as elaborated in Doornik and Hendry (2001). We report in Table 9 the T-nm tests.

Both the rank trace and maximum likelihood tests detect at least one cointegrating relationship and possibly two (the third one is only detected by one test). We consider two cointegrating relationships and will turn to a more refined procedure for selecting the cointegrating rank for the full model. On the basis of the rank estimation, the cointegrated model (rank 2) yields cointegration matrices α and β given in Table 10.

Table 10

Cointegration matrices of the closed model

	β		α	
BaC	1.00	–0.97	–0.064 (0.029)	–0.013 (0.050)
CCy	–3.71	1.00	–0.029 (0.022)	–0.124 (0.039)
CNC	–0.12	–1.14	0.006 (0.014)	–0.053 (0.025)
Cap	1.12	0.62	–0.185 (0.034)	–0.167 (0.060)
EnU	5.45	1.76	–0.025 (0.011)	–0.036 (0.018)
Fin	–2.56	–0.75	–0.001 (0.008)	0.024 (0.014)
TMT	0.11	–0.20	–0.388 (0.072)	0.107 (0.124)

Two groups of industries are identified by the two cointegrating relationships β in the closed system: the BaC, CNC, Cap and TMT sectors on the one hand, and the CCy, EnU and Fin sectors on the other. Within the first group, BaC, CNC and TMT are "substitutes" for each other (their β coefficients share a common sign in each vector β) and "complements" to Cap *across* cointegrating relationships. In the second group, CCy and Fin are "substitutes" for each other but "complements" to EnU across relationships. The complementarity/substitutability relationship is constant within groups across cointegrating vectors, but reverts between groups *across* cointegrating relationships. This suggests that the cointegrating relationships capture distinct effects of different types of shocks on the correlation across groups (even though they affect members within any one group equally). The degree of this complementarity/substitutability is slightly different in each cointegrating relationship, also pointing to the distinct magnitude of the shocks' impact.

The adjustment to the different cointegrating relationships is also revealing. Only the Cap, Fin and TMT sectors adjust to the first error correction relationship (as denoted by the significance of the α coefficients in each equation). The CCy, CNC and (again) Cap and TMT adjust to the other relationship. The BaC and EnU sectors do not appear to adjust to either of the error correction factors, and are therefore weakly exogenous, ie risk levels in the latter sectors do not respond to deviations from long-term risk "equilibria". Both sectors being at the first stage of the production chain suggests that they enjoy greater independence from the economic relationships tying the remaining sectors' long-term equilibria. A closer look at the sectors that adjust to deviations from long-term equilibria could provide an indication of the nature of the cointegrating factor. In this light, the latter group (CCY, CNC, Cap and TMT) appears to capture a cointegration resulting from consumption, whereas the second group (Cap, Fin and TMT) one from investment cycles.

Appendix C:
A forecasting framework

The error correction specification (cointegration relation) significantly restricts the model and its forecasts, as the variables in the cointegrating relationship will adjust to their long-run equilibrium. This is an important factor driving some of the short-run dynamics. Naturally, the forecasts are affected by this characteristic, tending to convergence to the level specified by the long-run components. Importantly from the results of the previous section, this long-run equilibrium will also embody the path that we exogenously assumed for the systemic variables, as they will drive the stochastic trend.

In order to make a forward assessment of the evolution of risk in the different industries, we require assumptions about the behaviour of the exogenous variables in our model. Indeed, Graph 2 above presents a *baseline scenario* for the six months following May 2003 (the last date for which data are available on EDFs), together with a *scenario of a deepening recession*. Future values on industrial output, oil prices and the US stock exchange consistent with these two scenarios are fed into the model to obtain out-of-sample forecasts.[36]

C.1 Forecasts with the integrated series

We first look at the forecasts from the restricted integrated model. The implied baseline scenario risk measures from June to November 2003 are displayed in Graph 9, from where it is clear that the model foresees the gradual reduction of the expected default frequency. This trend extends the very strong correction in April and May 2003, whose impact appears to persistently drive risk down across industries.

Standard error bars[37] are displayed around the forecast values to illustrate their uncertainty. Whereas the model suggests that risk will decrease to different degrees in all sectors, significant uncertainty surrounds the forecast values. With the exception of the Fin sector, the model still considers a deterioration in risk possible within a standard deviation. It must be noted, however, that much of the improvement forecast in the risk series following the substantial easing in risk that materialised in April and May 2003 is also reinforced by the high persistence in the industrial risk measures.

Indeed, much of the same persistence drives the forecasts under the more adverse recession scenario, where much of the same pattern applies to forecast industry risk. The high persistence in the risk series dominates the negative pull of the assumed systemic variables. Whereas risk in all industries is higher than under the baseline scenario, the change is small in relation to the levels in the risk measures. These forecasts are shown in Graph 10.

As already emphasised in Section 3.4 above, the limited effect of the systemic factors is also present in the forecasts: substantially distinct paths for the exogenous variables do not cause a reversal in the trend of the risk measures. However, their impact is not negligible, as noted by the higher forecast risk levels across industries under the recession scenarios. In particular, the possibility of a trend reversal in the financial sector (Fin) is well within a standard deviation under the more adverse conditions.

C.2 Forecasts with non-integrated series (in the $I(0)$ space)

The forecasting in the $I(0)$ space should be more accurate, as the cointegrating vectors become endogenous in a simultaneous equation model, and indeed the pattern for the risk forecasts depicted has some interesting differences.

The baseline forecast in the $I(0)$ space is given in Graph 11, showing no remarkable differences with the regression in the integrated space shown in Graph 9.

Similar comments apply to the forecast under the recessionary scenario.

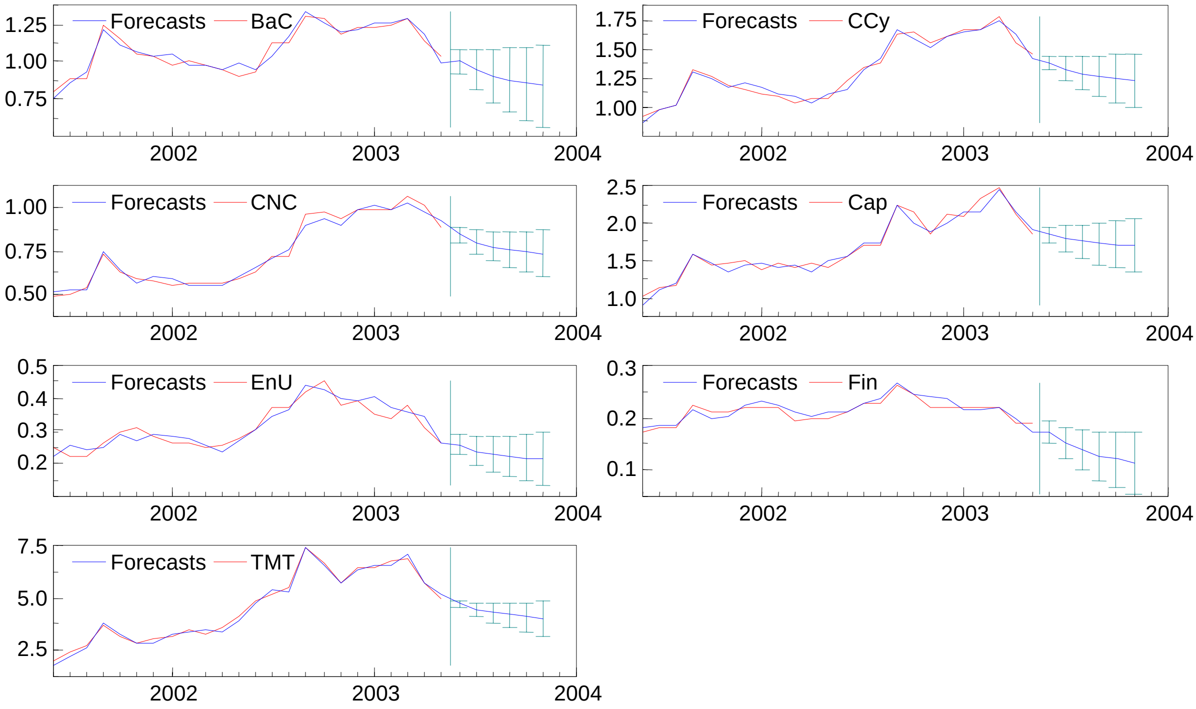
---

[36] The baseline scenario is based on the April 2003 ECB forecast exercise.

[37] On the basis of error variance only, ie does not include parameter uncertainty.

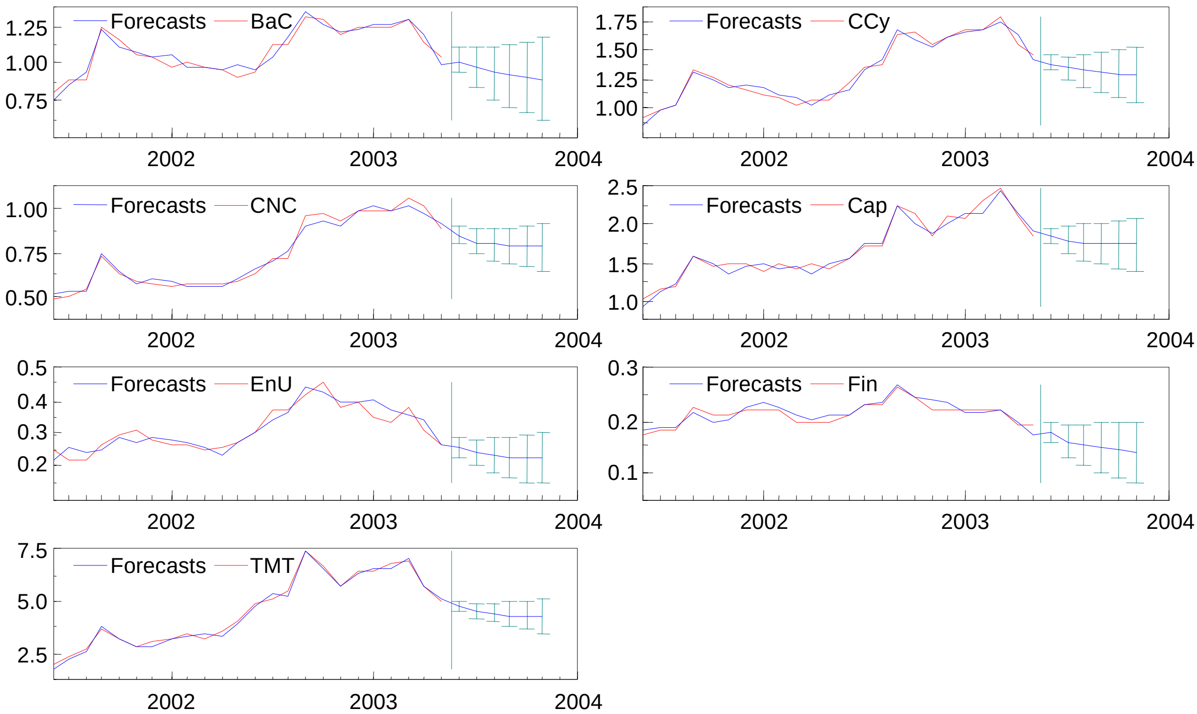
Graph 9

Baseline out-of-sample forecasts for industrial risk measures



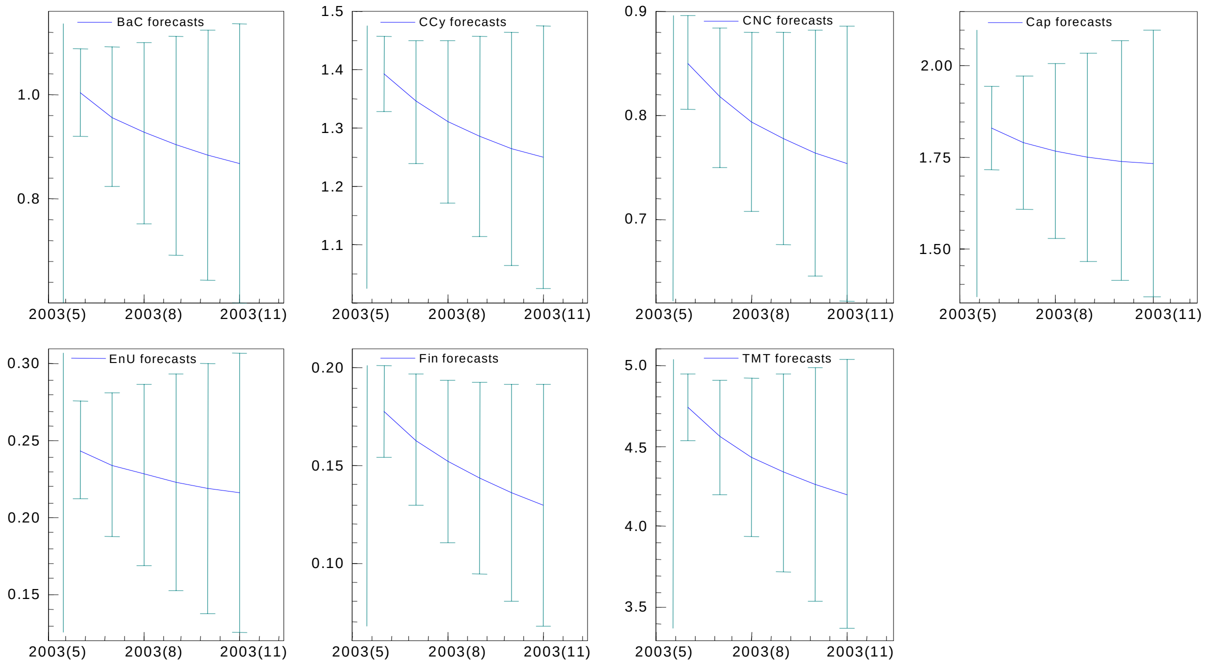
Graph 10

Recessionary out-of-sample forecasts for industrial risk measures



Graph 11

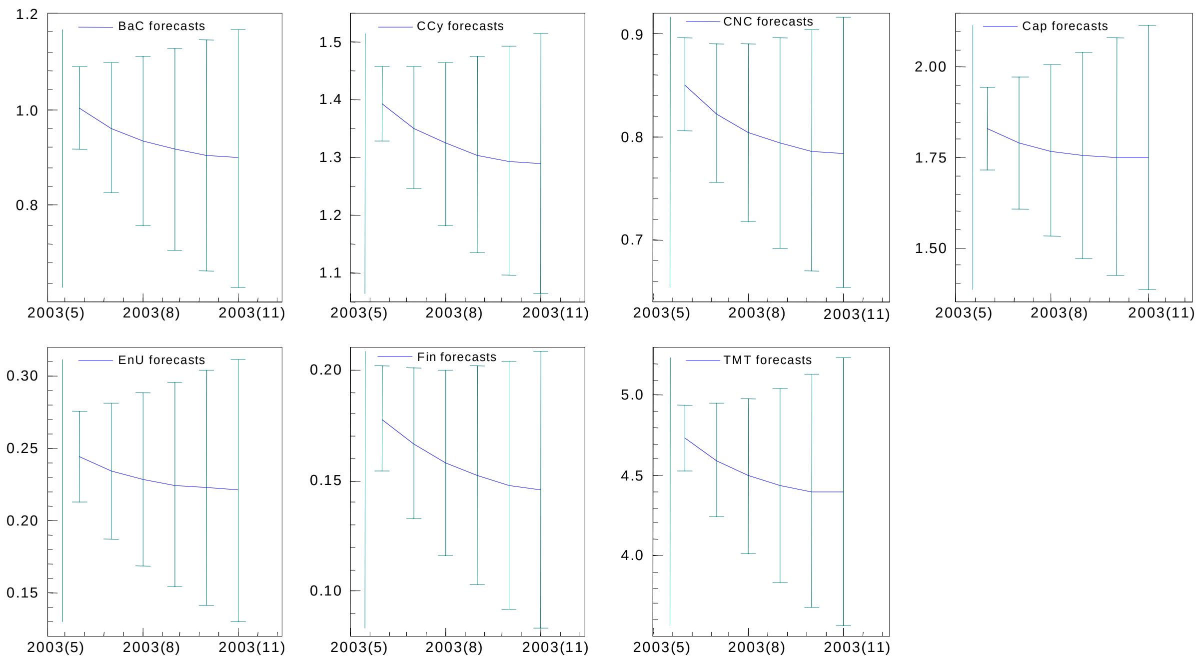
Baseline out-of-sample forecasts for industrial risk measures with error correction



Overall, the combined effect of strong common factors driving the dynamics of the system and the significant feedback mechanisms linking risk across sectors appear as overwhelmingly more predominant in determining the short outlook of sectoral risk.

Graph 12

Recessionary out-of-sample forecasts for industrial risk measures with error correction



References

Allen, L and A Saunders (2002): *A survey of cyclical effects in credit risk measurement models*, Stern School of Business, NYU, New York, discussion paper.

Bangia, A, F X Diebold, A Kronimus, C Schagen and T Schuermann (2002): "Ratings migrations and the business cycle, with applications to credit portfolio stress testing", *Journal of Banking and Finance*, 26(2-3), pp 445-74.

Carey, M (2002): "A guide to choosing absolute bank capital requirements", *Journal of Banking and Finance*, 26(5), pp 929-51.

Crosbie, P J and J R Bohn (2002): *Modelling default risk*, Moody's KMV, San Francisco.

Crouhy, M, D Galai and R Mark (2000): "A comparative analysis of current credit risk models", *Journal of Business and Finance*, 24(1-2), pp 59-117.

Delianedis, G and R Geske (1998): *Credit risk and risk neutral default probabilities: information about rating migrations and defaults*, Anderton Graduate School of Business, UCLA, Los Angeles.

Doornik, J A (1996): *Testing vector autocorrelation and heteroskedasticity in dynamic models*, Nuffield College, discussion paper.

Doornik, J A and D F Hendry (2001): *Modelling dynamic systems using PcGive 10*, vol II, Timberlake Consultants, third edition.

Doornik, J A, D F Hendry and B Nielsen (1998): "Inference in cointegrating models: UK M1 revisited", *Journal of Economic Surveys*, 12(5), pp 117-56.

Duffie, D and K J Singleton (2003): "Credit risk: pricing, measurement and management", *Princeton Series in Finance*, Princeton University Press, Princeton, NJ.

Enders, W (1996): *RATS handbook for econometric time series*, John Wiley & Sons, New York.

Fons, J S (1991): *An approach to forecasting default rates*, Moody's Investors Service, New York, discussion paper.

Friedson, M, C Garman and S Wu (1997): "Real interest rates and the default rates on high-yield bonds", *Journal of Fixed Income*, pp 27-34.

Gordy, M B (2000): "A comparative anatomy of credit risk models", *Journal of Business and Finance*, 24(1-2), pp 119-49.

Gropp, R, J Vesala and G Vulpes (2002): "Equity and bond market signals as leading indicators of bank fragility", *European Central Bank Working Paper Series*, 150.

Hansen, H and K Juselius (2002): *CATS in RATS: cointegration analysis of time series*, Estima.

Helwege, J and P Kleiman (1996): "Understanding aggregate default rates of high yield bonds", *Current Issues in Economics and Finance*, 2(6), pp 1-7.

Hendry, D F and K Juselius (2000a): "Explaining cointegration analysis: part I", *Energy Journal*, 21(1), pp 1-41.

——— (2000b): *Explaining cointegration analysis: part II*, Department of Economics, University of Copenhagen, discussion paper.

Ho, M S and B E Sorensen (1996): "Finding cointegration rank in high dimensional systems using the Johansen test: an illustration using data based Monte Carlo simulations", *The Review of Economics and Statistics*, 78(4), pp 726-32.

Jarrow, R A and F Yu (2001): "Counterparty risk and the pricing of defaultable securities", *Journal of Finance*, 56(5), pp 1765-99.

Johansen, S (2000): "Modelling of cointegration in the vector autoregressive model", *Economic Modelling*, 17(3), pp 359-73.

Jonsson, J G and M S Fridson (1996): "Forecasting default rates on high-yield bonds", *Journal of Fixed Income*, pp 66-77.

Koyluoglu, H U and A Hickman (1998): “A generalized framework for credit risk portfolio models”, *Default Risk*, discussion paper.

Kurbat, M and I Korablev (2002): “Methodology for testing the level of the EDF credit measure”, *Technical Report 020729*, Moody’s KMV, San Francisco.

Morin, S (2003): *On the modelling of EU banks’ loan loss provisions*, European Central Bank, Frankfurt am Main, discussion paper.

Pesaran, M H, T Schuermann, B J Treutler and S M Weiner (2003): *Macroeconomic dynamics and credit risk: a global perspective*, University of Cambridge, Cambridge, discussion paper.

Pesaran, M H and R P Smith (1998): “Structural analysis of cointegrating VARs”, *Journal of Economic Surveys*, 12(5), pp 55-90.

Rose, C and M D Smith (2002): *Mathematical statistics with mathematica*, Springer Verlag.

Saunders, A and L Allen (2002): *Credit risk measurement - new approaches to value at risk and other paradigms*, John Wiley & Sons, New York, second edition.

Trucharte, C and J Saurina (2002): *Analysis and prediction of the non-performing loans ratio for the Spanish banks: a possible approach for stress test exercises*, Banco de España, Madrid, discussion paper.

Wilson, T (1997a): “Portfolio credit risk (I)”, *Risk*, 10(9), pp 111-17.

——— (1997b): “Portfolio credit risk (II)”, *Risk*, 10(10), pp 56-61.

Macroeconomic conditions and banking performance in Hong Kong SAR: a panel data study

Stefan Gerlach, Wensheng Peng and Chang Shu[1]
Hong Kong Monetary Authority

1. Introduction

This paper provides a preliminary study of the determination of the net interest margin and the non-performing loan ratio, arguably the two most important measures or determinants of bank profitability, for all 29 retail banks in Hong Kong SAR. This sector does not include banks whose activities are primarily of an offshore or wholesale nature, and is thus representative of the banking business in Hong Kong. The study is based on annual data for the period 1994-2002 that are derived from information collected in the context of the supervisory activities of the Hong Kong Monetary Authority (HKMA). While the data set is rich in some dimensions (for instance, it contains data on many financial and income and expense ratios, such as non-performing loans and net interest margins), for confidentiality reasons it contains no information that would allow us to identify individual banks. Thus, we do not know how large a bank's assets are (although we do know if it is "small", "medium-sized" or "large"), how extensive a branch network it has, whether it is domestic or foreign-owned, etc.

The focus of the analysis is on the extent to which macroeconomic developments affect bank profitability and, in particular, whether that impact differs across banks. The paper is motivated by the fact that the banking sector plays a critical role in the economy. A strong and profitable banking system promotes broader financial stability and increases the economy's resilience to adverse macroeconomic shocks. At the same time, changes in macroeconomic conditions affect banks' performance and financial health. It is therefore of importance for the authorities responsible for the maintenance of financial and monetary stability to quantify the linkages between macroeconomic developments and the banking sector.

In the case of Hong Kong, this interest is enhanced by the fact that the Hong Kong dollar is linked to the US dollar through a currency board system, which implies that local interest rates are effectively beyond the immediate control of the HKMA. While this system has provided a firm nominal anchor to the economy since its introduction in 1983, monetary policy cannot be used to guard against large asset price swings. In particular, interest rates cannot be adjusted in the light of the state of the banking system. The currency board system therefore requires a careful use of regulatory policy and a strict regime of banking supervision. The effectiveness of this policy is evidenced most strikingly by the fact that the banking system remains generally sound despite a fall in property prices of almost 70% since 1998. A thorough understanding of the impact of business cycle movements on bank profitability is therefore of considerable interest.

There are a number of studies on banking performance in Hong Kong, most of which use aggregate data for the banking system. In particular, Shu (2002) examines the impact of macroeconomic conditions on the average asset quality of the banking sector. Peng et al (2003) study how changes in the Hong Kong dollar risk premium, measured by a widening of spreads between Hong Kong dollar and US dollar interest rates, may have influenced banks' aggregate net interest margin and asset quality. Gerlach and Peng (2003) find that bank lending is closely related to economic growth and fluctuations in property prices, and that regulatory measures have helped limit banks' exposure to swings in the property market. Two studies, Kwan (2002) and Jiang et al (2003), have used panel

[1] Corresponding author: Chang Shu, Research Department, HKMA, 84th floor, Two International Finance Centre, 8 Finance Street, tel: (852)2878 1657, fax: (852)2878 1897, e-mail: chang_shu@hkma.gov.hk. The views expressed in this paper are solely the authors' and not necessarily those of the HKMA.

data. By estimating cost frontiers, Kwan considers how the cost-efficiency of banks is determined by bank characteristics. In a paper closely related to this, Jiang et al relate bank profitability to macroeconomic conditions as well as bank characteristics. However, with access to public data on listed banks only, it covers a subset of the sector. Moreover, it does not include an analysis of any asymmetric effects of changes in macroeconomic and financial conditions across banks, because of data limitations.

The rest of the paper is organised as follows. Section 2 provides some stylised facts about the performance of Hong Kong's banking sector in recent years, and its relationship with macroeconomic developments. We show that changes in profitability are closely linked to the net interest margin and to the non-performing loan (NPL) ratio, which influences banks' provisioning decisions. Section 3 outlines the empirical strategy used in modelling these key determinants of profitability. Given that we have data for a cross section of banks for a number of years, we use a panel data approach that is common in studies of banking performance. Section 4 presents the estimation results and analysis. The main findings are that macroeconomic developments have played a large role in determining the profitability of banks in Hong Kong. Furthermore, the NPLs of smaller banks appear less sensitive to movements in real GDP than those of larger banks, but their net interest margin appears more sensitive. We also find, perhaps surprisingly, that the NPLs of banks holding more property loans have been relatively insensitive to property prices. Section 5 concludes.

2. Banking performance in Hong Kong: some stylised facts

While work to date has concentrated on studying developments in Hong Kong's banking sector as a whole, the focus of this paper is to explore whether larger and smaller banks are affected to different extents by macroeconomic conditions. For this purpose, the 29 banks are divided into three groups according to their asset size. The first of these groups contains five "large" banks defined as those with assets accounting for more than 5% of the retail bank sector; the second group contains 10 banks with an asset size representing between 5 and 1% of the sector; and the small bank group contains 14 banks with an asset size of less than 1% of the sector.

2.1 Profitability and the macroeconomic environment

To understand the role of macroeconomic factors in accounting for movements in profitability, it is useful to consider the macroeconomic indicators in Graph 1. Following a pronounced expansion in the mid-1990s, the Hong Kong economy fell into a recession as a result of the Asian financial crisis, with real GDP declining by over 5% in 1998. The economy rebounded strongly in 2000, but the recovery ended with the global economic slowdown in 2001. Subsequently, economic activity was generally sluggish notwithstanding strong performance in exports of goods and services. The developments also had a strong impact on the unemployment rate, which rose sharply from 2-3% in the pre-crisis period to 7.3% in 2002. Affected by both cyclical and structural factors, deflation started in 1998, and has persisted for over five years. Since bank loans are in nominal terms, an unexpected decline in the price level will increase the real debt burden, and may therefore affect borrowers' ability to repay and hence bank profitability. Furthermore, property prices have declined by over 60% from the pre-crisis peak level, exerting a significantly negative wealth effect on domestic demand. In addition to the impact through general macroeconomic performance, declines in property prices may have affected banks' profitability directly through a number of channels. These include deterioration in the quality of property-related assets such as mortgage loans and reduced demand for credit.

Graph 1 also shows that interest rates rose sharply during the Asian financial crisis, reflecting an increased risk premium.[2] Empirical estimates suggest that the spike in interest rates in 1997-98 reduced banks' net interest margins because of a faster and more complete pass-through to deposit rates than to retail lending rates (Peng et al (2003)). Helped by improved global market conditions as

[2] Since we plot annual data, the chart does not show the sharp increase in interest rates that occurred during the episode of severe speculative pressures in the autumn of 1998.

well as by a number of steps (the "seven technical measures") taken by the HKMA to strengthen the currency board system, interest rates subsequently stabilised. In recent years they have declined in line with the monetary easing in the United States. Despite these developments, real interest rates have remained high by historical standards as a result of deflation, and have been partly responsible for restraining the demand for bank credit.

It should be noted that these difficult macroeconomic conditions also coincided with interest rate liberalisation and increased competition in the banking sector that in turn led to changes in the structure of the banking system. Starting from 1994, the HKMA lifted rules on interest rates in stages. The liberalisation programme, coupled with the reduced demand for credit, has increased competition among banks, which can be seen from the downward trend in the Herfindahl-Hirschman index (Graph 2A).[3] The increased competition has led to a decline in lending spreads, particularly in the mortgage loan segment.[4] At the beginning of 1997, 84% of new residential mortgages were contracted at rates above the best lending rate (BLR). In contrast, nearly all new mortgage loans were made at rates below the BLR by about 2.5 percentage points in 2002 (HKMA (2002)).

2.2 Developments in profitability

As a preliminary to the discussion of profitability below, it is useful to consider what factors contributed, in an accounting sense, to movements in profitability. In accounting terms, profitability can be decomposed as:

$$\frac{BTP}{TA} = \frac{NI}{TA} + \frac{NII}{TA} - \frac{OV}{TA} - \frac{PROV}{TA}, \qquad (1)$$

where *BTP* denotes before-tax profits, *TA* total assets, *NI* net interest income, *NII* non-interest income, *OV* overhead and *PROV* loan loss provisioning.

Of the four components, much interest has focused on the ratio of net interest income to total assets, which is commonly referred to as the net interest margin (*NIM* = *NI*/*TA*).[5] Graph 2B depicts developments in overall profitability using the four components. Profitability for retail banks, defined as before-tax profits divided by total assets, fell sharply from around 1.8% during the boom period (1994-97) to 1% in 1999. It subsequently rebounded and reached about 1.4% in 2002. Variations in profitability appear to have been mainly driven by net interest income and loan provisions. Specifically, NIMs fell significantly in 1997-98 as the economy contracted and banks' funding costs soared. They recovered moderately between 1999 and 2000, but the subsequent economic slowdown and intense competition in the sector restrained any further improvement. By comparison, non-interest income (*NII*/*TA*) and overhead costs (*OV*/*TA*) have remained relatively stable.[6]

Graph 2B also shows that banks' loan loss provisions (*PROV*/*TA*) increased considerably in 1998-99 as asset quality deteriorated substantially. The sharp slowdown of the economy and higher borrowing costs caused severe financial difficulties for corporate and individual borrowers. The collapse of a number of large Mainland Chinese companies in 1998 exacerbated the situation. Provisions and non-performing loans declined in 2000-02 (Graph 2C), reflecting a number of factors including the

---

3 The Herfindahl-Hirschman index is an indicator of market concentration. It is calculated as the sum of the squares of individual banks' market shares.

4 Chart 2A includes a measure of the lending spread, which is calculated as the difference between the rate on new mortgage loans and a (weighted) average of deposit rates.

5 The *NIM* is the ex post spread, which differs from the ex ante spread calculated as the difference between the contractual rates charged on loans and rates paid on deposits. The ex post spread is more useful as it controls for the fact that banks with high-yield, risky credits are likely to face more defaults. Other things being equal, higher *NIM*s as a result of, for example, a fall in loan defaults, will increase bank profits, and thus improve the stability of the banking sector. However, a higher *NIM* may also reflect high intermediation costs due to insufficient competition or other institutional characteristics, and thus indicate inefficiency of the system.

6 The stability of non-interest income and overhead costs at the aggregate level obscures the fact that larger banks may be better able than smaller banks to manage these components in a countercyclical fashion to smooth profitability over time (see below).

economic recovery in 2000, and a more cautious lending stance by banks. Nevertheless, the NPL ratio remained higher than the pre-crisis levels.

2.3 Bank groups of different sizes

Developments in bank profits also vary across groups of banks classified by size. Graph 3A shows that while average profitability has been quite similar for different banks, the sensitivity of bank profitability to the state of the economy is inversely related to bank size. Thus, during the boom period of 1994-97, small banks were more profitable than larger banks. By contrast, during 1998-2002, when economic conditions were generally weaker, average profitability declined most in small banks. Although the profitability of smaller banks appears relatively more volatile than that of larger banks, the striking aspect of Graph 3A is that banks have generally remained profitable in recent years despite the very difficult market conditions.

Decomposition of profitability in the previous subsection suggested that movements in NIMs played a large role in accounting for shifts in bank profitability at sectoral level. Graph 3B therefore looks at the NIM by bank size, and shows that smaller banks generally maintained higher NIMs, but saw the largest declines in NIMs after 1997. A number of factors may explain the generally higher NIMs for smaller banks. First, they tended to have lower funding costs, as reflected in higher capital bases, and rely more on traditional lending business on the asset side, which led to a relatively high interest income as a share of total income (Table 1). The fact that smaller banks hold more capital should perhaps best be seen as recognition that their higher profit volatility may be associated with greater riskiness. Second, it may be the case that a higher NIM is associated with a higher loan risk profile, which raises operating costs entailed by monitoring and control. The small bank group indeed recorded higher operating costs in the period. The sharp decline in NIMs for smaller banks in recent years may reflect the relatively large weight of property-related loans in their portfolio, as lending spreads for mortgage loans declined significantly. Another possibility is that increased competition has required smaller banks to offer higher interest rates to attract customer deposits, and thus reduced their NIMs.

Next we turn to the NPL ratio. Graph 3C shows that loan quality worsened considerably for all three groups in 1998-2002 relative to 1994-97. Medium-sized banks saw the worst deterioration, and large banks recorded a slightly larger rise in NPLs than the small bank group. The bursting of the property "bubble" probably put the asset quality of the sector under significant stress. Banks in all groups had significant exposure to property lending, which accounted for around 50% of their portfolio. Although there was no systematic pattern as to which bank group was more exposed, the degree of exposure to property lending varied across banks. It should be noted, however, that a few factors mitigated the concentration risk associated with large exposure to the property sector. These factors included banks' observance of the HKMA's recommended loan-to-value ratio of 70% for residential mortgages, the low gearing ratio of property developers and the practice of pre-selling a large number of units (IMF (1999)). As a result, the delinquency ratio of residential mortgage loans has remained low relative to that of most other domestic credits.

Graph 3D indicates that non-interest income net of operating costs increased for large and medium-sized banks in 1998-2002 over 1994-97, but declined for the small bank group. This confirms that larger banks have managed to raise non-interest income and reduce operating costs in recent years to stabilise profits in the face of declining net interest income and increasing loss provisions.

Table 2 further shows the dispersion of profitability, asset quality and the NIM across the banks. The cross-bank dispersion of these variables rose in 1998, but started to fall back in 2002.

2.4 Summary

The analysis in this section suggests three broad conclusions. First, overall bank profitability dropped sharply following the Asian financial crisis and, notwithstanding some recovery in recent years, has remained below pre-crisis levels. The reduced profitability is related to relatively difficult macroeconomic conditions and increased competition in the banking sector. Second, bank profitability has been driven mainly by changes in NIMs and loan provisions that in turn were determined by asset quality. Third, smaller banks have recorded relatively larger declines in profits, attributable to a sharper fall in net interest margins as well as to rises in operating costs.

3. Empirical framework and methods

In the remainder of this paper, we carry out econometric analysis to examine how macroeconomic and financial conditions may have affected NPLs and NIMs, the two most important factors affecting bank profits in Hong Kong. Since we are interested in the behaviour of individual banks, it is natural to adopt a panel approach. We briefly describe the empirical framework and the estimation method used below.

Following Demirgüç-Kunt and Huizinga (1999, 2000) and similar studies in this area, asset quality, measured by NPLs for bank *i* at time *t* ($NPL_{i,t}$), is determined as follows:

$$NPL_{i,t} = f(MACRO_t, FIN_t, BANK_{i,t}) + error_{i,t}, \tag{2}$$

where *NPL* is the ratio of NPLs to total loans. *MACRO* denotes a set of macroeconomic variables reflecting the state of the economy, eg economic growth and inflation, *FIN* includes financial variables such as interest rates and changes in property prices, and *BANK* contains bank-specific variables such as asset size and sectoral concentration in lending. In particular, we examine whether shares of property-related and consumer loans affect the NPL ratio.

As there is no reason why the macroeconomic factors and financial variables must have the same impact on all banks, it is of interest to allow for interaction between the different variables used. For example, to test whether the impact varies systematically across banks, we include an interactive term between, on the one hand, the macroeconomic and financial variables and, on the other, the variable capturing the size of the bank. We also interact changes in property prices with the share of property-related lending in a bank's portfolio to examine how banks with different exposures to the real estate sector were affected by declines in property prices.

Similarly, the NIM equation is specified as:

$$NIM_{i,t} = g(MACRO_t, FIN_t, BANK_{i,t}) + error_{i,t}. \tag{3}$$

We consider a number of bank-specific variables that can be divided into three groups: (a) variables capturing the structure of assets and liabilities; (b) variables capturing the structure of income and expenses; and (c) sector concentration. As in equation (2), interactions between *BANK*, *MACRO* and *FIN* variables are allowed.

4. Empirical findings

4.1 Asset quality

Some estimated specifications for the NPL equation are presented in Table 3. The sample comprises 27 banks, since the NPL series are not available for two banks in the sample. We estimate all equations twice: first with a common intercept and then allowing for fixed effects.[7] The last two lines of the table give the test statistic and the associated p-value for a test for a common intercept.[8] As can

---

[7] The inclusion of a lagged dependent variable renders both the pooled and fixed effects estimators biased. Although, in our case, the time series dimension is not very small relative to the cross-sectional dimension, the bias can still be sizeable (Judson and Owen (1999)). Various methods have been developed to address this issue. Anderson and Hsiao (1981) suggest an instrumental variable (IV) estimation method that will lead to consistent estimates. Arellano and Bond (1991) propose a generalised method of moments (GMM) procedure that is more efficient than that of Anderson and Hsiao (1981). This literature is further generalised and developed by Ahn and Schmidt (1995), Arellano and Bover (1995) and Blundell and Bond (1998) to mention a few. In future work on more detailed data we intend to explore the importance of better estimation techniques.

[8] The test for a common constant for a panel model is often referred to as the test for fixed or individual effects. It is carried out by performing an F-test:

$$F = \frac{(RRSS - URSS)/(N-1)}{URSS/(Obs - N - K)} \sim F_{N-1, Obs-N-K}$$

The restricted model is the pooled regression, while the unrestricted model is the fixed effects model. *RRSS* and *URSS* are the residual sum of squares of the restricted and unrestricted models respectively, *N* is the number of banks, *Obs* the be seen, that hypothesis is rejected in all cases. Consequently, we only report results for the fixed effects regressions.

We first estimate the most general specification (Model 1), which encompasses all macroeconomic, financial and bank variables, but does not allow for any interaction. The results indicate that the variables measuring the shares of property-related (*PROP SHARE*) and consumer (*CONS SHARE*) lending are not significant. In Model 2, in which we exclude these two variables, all macroeconomic and financial variables are highly significant and have expected signs. Thus, increases in GDP growth (*GDP*), inflation (*INF*) and the rate of change of property prices (*PROP*) all reduce NPLs. By contrast, rises in short-term interest rates (*HIBOR*) increase NPLs.

While interesting, this model does not allow for any interaction between the macroeconomic/financial variables and bank characteristics. In Model 3 we therefore interact the macroeconomic and financial variables with bank size, which is arguably the single most important bank characteristic. This general model has a higher adjusted R^2 compared to the two previous models, suggesting that inclusion of the interactive terms improves the fit of the equation. However, a number of variables are not significant. In Model 4, we interact property price inflation with the share of property lending in total loans instead of size. This specification further improves the fit of the NPL model as evidenced by the adjusted R^2, which increases from 0.91 in Model 3 to 0.94. The final specification, Model 5, is obtained by eliminating the two insignificant variables in Model 4. Although the adjusted R^2 of Model 5 falls somewhat, all the remaining variables are highly significant.

Based on the specification of Model 5, a number of observations are worth noting. First, both *GDP* and *GDP*SIZE* are significant.[9] However, since the parameter on the interactive term is negative, the results suggest, perhaps surprisingly, that economic growth reduces NPLs of all bank groups, but more so for larger banks. This matches poorly with the earlier observation that asset quality of smaller banks deteriorated more than that of large banks in recent years. However, small banks differ from large banks in more ways than merely in size, and we return to this issue below.

Second, higher inflation also lowers NPLs. This may be so because it improves borrowers' ability to meet obligations by eroding the real value of the debt burden. Furthermore, under Hong Kong's currency board regime nominal interest rates are closely tied to US interest rates, implying that increases in inflation reduce the real interest rate. Inflation is also positively correlated with the state of the business cycle and might be interpreted as an additional indicator of the state of the economy.

Third, interest rates are positively related to NPLs. Declines in interest rates reduce the debt servicing burden, thereby helping to protect asset quality.

Fourth, rises in property prices reduce NPLs. One would expect that the size of the impact would depend on banks' exposures to the real estate sector. Thus, on the face of it, the positive sign on the interactive term between changes in property prices and the share of property lending is surprising, as it suggests that the impact is smaller for a larger exposure. However, an alternative explanation is that property prices should be seen as a measure of general economic conditions (rather than as an indicator specific to the property sector) and that property lending is less sensitive to changes in economic conditions than other types of bank credit.[10] As a result, a given change in property prices will affect a bank's NPL ratio less if its property-related lending is relatively large. To visualise this, suppose that the NPL ratio is determined as:

$$NPL_t = \underset{(-)}{\beta}(1 - \omega)X_t + \underset{(-)}{\delta}\omega X_t + \ldots, \tag{4}$$

where:

X_t: changes in property prices;

number of observations, and K the number of regressors. If the null hypothesis of a common intercept is rejected, the fixed effects model should be chosen for estimation.

9 As noted above, for confidentiality reasons we only have series of the weighted average asset size for the three groups, and their averages across time are used in calculating the impact of the growth variable.

10 This accords with our earlier observation that despite declining property prices and weak economic conditions, the default rate of residential mortgage loans has remained low relative to that for most other bank lending.

ω: fraction of loans related to the property sector;

β: sensitivity of NPLs among non-property loans to property prices;

δ: sensitivity of NPLs among property loans to property prices.

The above equation can be rewritten as:

$$NPL_t = [\beta + (\delta - \beta)\omega]X_t \qquad (5)$$

This equation suggests that the impact of changes in property prices varies with ω, and is given by $\beta+(\delta-\beta)\omega$. The term $(\delta-\beta)$ captures relative sensitivity (riskiness) of property loans. Specifically, property loans are less risky (sensitive to property price changes) than other types of lending if $\delta-\beta>0$, which is the case for Hong Kong according to our estimates.

4.2 Net interest margin

Table 4 presents estimates of the NIM equation. We first include all the *MACRO*, *FIN* and *BANK* variables (Model 1).[11] As the model is probably overfitted, only *GDP*, *INF* and *NIEXPENSE* (which we interpret as a measure of banks' operating costs) are significant and have the expected signs. Dropping insignificant variables leads to Model 2 in which *GDP*, *INF*, HIBOR and *NIEXPENSE* remain important and *HIBOR* is also significant. However, the adjusted R^2 declines, suggesting that this model fits less well. In Model 3 we interact *SIZE* with the *MACRO* and *FIN* variables. This model fits better, as indicated by a higher adjusted R^2. All interactive terms in the equation are highly significant, and have the expected signs. This provides strong evidence that the NIMs of smaller banks respond differently to changes in economic conditions than those of larger banks. Finally, the test statistics in the last row of the table confirm that the fixed effects should be allowed for in estimation.

The estimates of Model 3 indicate that economic growth and inflation lead to higher NIMs, probably by reducing NPLs as suggested by the earlier estimates. In addition, loan demand is likely to rise in a period of expansion, giving banks more pricing power in lending. In this light, sluggish economic growth and deflation in recent years have contributed to the narrowing of NIMs.

The interactive terms suggest that the effects of macroeconomic developments on NIMs vary depending on the size of banks, with smaller banks being more affected. It could be the case that when loan demand increases, smaller banks may be prepared to expand lending more aggressively than larger banks by taking on more risky projects with higher returns.

Changes in interest rates also tend to have asymmetric effects across banks. The interactive term between the interest rate and *SIZE* suggests that smaller banks are more affected by changes in interest rates. One explanation for this finding is that the smaller banks have a higher capital base, which reduces overall funding costs. As a result, they can sustain higher NIMs when interest rates rise. To test this hypothesis, an interactive term between the interest rate and the capital base variable is added (Model 4). This variable turns out to be significant and of the expected sign.

Finally, operating costs are found to be positively related to the NIM. There are two possible explanations. First, banks may be able to pass changes in operating costs on to customers by varying lending spreads. Second, a higher NIM may be associated with a higher risk profile of loans. This in turn raises operating costs entailed by monitoring and risk control.

5. Conclusion

Using a confidential supervisory bank-level data set, this paper has examined the determinants of banking performance in Hong Kong SAR, with a focus on the impact of macroeconomic developments

[11] It is difficult to measure changes in the degree of competition in the banking sector. Some preliminary measures such as asset concentration ratios are tried, but turn out to be insignificant.

on NIMs and NPLs. Corroborating earlier studies in the literature, the empirical analysis finds that macroeconomic developments and financial conditions affect banking performance.

A specific focus of the paper was to explore whether bank-specific factors may lead to asymmetric effects of macroeconomic developments across banks. The evidence generally suggests that the NIMs of smaller banks are more, but their NPLs are less, exposed to changes in GDP growth. Understanding the reasons for these differences should be high on the research agenda.

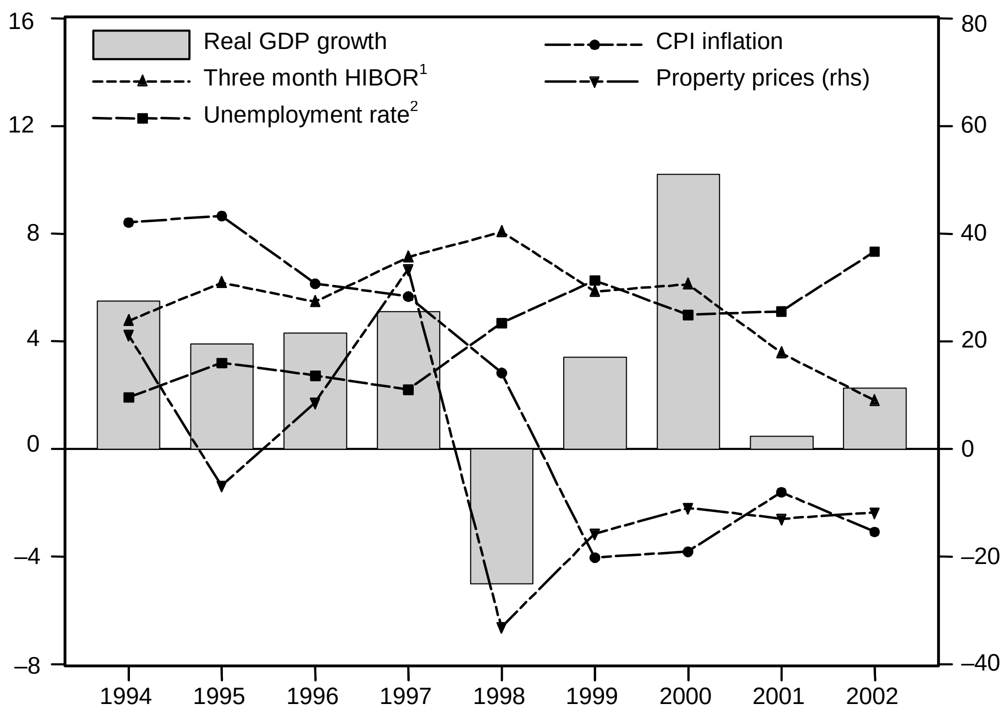
The estimates further suggest that the sharp decline in property prices may have also put banks under stress due to the large exposure to property-related lending. However, property loans appear to be less risky than other types of loans, in that their quality is less sensitive to fluctuations in macroeconomic conditions and property prices. This reflects a combination of factors that mitigate risks associated with property lending, including the HKMA's guideline of a maximum loan-to-value ratio of 70% for residential mortgage loans, and the low gearing ratio of property developers.

This study is preliminary and more work is required. Several extensions seem natural and useful. First, it would be of interest to use quarterly data to obtain a clearer sense of the dynamic responses of bank profitability to movements in real GDP growth and inflation. If real economic growth rebounds in Hong Kong, will banking sector profitability respond after two, four or eight quarters? The annual data used here are too coarse to permit such an analysis. Second, it would be important to explore which macroeconomic time series have the strongest links to the profitability of the banking sector. While we have used real GDP growth, property prices and CPI inflation in this study, it is possible that other time series (such as unemployment and consumption spending) may be more relevant. Third, it would be desirable to sharpen the estimates by taking into account a greater variety of bank characteristics. For instance, do banks with a large number of branches have higher costs and lower profits? Or do banks with a strong retail network obtain funds more cheaply and have greater profits? In future work we hope to shed some light on these issues.

Graph 1

Macroeconomic indicators

Year over year, in per cent

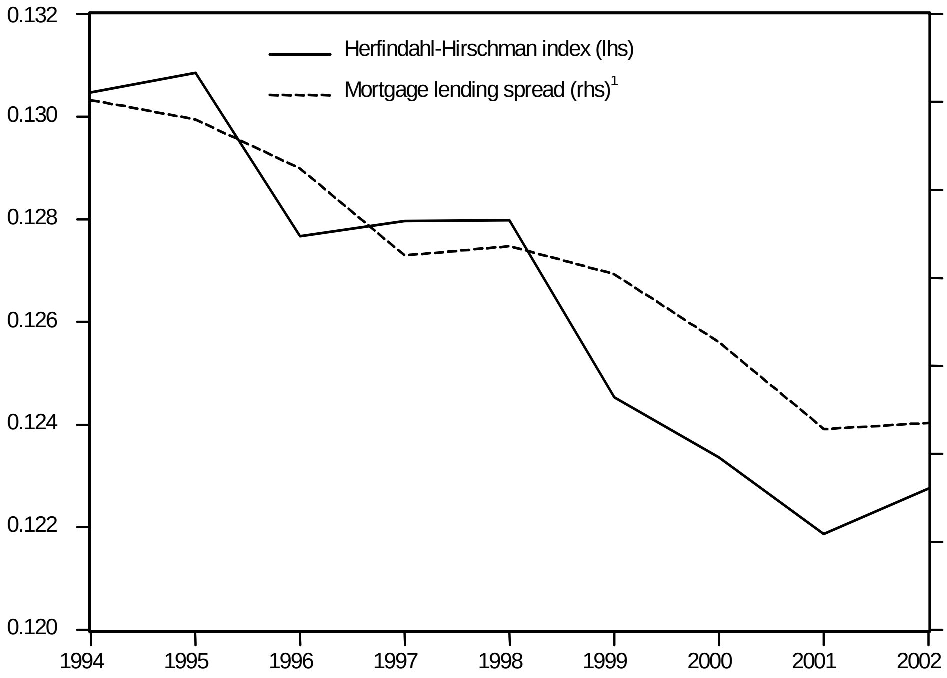


[1] In percent per annum. [2] In per cent.

Graph 2

Bank indicators

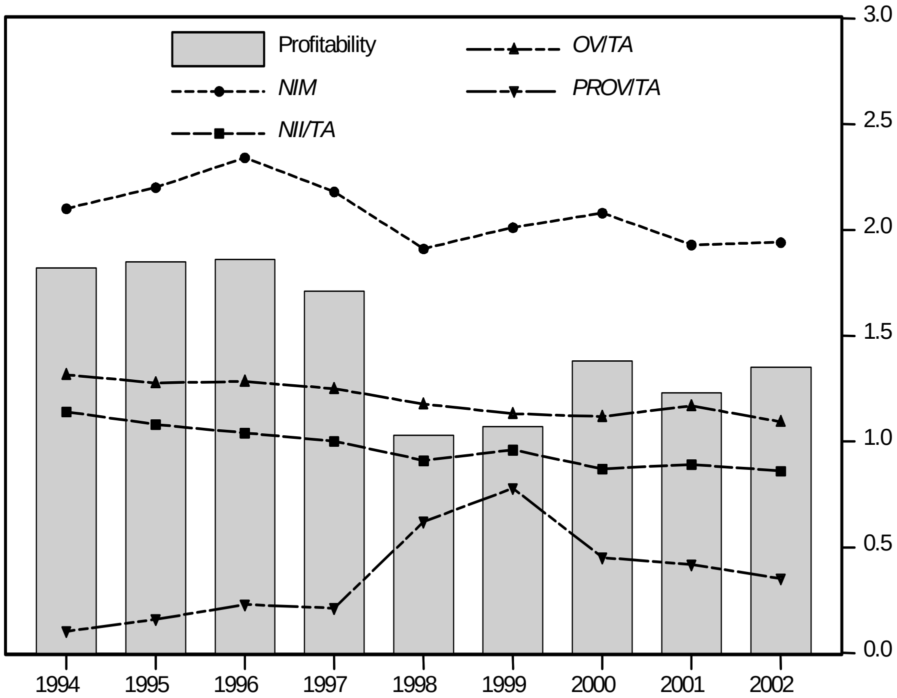
A. Market concentration and competition



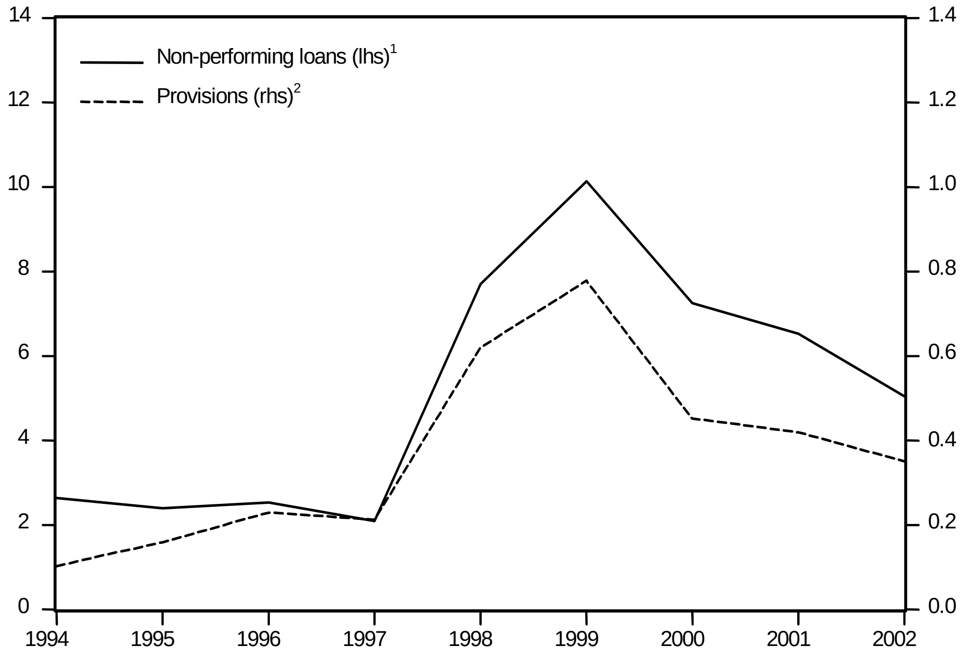
[1] In percent.

B. Decomposition of profitability

As a percentage of total assets



C. Provisions and non-performing loans



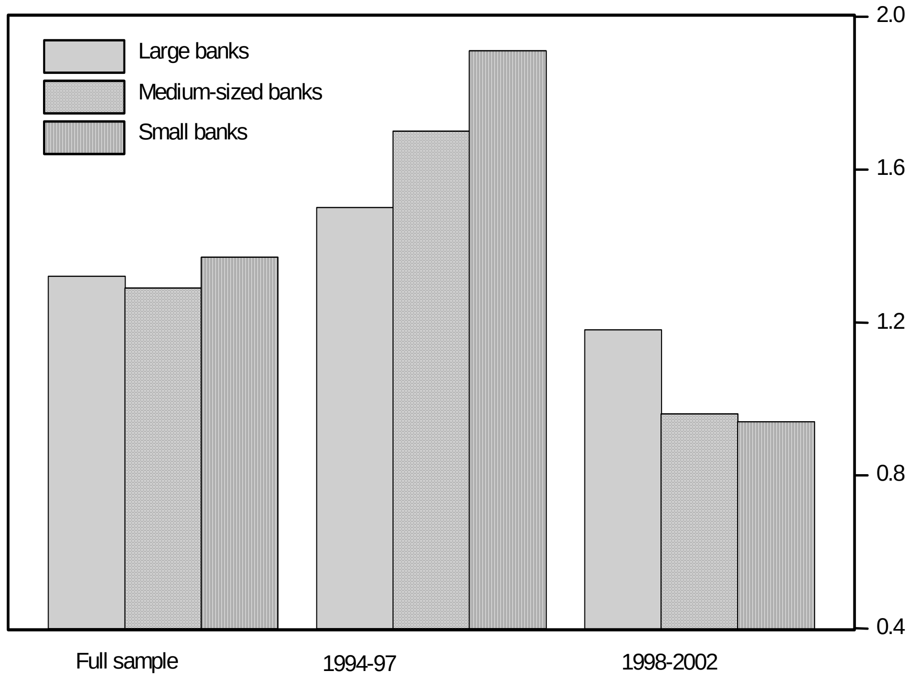
[1] As a percentage of total loans. [2] As a percentage of total assets.

Graph 3

Profitability, NIM, NPLs and bank size

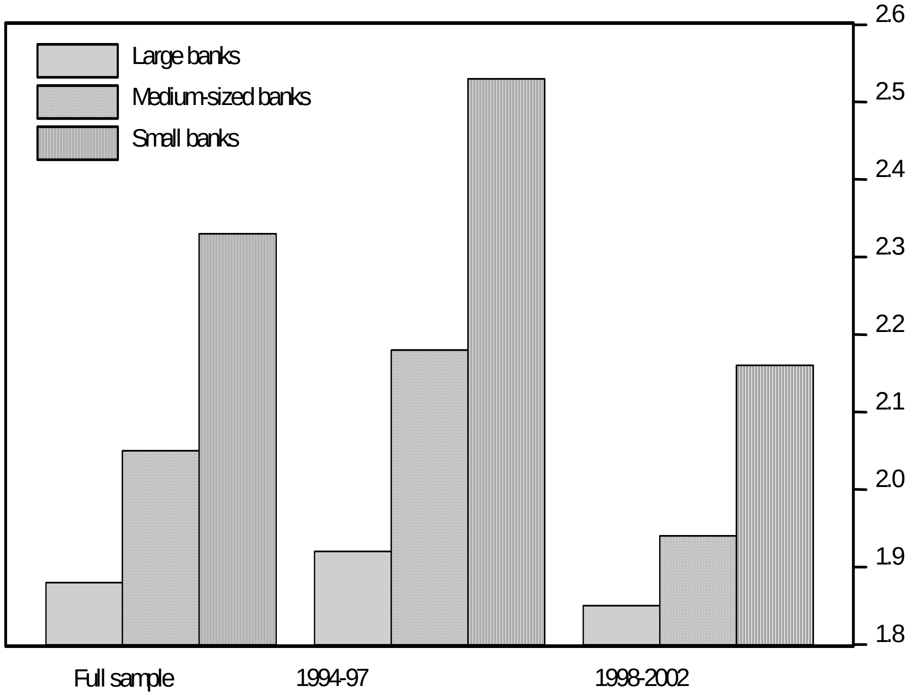
A. Profitability

In per cent



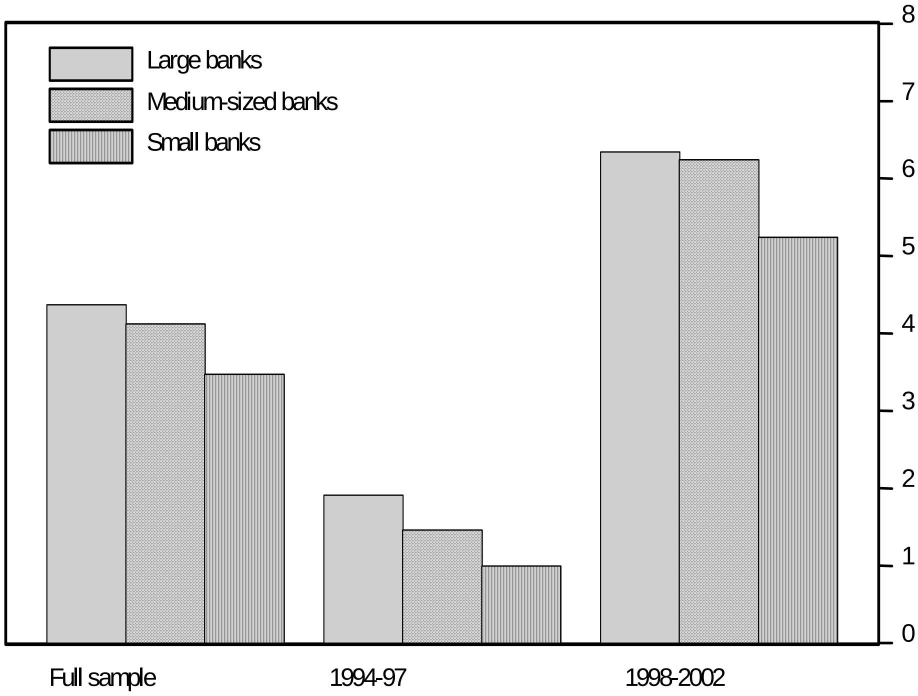
B. Net interest margin

In per cent



C. Non-performing loans

In per cent



D. Non-interest income net of operating costs

In per cent

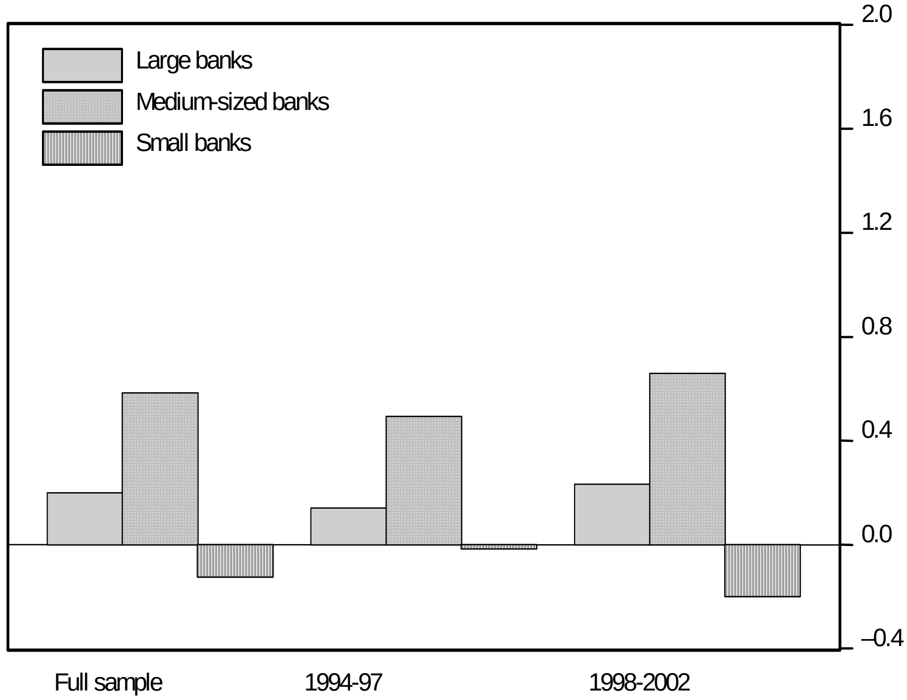


Table 1

Retail banks' business structure (1994-2002)

	All	**Large**	**Medium-sized**	**Small**
Asset portfolio				
Equity capital/total assets	13	4	10	18
Loans/total assets	46	40	49	46
Deposits/total assets	66	73	67	64
Income and expense				
Operating expenses/total expenses	25	25	21	29
Interest income/total income	87	85	88	88
Provisions/total loans	2	2	3	2
Lending portfolio				
Property loans/total loans	50	51	51	48
Consumer lending/total loans	7	7	6	9

Table 2

Dispersion of profitability, NIMs and NPLs

Bank-specific variables	Profitability				Net interest margin				Non-performing loans			
	Mean	**Std dev**	**Max**	**Min**	**Mean**	**Std dev**	**Max**	**Min**	**Mean**	**Std dev**	**Max**	**Min**
1994	1.9	0.7	3.9	0.5	2.3	0.8	4.6	0.6	1.1	1.1	5.5	0.1
1995	1.9	0.7	3.6	0.6	2.4	0.8	4.3	0.6	1.1	1.0	4.4	0.1
1996	1.8	0.7	3.4	0.3	2.4	0.8	4.8	0.5	1.5	1.2	4.2	0.0
1997	1.8	0.7	3.0	–0.1	2.4	0.8	4.2	0.7	1.3	0.9	3.9	0.1
1998	1.1	0.9	2.7	–0.8	2.2	0.7	3.4	0.7	6.2	5.5	29.0	0.9
1999	0.9	1.0	2.7	–2.8	2.2	0.7	3.4	0.3	8.7	6.6	31.3	0.5
2000	1.4	1.1	6.0	0.2	2.3	1.0	6.8	1.0	6.6	4.1	16.3	0.5
2001	0.9	0.9	2.1	–2.5	2.0	0.5	3.3	0.6	4.7	2.9	12.4	0.6
2002	0.9	0.7	2.3	–1.1	1.8	0.5	3.1	0.8	3.5	1.8	8.5	0.7

Table 3

Determinants of NPLs

Sample period: 1995-2002

	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**
NPL_{t-1}	0.33 *** (5.34)	0.36 *** (5.89)	0.36 *** (5.90)	0.33 *** (5.71)	0.34 *** (5.88)
GDP_t	–0.15 *** (–4.49)	–0.15 *** (–4.84)	0.59 (1.54)	0.81 *** (3.24)	0.82 *** (3.14)
$(GDP*SIZE)_t$	– –	– –	–0.04 ** (–1.90)	–0.05 *** (–4.02)	–0.05 *** (–3.37)
INF_t	–0.32 *** (–8.19)	–0.30 *** (–8.52)	–0.82 ** (–2.23)	–0.50 ** (–2.03)	–0.32 *** (–9.61)
$(INF*SIZE)_t$	– –	– –	0.03 (1.44)	0.01 (0.73)	– –
$PROP_t$	–0.03 *** (–3.55)	–0.02 *** (–3.50)	0.03 (0.27)	–0.13 *** (–4.24)	–0.12 *** (–4.28)
$(PROP*SIZE)_t$	– –	– –	0.00 (–0.54)	– –	– –
$(PROP*PROP\ SHARE)_t$	– –	– –	– –	0.19 *** (3.39)	0.20 *** (3.48)
$HIBOR_t$	0.55 *** (10.91)	0.57 *** (11.44)	1.14 * (1.69)	0.83 (1.44)	0.58 *** (13.22)
$(HIBOR*SIZE)_t$	– –	– –	–0.03 (–0.85)	–0.01 (–0.44)	– –
$PROP\ SHARE_t$	0.00 (0.01)	– –	– –	– –	– –
$CONS\ SHARE_t$	–0.06 (–1.61)	– –	– –	– –	– –
Adjusted R^2	0.88	0.89	0.91	0.94	0.92
Number of banks	27	27	27	27	27
Number of observations	209	209	209	209	209
Test for common intercept	1.92	2.08	2.13	2.23	2.37
p-value	[0.01]	[0.00]	[0.00]	[0.00]	[0.00]

Note: t-values are in ( ), p-values in [ ]. *, ** and *** indicate that variables are significant at 10%, 5% and 1% levels respectively.

Table 4

Determinants of the net interest margin

Sample period: 1995-2002

	Model 1	Model 2	Model 3	Model 4
NIM_{t-1}	0.39 ***	0.38 ***	0.30 ***	0.27 ***
	(6.32)	(6.36)	(5.38)	(4.62)
GDP_t	0.02 ***	0.02 ***	0.14 ***	0.12 ***
	(4.68)	(6.10)	(3.13)	(2.78)
$(GDP{*}SIZE)_t$	–	–	–0.10 ***	–0.01 **
	–	–	(–2.71)	(–2.38)
INF_t	0.01 **	0.01 **	0.15 ***	0.15 ***
	(2.41)	(2.93)	(3.60)	(3.69)
$(INF{*}SIZE)_t$	–	–	–0.01 ***	–0.01 ***
	–	–	(–3.53)	(–3.59)
$PROP_t$	0.00	–	–	–
	(–0.80)	–	–	–
$HIBOR_t$	0.01	0.04 ***	0.73 ***	0.59 ***
	(1.22)	(4.69)	(7.12)	(5.02)
$(HIBOR{*}SIZE)_t$	–	–	–0.04 ***	–0.03 ***
	–	–	(–6.99)	(–5.13)
$(HIBOR{*}EQUITY)_t$	–	–	–	0.18 **
	–	–	–	(2.42)
NII_t	–0.01 **	–	–	–
	(–2.02)	–	–	–
$NIEXPENSE_t$	0.60 ***	0.56 ***	0.58 ***	0.55 ***
	(7.65)	(7.41)	(8.29)	(7.80)
$PROP\ SHARE_t$	0.00	–	–	–
	(–0.27)	–	–	–
$CONS\ SHARE_t$	–0.00	–	–	–
	(–1.05)	–	–	–
Adjusted R^2	0.97	0.96	0.98	0.98
Number of banks	29	29	29	29
Number of observations	232	232	232	232
Test for fixed effects	4.61	4.31	5.06	5.20
p-value	[0.00]	[0.00]	[0.00]	[0.00]

Note: t-values are in ( ), p-values in [ ]. *, ** and *** indicate that variables are significant at 10%, 5% and 1% levels respectively.

List of variables

Dependent variables

NPL:	ratio of classified loans to total loans
NIM:	ratio of net interest income to total assets

Macroeconomic variables

GDP:	GDP growth
INF:	CPI inflation

Financial variables

PROP:	changes in property prices
HIBOR:	three-month Hibor

Bank variables

SIZE:	logarithm of asset size
EQU:	ratio of equity capital to total assets
PROVISION:	ratio of provisions to total assets
NII:	ratio of non-interest income to total assets
NIEXPENSE:	ratio of non-interest expenses to total assets
PROP SHARE:	ratio of property loans to total loans
CONS SHARE:	ratio of consumer loans to total loans

References

Ahn, Seung Chan and Peter Schmidt (1995): “Efficient estimation of dynamic panel data models”, *Journal of Econometrics*, 68, pp 5-27.

Anderson, T W and Cheng Hsiao (1981): “Estimation of dynamic models with error components”, *Journal of the American Statistical Association*, 76, pp 598-606.

Arellano, Manuel and Stephen Bond (1991): “Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations”, *Review of Economic Studies*, 58, pp 277-97.

Arellano, Manuel and Olympia Bover (1995): “Another look at the instrumental variables estimation of error-component models”, *Journal of Econometrics*, 68, pp 29-51.

Blundell, Richard and Stephen Bond (1998): “Initial conditions and moment restrictions in dynamic panel data models”, *Journal of Econometrics*, 87, pp 115-43.

Demirgüç-Kunt, Asli and Harry Huizinga (1999): “Determinants of commercial bank interest margins and profitability: some international evidence”, *World Bank Economic Review*, 13(2), pp 379-408.

——— (2000): “Financial structure and bank profitability”, *World Bank Working Paper*, 2430.

Gerlach, Stefan and Wensheng Peng (2003): “Bank lending and property prices in Hong Kong”, *HKIMR Working Paper*, 12/2003.

Hong Kong Monetary Authority (2002): *Annual Report 2002*.

International Monetary Fund (1999): *People’s Republic of China - Hong Kong Special Administrative Region*.

Jiang, Guorong, Nancy Tang, Eve Law and Angela Sze (2003): “Determinants of bank profitability in Hong Kong”, *Hong Kong Monetary Authority Research Memorandum*, http://www.info.gov.hk/hkma/eng/research/RM_on_Bank_Profitability.pdf.

Judson, Ruth A and Ann L Owen (1999): “Estimating dynamic panel data models: a guide for macroeconomists”, *Economics Letters*, 65, pp 9-15.

Kwan, Simon (2002): “The X-efficiency of commercial banks in Hong Kong”, *HKIMR Working Paper*, 12/2002.

Peng, Wensheng, Kitty Lai, Frank Leung and Chang Shu (2003): “The impact of interest rate shocks on the performance of the banking sector”, *HKMA Quarterly Bulletin*, 35, pp 20-7.

Shu, Chang (2002): “The impact of macroeconomic environment on the asset quality of Hong Kong’s banking sector”, *Hong Kong Monetary Authority Research Memorandum*, http://www.info.gov.hk/hkma/eng/research/RM20-2002.pdf.

Previous volumes in this series

No	Title	Issue date
BIS Papers No 21	Real estate indicators and financial stability	April 2005
BIS Papers No 20	Fiscal issues and central banking in emerging economies	October 2003
BIS Papers No 19	Monetary policy in a changing environment	October 2003
BIS Papers No 18	Monetary stability, financial stability and the business cycle: five views	September 2003
BIS Papers No 17	Regional currency areas and the use of foreign currencies	Revised September 2003 May 2003
BIS Papers No 16	Guide to the international banking statistics	April 2003
BIS Papers No 15	China's capital account liberalisation: international perspectives	April 2003
BIS Papers No 14	Guide to the international financial statistics	February 2003
BIS Papers No 13	Comparison of creditor and debtor data on short-term external debt	December 2002
BIS Papers No 12	Market functioning and central bank policy (2001 Autumn Economists' Meeting)	June 2002
BIS Papers No 11	The development of bond markets in emerging economies (2001 Deputy Governors' Meeting)	June 2002
BIS Papers No 10	Central bank intervention and market expectations	April 2002
BIS Papers No 9	Comparing monetary policy operating procedures across the United States, Japan and the euro area	December 2001
BIS Papers No 8	Modelling aspects of the inflation process and the monetary transmission mechanism in emerging market countries	November 2001
BIS Papers No 7	Electronic finance: a new perspective and challenges	November 2001
BIS Papers No 6	The financial crisis in Japan during the 1990s: how the Bank of Japan responded and the lessons learnt	October 2001
BIS Papers No 5	The changing shape of fixed income markets: a collection of studies by central bank economists	October 2001
BIS Papers No 4	The banking industry in the emerging market economies: competition, consolidation and systemic stability (2000 Deputy Governors' Meeting)	August 2001
BIS Papers No 3	Empirical studies of structural changes and inflation	August 2001

All volumes are available on request from the BIS and are also on our website www.bis.org.