Committee on the Global Financial System

Risk measurement and systemic risk

Proceedings of the Third Joint Central Bank Research Conference

October 2002

BANK FOR INTERNATIONAL SETTLEMENTS

This volume contains papers presented and papers based on presentations at the Third Joint Central Bank Research Conference on Risk Measurement and Systemic Risk held at the BIS in March 2002. The views expressed in this volume are those of the authors and do not necessarily reflect the views of the BIS or the central banks represented at the conference. Authors retain the copyright for their individual papers.

Copies of publications are available from:

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).

ISBN 92-9197-638-5

Preface

The Third Joint Central Bank Research Conference on Risk Measurement and Systemic Risk took place at the Bank for International Settlements (BIS) in Basel on 7 and 8 March 2002. The conference was organised by the BIS on behalf of the Committee on the Global Financial System (CGFS),[1] in cooperation with the Bank of Japan, the Federal Reserve Board and the European Central Bank. The two earlier conferences were hosted by the Federal Reserve Board and the Bank of Japan in 1995 and 1998, respectively.

Staff from the Bank of Japan (Naohiko Baba and Tokiko Shimizu), the Federal Reserve Board (Michael Gibson and Matthew Pritsker), the European Central Bank (Philipp Hartmann and Jukka Vesala) and the BIS (Ingo Fender and Allen Frankel) were the principal organisers of the conference. With regard to administrative matters, crucial contributions to the successful organisation of the event were made by Beate Diemer, Siegfried Eger, Hermann Greve, Thomas Lejeune, Cynthia Lloyd and Bridget Thomas. Ingo Fender and Jacob Gyntelberg edited the present volume and staff from the BIS's Information and Publication Services and Language Services helped to prepare it for publication.

This volume contains papers that either were presented or interpret presentations at the conference. Authors retain their copyright. The following chapter summarising the conference was authored by Ingo Fender.

One of the main goals of the conference was to bring together the business, research and policy communities to foster active exchange on issues related to risk measurement and systemic risk. It was against this background that the conference organisers gathered a group of attendees from the risk measurement-minded quarters of each of these three communities. The organisers wish to express their appreciation to all those who agreed to attend the conference, be it as paper presenters, session chairs, discussants or participants in the open discussion. The conference's 18 papers, grouped in six sessions, were selected from more than 130 submissions. To foster interaction and to give the discussion of the conference papers a practical perspective, session chairs were drawn from the central bank community, while industry representatives were asked to serve as discussants.

While being somewhat unusual, this arrangement seems to have worked rather well in terms of promoting exchange of ideas. Authors, that is academics and central bank researchers, had the opportunity to present their research to a relatively senior audience of policymakers and risk management professionals. In turn, these practitioners offered their views on various issues of practical relevance, providing a valuable angle on current findings and possible guidance for future research. While the organisers of future conferences might like to set aside more time for open-floor discussions, it seems that a worthwhile tradition has now been established to further research on the important topic of risk measurement and systemic risk through interaction at Joint Central Bank Research Conferences.

---

[1] The Committee on the Global Financial System (CGFS) is a central bank committee established by the Governors of the G10 central banks. It monitors and examines broad issues relating to financial markets and systems, with a view to elaborating appropriate policy recommendations to support the central banks in the fulfilment of their monetary and financial stability responsibilities. In carrying out its tasks, the Committee places particular emphasis on assisting the Governors in recognising, analysing and responding to threats to the stability of financial markets and the global financial system. The CGFS is chaired by Yutaka Yamaguchi, Deputy Governor of the Bank of Japan.

Table of contents

Risk measurement and systemic risk: a summary

1. Overview

Research on risk measurement and systemic risk-related issues, the focus of the conference, has progressed substantially since 1995, when the first in this series of conferences was held. At the first conference, centre stage was taken by the value-at-risk (VaR) methodology, which was then gaining ground in academia and at leading financial institutions. Some papers explored how risk could be quantitatively measured and what the meaning of such measures would be. Shortly thereafter, in 1997, the Asian crisis erupted, triggered by and itself triggering events that were beyond the bounds envisioned by standard VaR methodology. As a result, discussions at the second conference in 1998 very much focused on the implications of the Asian crisis for risk measurement methodologies as well as market microstructure theory's lessons for market dynamics in times of stress.

In his opening remarks, Andrew Crockett explained the rationale for the focus of this third conference and its emphasis on questions relating to the nature and sources of market liquidity, recent advances in risk measurement methods, sources of banking crises and contagion effects across regions and markets. As for the first two conferences in the series, the goal was to foster the exchange between the policy and research communities. To this end, the co-organisers brought together a broad mix of attendees: academics, public sector officials and industry professionals as well as central bank staff. Overall, the conference generated a set of interesting discussions which sought to both assess and further the current state of knowledge on issues related to risk measurement and systemic risk and to identify areas of policy interest and for future research. These discussions focused on three broad topics, which are summarised below under three headings.

2. Systemic banking crises, contagion and monitoring

The series of banking and currency crises that emerged in various parts of the world during the past two decades or so suggests that financial stability is not to be taken for granted. In view of this, the conference organisers encouraged submission of research concerned, among other things, with the sources of financial market instabilities and, by extension, ways to avoid financial crises. Much of this literature has focused on issues of banking stability and the notion of "systemic risk", ie the danger that problems in a single financial institution might spread and, in extreme situations, such contagion could disrupt the normal functioning of the entire financial system.

Banking stability and systemic crises

Diamond and Dybvig,[1] in their seminal paper, present a theory of banking based on liquidity risk sharing, with banks emerging as providers of the required liquidity insurance. They show how, under asymmetric information, bank runs can emerge in such a fractional reserve banking system. However, while allowing for the possibility of bank runs, the Diamond/Dybvig (DD) model is not able to explain the causes of banking crises: bank runs, in their world, are essentially self-fulfilling prophecies or "sunspot" events.

Extensions of the DD model, as surveyed by Allen and Gale's contribution to this proceedings volume, have therefore introduced uncertainty about asset returns to proxy for the impact of the business cycle on the valuation of bank assets. In these models with aggregate shocks to asset returns, financial crises are driven by fundamentals. Shocks to asset returns, by reducing the value of bank assets, raise the possibility of banks being unable to service their commitments. Depositors, anticipating such difficulty, will tend to withdraw their funds early, possibly precipitating a crisis.

[1] D Diamond and P Dybvig, "Bank runs, deposit insurance, and liquidity", *Journal of Political Economy* 91, 1983, pp 401-19.

Despite its widespread use in theoretically analysing financial instability, the DD model and its various extensions do not provide a completely plausible description of actual patterns of banking crises. Runs by depositors are rare. Therefore, banking crises have more typically started when the interbank supply of credit was sharply cut or withdrawn. In addition, a purely bank-centric approach to systemic risk may no longer be appropriate, given that financial markets tend to play a significant role as propagation channels for disturbances involving the banking system and the real economy. This is why Yutaka Yamaguchi, in his luncheon address, set out the need for any comprehensive analysis of systemic risk to go beyond the narrow confines of the banking system, to cover the interrelations between the banking system, financial markets and the real economy. Indeed, one of the recurring themes of the conference was that much of the literature on banking crises and contagion, the topics of the first two conference sessions, remained overly focused on a set of specific assumptions and modelling conventions. As a result, while being more tractable, these models have provided only limited analytical assistance to the policy community.

In the latest version of their 1998 model,[2] the main focus of the first presentation at the joint research conference, Allen and Gale introduce a market for long-term assets into the analysis, enabling banks to liquidate these assets. Contrary to the original DD model, liquidation costs are therefore endogenous. As a result, asset markets provide a transmission mechanism that serves to channel the effect from the liquidation of assets by some banks to other banks in the economy. If a sufficient number of banks are forced to liquidate their assets and the demand for liquidity rises above a certain level, asset prices will move sharply. This may, in turn, force other banks into insolvency and exacerbate the original crisis. As a result, the model, compared with earlier theories, provides a more realistic explanation of how and why financial crises may develop. It also highlights the importance of asset market liquidity for the evolution and, eventually, the avoidance of financial crises.

Carletti et al, in their presentation, tackled another major shortcoming of many analyses based on the traditional Diamond/Dybvig approach: the failure to recognise the role of interbank credit. In their model, banks compete in the loan market, while the interbank market serves as an insurance mechanism against deposit withdrawals due to liquidity shocks. This setup enables the authors to investigate the influence of bank mergers on reserve holdings and the interbank market and, ultimately, aggregate liquidity risk. Mergers affect bank balance sheets via increased concentration and potentially enhanced cost efficiency, while also altering the structure of liquidity shocks. The model highlights the importance of functioning interbank markets for financial stability and sheds some light on potential trade-offs between antitrust and supervisory policies. In the discussion, some conference participants commented on the practical relevance of the model. In particular, it was noted that nowadays central banks were usually ready to provide liquidity elastically to accommodate temporary fluctuations in liquidity. Given this willingness, it was argued, bank liquidity crises would be of limited importance. However, it was felt that the paper generated important insights into how mergers might affect liquidity in the money market and, by extension, how this would influence the execution of monetary policy operations.

The final presentation of the first conference session, which is summarised in Giannetti's contribution to this volume, shifted the focus to the emerging markets. Specifically, she argued that underdeveloped financial markets, characterised by a lack of transparency, and easy access to foreign capital can help to explain overlending and crisis phenomena in emerging financial markets. According to Giannetti, overlending due to investor moral hazard, that is the existence of explicit or implicit guarantees, is merely a special case of a broader crisis model. In her model, based on incomplete investor information on the average quality of investment opportunities and the existence of soft budget constraints due to capital inflows, bank-financed investors will rationally not require a risk premium until losses become substantial, even without guarantees on deposits. Based on this insight, the paper suggests that well developed capital markets, by increasing the number of creditors, can eliminate excessive reliance on bank-firm relationships and soft budget constraints, which will reduce the probability of financial crises. This, in turn, lends support to the often advocated "sequencing" policy prescription, demanding that countries should have appropriate financial structures in place before removing capital controls and passively accommodating foreign investors.

---

[2] F Allen and D Gale, "Optimal financial crises", *Journal of Finance* 53, 1998, pp 1245-84.

Contagion

During the second session, the conference's focus moved on to contagion across markets and countries, an issue which, despite its importance for financial market stability, remains less than completely understood. Contagion is at the heart of any analysis of financial crises, because it is contagion that makes the initial shock a truly systemic event. Therefore, echoing Tommaso Padoa-Schioppa's luncheon speech, to understand financial sector risks, one has to deal with the origin of these risks as well as the channels of propagation. Padoa-Schioppa noted that the increasing use of complex risk transfer instruments and speed of financial market transactions add to the complexity and rapidity of the potential propagation of shocks, making these risks difficult to gauge. Based on these considerations, contagion can be viewed as the propagation mechanism that causes small idiosyncratic or systematic shocks to have systemic consequences.[3]

The session started with two empirical presentations, by Kaminsky and Reinhart and Dungey et al, investigating contagion by using stock market and bond market data, respectively. Kaminsky and Reinhart's presentation involved an analysis of daily stock market behaviour for a number of emerging and mature markets. Specifically, their paper looks at empirical return distributions in different countries and regions, conditional on extreme returns in financial centres or emerging markets, to identify where shocks originate and how they spread through the system. By comparing these distributions, the authors discover that the distribution of returns around the globe changes only in those periods that are characterised by turmoil in large financial centres (notably the United States, Germany and Japan). While shocks might spill over regionally, via trade links, centres have to be affected for financial turmoil to be become a global phenomenon. That is, shocks to the periphery seem to spread to other peripheral areas via their impact on financial centres. A shock that never reaches a centre is likely not to become a systemic event.

Mardi Dungey and her co-authors employ a somewhat different approach. They identify contagion by looking at daily movements in bond spreads for the LTCM crisis period in an effort to quantify the effects of unanticipated regional shocks across borders. The resulting contagion measure controls for common global shocks, country specific shocks and regional factors. The authors find contagion originating from the Russian default, with the measured level of the effect larger for emerging economies. However, the proportion of total volatility attributable to contagion varies widely across countries and is not always more substantial for developing countries. Thus, while contagion tends to be viewed as mainly a concern for developing countries, the evidence from the Russian and LTCM crises suggests this is not necessarily the case. In fact, contagion effects are found to be widely distributed across both developed and developing markets, making contagion a phenomenon reserved not only for developing countries.

In the discussion of the two empirical contagion papers, however, doubts were expressed about whether the data and methodologies used in these and similar empirical models were always suitable for identifying the effects of contagion. It was pointed out, for example, that Kaminsky and Reinhart's decision to look at daily stock price returns of emerging economies to establish how turmoil in an emerging market spills over to other markets was open to criticism. In particular, it was observed that the definition of what a crisis is and when it started might change when equity prices rather than exchange rate data are used. For example, during the Asian crisis, days of crisis in stock and foreign exchange markets tended to differ, as stock markets in Asia welcomed the initial depreciation of local exchange rates as a necessary adjustment. As a result, turmoil in their paper might be very different from what is commonly perceived as a crisis, limiting the value and comparability of their findings. In addition, doubts were voiced as to whether the paper actually addressed the issue of contagion, given its focus on patterns of spillovers in stock markets. In particular, conference participants suggested that future work might look more closely at causalities by trying to infer the direction of spillovers. Finally, it was proposed to apply the two papers' methodologies to recent cases of limited or non-contagion. This was seen as potentially useful in testing the hypothesis that, during recent episodes, investors have been more discriminating in their reactions than in the past.

In the third presentation of the contagion session, Cipriani and Guarino elaborated on social learning and informational herding as a source of financial crises. While in the previous presentations, such as

[3] See O DeBandt and P Hartmann, "What is systemic risk today?", in *Bank of Japan, Risk Measurement and Systemic Risk: Proceedings of the Second Joint Central Bank Research Conference*, Tokyo, 1999, pp 37-84.

the one by Allen and Gale, financial crises and contagion were essentially based on developments in fundamentals or sunspot phenomena, Cipriani and Guarino advance reasons for crises in the absence of sunspots and despite sound fundamentals. Essentially, the authors introduce the possibility that crisis phenomena might reflect a learning process between traders, independent of any change in fundamentals. By doing so, the approach provides a possible underpinning for the centre-periphery results found by Kaminsky and Reinhart, as increased effects on peripheral markets could now be interpreted as the result of a higher trading frequency at the centre. The authors also implicitly challenge insights from other areas of research, such as second-generation speculative attack models.

According to Cipriani and Guarino, a possible explanation of why sound fundamentals may not be reflected in asset prices is that information about these fundamentals may be spread among investors, with prices failing to fully aggregate it. In particular, this would happen if investors, instead of acting according to their own private information, simply decided to follow the actions of previous traders, a phenomenon known as informational herding. Specifically, the authors use an information cascades model with flexible prices to show that sequential trading under incomplete information can lead to a permanent deviation of prices from fundamentals. In such a model, prices may fail to aggregate private information and may, due to asymmetric information, lead to all traders taking the same action. Under specific conditions, traders choose to essentially disregard their own private information, not allowing asset prices to reflect fundamental values. Furthermore, in a multi-market sequential trading framework, it can be shown that sell orders in one market can affect the price path of another market, making its price settle at lower value. While such informational spillovers are to be expected, due to correlation between fundamentals, sequential trading can explain contagion across markets as correlation between the prices of two assets can be higher than correlation between fundamentals. Informational herding effects can, therefore, spill over from one asset market to the other, providing a potential explanation for contagion across markets.

Conference participants, however, raised doubts about the validity of the two core assumptions behind the model - the existence of “gains from trade” and the restriction that trades occur only sequentially. In particular, it was noted that gains from trade implied that market participants would be willing to trade at a loss. With regard to future research, it was suggested that empirical implementations of the cascades approach could shed some light on contagion effects, eg during the Asian crisis. For this to be possible, however, cascade models would have to be reworked to generate verifiable theoretical predictions on, for example, conditions under which informational cascades were more likely to occur.

Systemic monitoring

Systemic events can impose substantial social costs on the affected economies, as bank runs, for example, will disrupt credit relations and allocative efficiency, in turn leading to non-trivial direct and indirect effects on economic performance in the form of output losses. Practical aspects of systemic monitoring and the analysis of systemic risks are, therefore, high on the policy agendas of central banks and other members of the regulatory and supervisory community. For this reason, the last conference session featured two papers that added a practical angle to the discussion on banking crises and contagion, by showing how financial market and banking data can be used to monitor the fragility of real-world banking sectors.

Against this background, Gropp et al explore how market-based indicators can be usefully employed to predict banking fragility by adding to the information gained from more traditional, balance sheet-based indicators. To this end, the authors analyse the indicator characteristics of Merton-type distances-to-default and subordinated bond spreads in signalling material weakening of banks’ financial conditions. They demonstrate that useful and well-behaved indicators can be derived from stock market data, while, so far, the focus has been much on subordinated debt spreads. They also find that these market-based indicators, with different leads, are useful in predicting banking fragility and that they even add information relative to more traditional indicators based on balance sheet information. The authors thus suggest the use of market-based indicators in supervisors’ early warning models, a potentially promising future enhancement of supervisors’ ongoing monitoring efforts.

Blåvarg and Nimander, in their paper, give valuable insights into the Riksbank’s monitoring of systemic risk in the Swedish banking system. In particular, to monitor counterparty exposures in the domestic interbank market, the Riksbank uses data detailing the largest uncollateralised exposures of the four major players in the Swedish banking system. The approach involves exposing a proxy for the Swedish banking system, ie the four biggest banks, to solvency shocks originating from outside the interbank market and assessing how the system is affected via interbank exposures. The authors find that domestic direct contagion effects are less than what might have been expected in the Swedish banking system, given its degree of concentration. In most cases where one of the large banks is assumed to fail, other banks are found not to suffer direct losses that would reduce their capital ratio significantly below the regulatory level. Similar results are found for the risk of direct contagion from abroad, which mainly arises from foreign exchange settlement exposures. Conference participants suggested that the approach presented might be extended to explicitly take into account correlated shocks due to common exposures. This, and coverage of possible second-round effects of a given primary shock, were avenues suggested for future research.

3. Market liquidity

As argued above, much of what was discussed during the first two sessions revolved around the concept of market liquidity and its relevance for financial stability. Banks, the epicentre of instability in the models surveyed by Allen and Gale, are providers of insurance for liquidity risk. They serve this function by following a liquidity immunisation strategy, implemented via individual asset markets and interbank credit markets, to guard themselves against the possible effects of forced asset liquidation. This, in turn, explains the organisers’ motivation for placing particular emphasis on papers seeking to explicate the specifics of liquidity provision in various microstructural seetings and across various asset markets.

Against this background, the papers presented in the two liquidity sessions all addressed issues of liquidity provision and, in various ways, all supported the view that market liquidity can affect market performance, while, in turn, being affected by market microstructure. In the first liquidity session, the paper by Cohen and Shin explored the short-run variability of US Treasury note prices using order flow data from the US Treasury market. The paper by Tien investigated the determination of exchange rates using currency futures data disaggregated by type of trader. Finally, Pritsker employed a theoretical asset pricing model to demonstrate the possibility that the asset holdings of large investors might matter for asset price determination. In the second liquidity session, Danielsson and Payne examined the microstructural specifics of liquidity provision on an electronic foreign exchange trading platform, while Harrison and Wong and Fung looked into the microstructure of the primary corporate bond and the equity markets, respectively.

Positive feedback in the Treasury market

Cohen and Shin explore the empirical relevance of strategic interaction among market participants. In particular, they are interested in whether the distributions of returns/liquidation values are more dispersed than they would be if risks were truly exogenous. A direct implication of such a finding would be that individual market participants are likely to underestimate potential price movements resulting from shocks to markets and, therefore, predictably underestimate the riskiness of their own exposures. The empirical part of the paper investigates return and order dynamics in the US Treasury bond markets to find that signed order flow has a strong impact on prices. While this is fully in line with what one would have expected based on standard market micostructure models, the authors also find that there is often also a strong effect in the other direction, ie prices affecting order flow. As this is found to be more likely in turbulent times, bond markets seem to behave in meaningfully different ways depending on market conditions. This effect in the price-order flow pattern, so the authors claim, may be attributed to constraints on traders' behaviour, such as those imposed by risk management systems or position limits. As these and similar constraints can give rise to “strategic complementarities”, the most basic concept of strategic interaction, the actions of individual traders may become mutually reinforcing, introducing feedback from prices to order flow. The specific issue of how VaR constraints might affect asset prices and volatilities, an interesting topic against the background of Cohen and Shin’s findings, was taken up again by Berkelaar et al and is covered below.

Market microstructure and FX market liquidity

Tien, in his paper, shifted the attention to the foreign exchange (FX) market and investigated FX premia based on hedge demand, where risk (forward) premia are driven by income shocks and risk-averse agents' attempts to hedge these shocks by trading foreign currency. The model is tested using data on hedging demand in currency futures markets and the author finds evidence indicating that FX risk premia based on hedge demand explain, on average, some 45% of the variation in currency returns at a monthly horizon. Therefore, risk premia are present and identifiable in the foreign exchange market and, more importantly, risk sharing can explain a significant proportion of the observed variation in exchange rates. This, in turn, suggests that the FX market is an efficient mechanism for allocating risk across the economy. These results may also help to explain the depth and liquidity of the major currency markets, since traders should be more willing to trade in situations where counterparties are not likely to be better informed. In the discussion of the paper, it was suggested that the definition of hedgers used in disaggregating the data by type of trader might be a problem that could potentially skew the results. It was separately noted that, by extension, Tien's findings also supported opposition to recent proposals for the introduction of Tobin taxes in the foreign exchange markets. This is because such a tax would interfere with the needs of those market participants seeking to hedge their income risk - an unnecessary burden from a viewpoint of allocative efficiency.

Daníelsson and Payne, in their presentation, remain in the realm of the FX market by empirically investigating liquidity provision on electronic FX broking systems. Such electronic trading platforms, having captured a sizeable market share in the inter-dealer FX market, have recently attracted considerable interest, particularly as they rely on electronic order books and, thus, on limit orders as the ultimate source of liquidity.[4] A deep limit order market would, therefore, be characterised as having a large volume of differently priced buy and sell limit orders outstanding, waiting to be "hit" by market orders arriving in the market. Such a market would thus be able to absorb large numbers of limit orders without significant price movements, while being able to restore the depth of the order book once a market order is executed.

In their latest paper, which was the basis of their presentation, Daníelsson and Payne use DEM/USD Reuters data for a particular week in October 1997 to look at the dynamics of market liquidity. In particular, the authors seek to establish the conditions driving liquidity supply and demand in the market. They find that market order activity has strong and persistent effects on subsequent limit order activity in electronic order books. In addition, they show that the order book is dynamically illiquid in the sense that, subsequent to market order arrival, further liquidity is removed from the other side of the order book as buy orders cause liquidity suppliers to reprice limit orders, leading to a reduction in sell side depth. The order book, therefore, "thins out" as liquidity suppliers seem to guard themselves against being picked off by traders with superior information, a finding in line with market microstructure models based on asymmetric information. In addition, the authors find depth to be negatively related to volatility and unexpected volume, while being positively related to expected volume. This, in turn, suggests that liquidity suppliers are risk-averse and concerned about informed trades by market order traders. On the other hand, as remarked in the subsequent discussion at the conference, such correlation between volatility and market depth could also be a reflection of liquidity providers not being anxious to enter a market where there is not a sufficient background level of volatility to justify their presence. Overall, this line of research was seen as an important contribution to the existing empirical market microstructure literature. Nevertheless, it was noted that there was much more research to be done before arriving at an informed understanding of liquidity generation in order-driven markets.

Market microstructure and stock market liquidity

The paper by Wong and Fung looks into the liquidity of equity markets, using a unique set of 30-second tick-by-tick data from the Hong Kong Stock Exchange. Various conventional liquidity

---

4 See Committee on the Global Financial System, *Structural aspects of market liquidity from a financial stability perspective: a discussion note prepared by the CGFS for the March 2001 meeting of the Financial Stability Forum (FSF)*, Bank for International Settlements, Basel, 2001.

indicators are constructed to evaluate how liquidity has evolved since the 1997 Asian crisis and to examine the determinants of changes in liquidity. The analysis shows that, having deteriorated during the Asian and Russian financial crises, market liquidity has broadly recovered to pre-crisis levels. In addition, to more fully gauge the dynamics of market liquidity, a GARCH model is developed for five selected stocks to relate the sensitivity of their price movements to net order flows. Based on this model, the authors establish that market liquidity deteriorated sharply during the crises, followed by an apparent recovery in the post-crisis period.

Overall, given the correlation of stock market liquidity with cost and risk factors established in the paper, the authors find their results to be consistent with market microstructure models based on inventory control, predicting that market depth is negatively correlated with price volatility. In such a model, as limit orders are essentially options that can be exercised by submitting a market order, heightened volatility would imply an increased risk for the limit order provider to deviate from his optimal inventory position - which would, in turn, lead to declining order book depth. However, as in Danielsson and Payne, it was noted that a negative correlation between depth and volatility could well be consistent also with microstructure models based on asymmetric information and risk-averse liquidity providers interacting with informed traders. Finally, interest was expressed in further studies to fully reconcile the various theories on how market microstructure might affect market prices with real-world market structures and transactions data from different markets.

Issue size and bond market liquidity

Issue size is known to be an important determinant of bond market liquidity and the issuer's funding cost. For this reason, an empirical study by Harrison investigated the issue size-liquidity linkage by looking into the impact of liquidity shocks on the composition of firms entering the corporate bond market. As much previous research on bond market liquidity has focused on secondary markets, examining the primary bond market provides additional insights into what issue and issuer characteristics may be fundamental liquidity factors. To this end, Harrison's approach focuses on the role of issue size and its sensitivity to illiquidity. That is, unlike other authors, he looks at the effect of market stress on liquidity, rather than the causes of market stress and illiquidity. Using multivariate regressions to control for observable issue and issuer characteristics, he finds that issue size, and certain measures of issuer familiarity, are priced liquidity factors. Primary markets, therefore, seem to recognise and price information problems and related factors of liquidity determination at issuance. In particular, the price depends crucially on whether the economy is experiencing an illiquidity shock. When liquidity is at a premium, larger bonds by well known issuers are much more prominent, squeezing issues by smaller, less known firms and the high-yield market in particular. Overall, it seems, with multiple issues and large issues being discounted, that the prospect of wider ownership translates into more trading and more liquidity for these securities.

In the discussion, there was agreement that, while the paper was more or less agnostic about what exactly explains the link between size and liquidity, it would be worthwhile to examine the issue further. In particular, it was felt that size might well proxy for some very specific factor not (yet) captured in the paper. In addition, it was suggested that the hedging of corporate bond inventories might influence the econometric results, if not properly controlled for. If inventory hedging becomes more expensive, dealers will become more reluctant to bring new issues to the market, making hedging a core factor behind the activity in the primary market. From a policy viewpoint, Harrison's findings were seen to suggest that, as conditions in the primary bond market tend to reflect conditions in the secondary market, those monitoring liquidity can also turn to the primary market to gauge liquidity developments. For example, as the composition of issuers tends to change rather dramatically in response to periods of illiquidity, it may be interesting to look at who is coming to the market instead of just looking at the overall amount issued.

Large investors and market liquidity

Pritsker, to address questions related to liquidity determination, constructs an imperfect competition model of asset pricing without focusing on a particular market. A key innovation of his approach, when compared to the studies surveyed in his contribution to this volume, is the assumption that institutional investors incorporate the price impact of their actions into their own decision-making. Imperfect competition and the existence of agents with differently sized endowments, ie asset holdings, imply that large agents face costs, due to illiquidity, when trying to rebalance portfolios. As a result, large traders will be hesitant to trade away from their endowments. That is, in response to acquiring an appreciation of the possible consequences of their actions on market outcomes, large traders will tend to sell less of their endowments when subject to liquidity shocks. In turn, observed market returns on assets will be directly related to the size of large traders' endowments. The model thus endogenously generates trading costs and explores their implications for asset pricing and market liquidity. Therefore, this line of research is likely to be useful in further exploring issues such as why some shocks are more contagious than others, or why some assets are more liquid than others.

One conjecture that would follow from this analysis, as remarked in Peter Praet's discussion of the Pritsker paper, is that large traders will not only sell less of their endowments but will also be biased towards holding liquid, blue-chip stocks in their portfolios. It was also noted that, given that Pritsker suggests that large market players may have an incentive to hide their asset endowments, extending his model to an environment with asymmetric information concerning investors' holdings could yield interesting new results. Such an approach could, for example, be used to compare the "full disclosure" case adopted in Pritsker's model with situations of zero disclosure - an interesting undertaking from a policy perspective.

4. Practical risk measurement and management

Over the course of the conference, several participants noted that, with regard to practical risk measurement, substantial progress had been made since the first conference in 1995. There was agreement, however, that further improvements were necessary in terms of modelling the tails of return distributions, improving the treatment of liquidity risk, and integrating the measurement of market and counterparty credit risk. The performance of risk measurement systems in times of stress and possible shortcomings of conventional methods in dealing with such situations received particular attention throughout the discussions. Against this background, conference participants commented on the need for the use of other techniques, such as stress testing, to address the shortcomings of the more traditional risk measures, an issue that also received the attention of two recent CGFS reports.[5]

The papers in the fifth, the technical, conference session applied cutting-edge statistical techniques to specific issues of financial risk measurement. One paper, by Diebold et al, showed how high-frequency data can be employed to construct volatility forecasts which, in turn, can be used as an input for firms' risk measurement. The authors integrate high-frequency intraday FX data into the measurement and modelling of daily and lower-frequency volatility and return distributions, overcoming the problems of more restrictive, traditional approaches in terms of dealing with intraday frequencies. The relevance of the study stems from the fact that volatility forecasting is a prominent feature of many practical financial decisions such as asset allocation, market timing and derivatives pricing.

The second paper, by Yamai and Yoshiba, compared two popular summary measures of financial risk, value-at-risk (VaR) and expected shortfall, using extreme value theory (EVT). The authors use simulated asset returns with extreme correlations and fat-tailed distributions to compare the performance of the two measures under market stress and evaluate whether the measures take account of extreme losses in the tail of the underlying distributions (tail risk) and whether they can be accurately measured using limited data (estimation error). In the open discussion, while agreeing with some of the advantages of expected shortfall measures under conditions of stress, some participants raised doubts as to whether expected shortfall could be a practical measure to be actually used by banks. In particular, it was noted that, while VaR had very good statistical properties, not much was known about the distribution of expected shortfall and that using expected shortfall for backtesting might pose problems.

Lucas et al, in their presentation, used EVT to describe how the tail of the loss distribution in portfolio credit risk models depends on modelling assumptions and parameter choices. While tail index and

---

[5] See Committee on the Global Financial System, *Stress testing by large financial institutions: current practice and aggregation issues*, Basel, 2000, and *A survey of stress tests and current practice at major financial institutions*, Basel, 2001.

quantile estimators, like VaR, are now commonly used to assess the tails of return distributions, application of these statistical techniques in calculating extreme credit loss quantiles is less common. In their paper, the authors investigate whether the application of extreme value theory to the tails of portfolio credit losses generates EVT quantiles that are accurate enough to be useful for credit risk managers. To this end, alternative tail approximations are considered for two special cases of a generalised model for portfolio credit losses. The results suggest that one has to be careful in applying EVT for computing extreme quantiles efficiently. The applicability of EVT in characterising the tail shape appears to depend crucially on the exact distributional assumptions for the systematic and idiosyncratic credit risk factors. These factors are seen to limit the applicability of standard EVT methods in the credit risk context. This leads the authors to suggest that more care should be taken when using EVT for credit risk management purposes. In the discussion, this last implication of the paper triggered some controversy, as EVT is already widely used throughout the banking sector to model various types of financial risk, including credit risk. With regard to future work, it was suggested that the authors could consider extending their current, one-factor approach to a multi-factor setting to enhance the applicability of their research.

Finally, as part of the sixth and last conference session, Berkelaar et al investigated how the application of standardised, VaR-based risk management tools might reduce market participants' risk-taking in normal circumstances at the expense of increasing exposure to extreme events. Their paper, therefore, sheds light on how practical risk management using now-standard statistical techniques might affect market dynamics and equilibrium prices. To this end, the authors extend earlier research to find that, in a world with VaR-constrained agents, market volatility (as well as implied options volatility) is generally reduced, generating a stabilising effect for the economy as a whole for most states of the world. However, in extremely bad states, agents have an incentive to gamble by taking large exposures, pushing up market risk and creating a hump in the equilibrium price function. As a consequence, losses for most states are thus reduced at the expense of the remaining states where, with the probability of extreme losses fixed via the VaR constraint, losses will be larger than in the unconstrained case. While this was seen as an interesting and potentially important insight, in the course of the discussion conference participants noted that the results were based on highly restrictive assumptions, such as the strict application of VaR limits and the absence of other risk-related constraints. Given these assumptions, the model was seen as being based on an overly rigid notion of risk management. This raised doubts about the direct practical relevance of this particular model's insights. It was, hence, left for future research to investigate the topic further.

Part 1

Introductory remarks and luncheon addresses

Introductory remarks

by Andrew Crockett,
General Manager, Bank for International Settlements, and
Chairman of the Financial Stability Forum

Introduction

This conference is, as you all know, the third in a series devoted to the same theme. The previous two were held at the Federal Reserve Board in 1995 and at the Bank of Japan in 1998.

In his remarks at the first joint conference in 1995, Alan Greenspan referred to *risk measurement and systemic risk* as parts of a newly evolving area of research in finance and economics. He foresaw that such research would influence the way business would be done in both the private and public sectors. Research on risk measurement-related issues indeed strongly influenced the character of regulatory and policy initiatives as well as of industry practice during recent years.

Making good policy depends on having a clear awareness of the limitations of our knowledge. We do not have all the answers, so we need more research into what we do not fully understand. The focus, of course, is on the practical implications, for both financial regulators and market practitioners, of such research.

Research on risk measurement has enabled private sector institutions to put in place practical risk measurement and management tools to manage their portfolios more efficiently. Market participants have become more able to differentiate among sectoral and country-related risks, and to take pre-emptive precautionary measures. This may help explain the resilience of the global financial system in the face of the economic slowdown, and the apparent absence of contagious effects in the immediate aftermath of the Argentine debt default.

Official sector monitoring of potential risks has also been improved. There is now a much greater awareness of the need for coordination between various policymaking institutions and interests. The establishment of the Financial Stability Forum (FSF) is both a cause and a consequence of this greater awareness. The FSF brings together a wide cross section of representatives of institutions involved in financial stability-related issues. In doing so, it has helped raise awareness of the interrelationships of various aspects of financial stability and to promote the exchange of information and identification of gaps. Taken together, all this should improve our ability to reduce the incidence of financial crises, and to deal with those that nevertheless occur.

We in the BIS are particularly happy to host this conference in Basel, because its subject matter is so close to the heart of the activities of the BIS and the broader central bank community. The CGFS, in particular, has always had as its mission to understand and help shape the structural characteristics of financial markets.

I should say, finally, that now is a particularly appropriate time to address issues of risk measurement. It is true that the global financial system has, overall, shown a remarkable degree of resilience in the face of a confluence of economic shocks. However, recent developments have also exposed concentrations of systemically relevant financial risk exposures. In addition, much of what has been done by the official and private sector to anticipate and manage financial sector risk is now being seriously tested for the first time.

Market developments in response to the crisis in Argentina and the bankruptcy of Enron can be seen as both evidence of the substantial advances in risk management practices and a reminder that there is, and always will be, substantial room for improvement with regard to the way risks are being treated. Issues such as the nature of systemic banking crises, the sources of market liquidity, and how to further refine risk measurement methods, therefore, remain firmly on the policy and research agendas.

So much for the history and purpose of this series of conferences. Let me now consider some of the issues. I will first try to elaborate on what I will call the "***endogeneity of risk***" in the financial system. In my view, this concept is key to understanding the concept of financial instability. I will take the term financial instability to encompass two closely related phenomena: the potential for large destabilising movements in asset prices and the possibility of financial institutions' distress or failure. Of course, periods of stress in financial markets can result from the knock-on effects of singular events involving individual market participants. More normally, however, generalised financial distress arises when groups of market participants are exposed to common risk factors. These factors, in turn, may be exogenous to financial decision-making processes. But, not infrequently, they are the consequences of endogenous forces within the financial system that tend to amplify the impact of exogenous developments or may even generate crisis situations themselves.

In what follows, I will argue that a more comprehensive approach to risk measurement is key to understanding these issues. Against this background, I will have a few words to say about some of the more specific topics addressed in the papers to be presented over the next two days, namely ***banking stability and contagion***, ***market liquidity***, ***exposures to extreme events***, and ***monitoring of systemic risks***.

The endogeneity of risk

Until recently, risks were essentially compartmentalised, with various categories of market and credit risk each being modelled and managed separately. In addition, under what might be called the "old view", sources of risk were seen as largely exogenous. Risk measurement and management systems were essentially based on the assumptions of atomistic markets: markets are made up of a very large number of independent agents, each of them too small to matter. The consequence of this implicit assumption was that risks were seen as independent of market participants' own actions.

Increasingly, however, risk is now seen as multidimensional. Advances in modern finance theory and information technology have identified and defined a multitude of risks, including - as well as market and credit risk - liquidity, operational, legal and reputational risk. Previously combined categories of risk, such as market risk, have been broken down into component categories. And correlations among risk factors have been realised to be of critical importance in the actual measurement of a portfolio's overall risk profile. Consequently, formal statistical models have been generated for the measurement and appropriate management of these risks. This development towards model-based, statistical risk measurement and management has greatly improved financial decision-making, by enabling market participants to more thoroughly understand their exposures. As a result, it can be argued that risk-taking decisions by market participants now conform more closely to their actual risk-bearing capabilities. This should have improved market efficiency, in terms of both pricing and resource allocation, as well as financial stability. However, risk management techniques are constantly evolving as conditions change. Each "crisis" brings to light new weak spots that need to be addressed.

Furthermore, as I said earlier, risk is now seen as endogenous. The environment is not given, but is the product of the actions of individual agents. As a result, systemic stability is critically determined by the collective behaviour of individual market players. Under this "new view", strategies of market participants, including policymakers and regulators, need to take account of any feedback of their collective actions on the conditions under which individual market participants operate. These insights have flowed from the game-theoretic contributions of recent years.

Decision-making processes, therefore, have to take account of the possibility that actions and policies that are reasonable or desirable from an individual perspective may result in unwelcome consequences at the system level. Financial firms need to manage risk with an eye on how their own behaviours are likely to influence those of other market participants. And supervisors need to analyse the interaction between individual incentives and systemic outcomes.

For example, it would be natural for market participants to cut exposures as market prices fall to match their "value-at-risk" to their diminished capital position. Such behaviour, especially by players whose positions are large relative to the overall market, would tend to deepen the decline in prices. This, in turn, might feed into other players' decision-making, potentially triggering further sales and a vicious circle that could end in a drying-up of market liquidity and a spreading of financial stress.

How serious in practice this phenomenon of endogeneity is depends on a number of factors, including the number of players and the diversity of their behaviour. It has been argued, for example, that the now widespread use of a relatively small number of similar risk management systems may induce significant numbers of market participants to respond to market developments in similar ways. This is not to say that the move towards more sophisticated, statistical risk management models should be abandoned. By no means. These approaches have, for good reason, been widely adopted throughout the financial community - a development, by the way, which has been much supported by Basel-based bodies. Still, the fact that similar models are being used is a fact that is relevant, both to the optimal course of action for individual firms, and to the incentives embodied in supervisory guidance.

Let me repeat a point made earlier: there is always room for improvement in terms of understanding the limitations of what we know and of how this knowledge is being applied. It is for this reason that issues such as the appropriate treatment of operational and liquidity risks or the formal integration of market and counterparty credit risk have attracted growing interest. For the same reason, we still have to more fully understand the nature of systemic banking crises, the dynamics of market liquidity, and contagious effects across markets and countries.

There is a lively debate on these and related issues in academic as well as central bank and practitioners' circles, which I am sure will be taken further during this conference, in the light of the interesting papers that will be presented on these topics.

Specific issues

Let me turn now to some of these specific issues to be addressed in the conference:

Banking stability and contagion

Thinking about the nature, causes and transmission of crises has developed a great deal in recent years, building on the original Diamond and Dybvig model and other studies on banking crises and contagion. However, the nature and causes of systemic crises and of contagion across markets and countries are still only partially understood. Theoretical as well as empirical work on contagion is, therefore, still necessary, particularly as contagion continues to mean different things to different people. Some of the papers to be discussed today explore these issues. The models presented in these papers examine how financial turmoil might "travel" from one country or market to another. For example, sequential trading in the presence of asymmetric information may trigger contagious asset price movements. Movements in asset prices are important in determining the probability of bank runs. The way in which bank mergers take place can affect bank balance sheets and, in turn, system stability. Finally, the degree of development of bond markets can be shown to influence the effectiveness of financial market discipline and thus reduce overlending. An important policy implication of this analysis is the role of market development in helping to avoid emerging market crises.

Ultimately, thinking about these models helps improve our understanding of the real world. In turn, this understanding should eventually be reflected in risk management tools and prudential policies. I have myself spoken more than once on the need to add a degree of macroprudential orientation to existing regulatory and supervisory frameworks. But what are the appropriate instruments? Some have pointed to stress testing techniques or provisioning practices. Stress tests, for example, are used to supplement traditional risk measurement approaches, like value-at-risk. They are, therefore, recognition of the limited ability of such statistical models to accurately capture exposures under exceptional circumstances. These are just some of the questions we will be addressing during this first day of sessions.

Endogeneity of risk and market liquidity

I have already talked about how important it is, for market participants and policymakers alike, to understand the implications of endogeneity of risk. Nowhere is this endogeneity clearer or more important than in the matter of liquidity risk. Liquidity is, almost by definition, the combined result of the actions of a multiplicity of market players. Its availability depends on the existence of a diversity of market views, something that is in turn influenced by the evolution of risk management practices.

We know that, at times, market liquidity can evaporate, making trading impossible or, at least, much more difficult. In response, market participants, partially due to events like the LTCM crisis, have come to grasp the importance of liquidity risk. But work still needs to be done to more fully understand the sources of market liquidity and to deal with liquidity risk in a more sophisticated fashion, for example, by applying stress testing techniques. I am encouraged to see that some of the papers to be presented at this conference address, in various ways, these issues. Other related topics include, for example, the potential importance of large investors in the determination of market prices and liquidity, and the microstructural specifics of liquidity provision on electronic FX trading platforms and in the Treasury and corporate bond and equity markets.

Exposure to extreme events

A number of the conference papers present efforts to enhance our understanding of the tails of statistical distributions of returns. And, indeed, it is the tails of the distributions that, from a financial stability perspective, matter most. For it is in times of stress, rather than in normal times, that traditional risk measurement models tend to convey imprecise or misleading information. One of the papers to be presented tomorrow compares two popular summary measures of financial risk, value-at-risk and expected shortfall, while another describes how the tail of the loss distribution in portfolio credit risk models depends on modelling assumptions and certain parameter choices. Developing our understanding of these issues is central to moving beyond summary statistics such as VaR as a sufficient expression of the risk profile of an enterprise or a trading activity. Doing so, however, involves addressing even more complex issues. For example, how might the strategic interaction of market participants and use of standard measures of risk lead market participants to underestimate the true risk of their positions? Can this tendency to underestimate be quantified? Can offsetting incentives be designed? We understand by now that the strict application of certain risk management tools such as VaR can reduce risk-taking in normal circumstances at the expense of increasing exposure to extreme events. This could well make crises much worse once they strike. Of course, we don't want to "turn back the clock". What we need to do, however, is to understand the potential implications of what is being done and to avoid that the processes used are implemented in an overly rigid fashion, potentially impairing the scope for independent judgment by the decision-maker.

Monitoring of systemic risks

It should be clear by now that the analysis of systemic risks is high on the policy agendas of central banks. Some of the papers to be presented tomorrow show how financial market and banking data can be used to monitor the fragility of banking sectors. One paper, for example, attempts to show how Merton-type, market-based indicators can be usefully employed to predict banking fragility by adding to the information gained from more traditional, balance sheet-based indicators. In this regard, I find it particularly useful to have the opportunity to hear how a central bank assesses potential contagion risks in the banking sector in practice - by monitoring counterparty exposures in the interbank market using unique data detailing the largest uncollateralised exposures of the four major market players.

Conclusion

Let me again highlight the main goal of the conference, which is to bring together the research and policy communities in order to achieve a "virtuous loop" of interaction that provides feedback from the policy agenda to research and back to the policy agenda. I am sure that this conference will take us a step further in this regard, and I look forward to a stimulating two days of discussions.

Triangular view of systemic risk and central bank responsibility

Speech by Yutaka Yamaguchi,
Deputy Governor of the Bank of Japan and Chairman of the
Committee on the Global Financial System

Introduction: A brief history of the Systemic Risk Conference

I am very pleased to join you at the third conference on "Risk Measurement and Systemic Risk". This theme has gained increasing importance since the first conference in 1995. The keen attention paid to it by the international community is evident in the fact that a number of international forums now endeavour to spot potential financial vulnerabilities which might lead to systemic crisis in the global market. Actually, the term "systemic risk" makes me a bit uneasy as it unfortunately has too realistic a connotation in my own country. I am looking forward to taking home new insights on this subject, and, as Chair of the Committee on the Global Financial System co-hosting this conference, I would be happy if you could do the same.

The aim of this series of conferences is to enhance our understanding of the mechanism through which a systemic shock is generated and transmitted. Meanwhile, during the six years since the first conference, we have witnessed significant changes in the world of finance. As a result, the focus of the conferences has changed over the years. If I may generalise at the risk of oversimplification, the centre stage of the first conference in 1995 was occupied by VaR methodology, which was then gaining acceptance at leading financial institutions. Reporters explored how risk could be quantitatively measured and what would be the real-life meaning of such measures. Well, in real life, crisis erupted in Asia in 1997 triggered by events that were largely beyond the bounds of standard VaR methodology. Naturally, discussions at the second conference in 1998 were much influenced by the Asian crisis. We began to realise that market microstructure theory could shed light on market dynamics in times of stress. Our third conference today carries this theme further, with many papers paying attention to what creates stress and how stress is contagious.

Triangular view of systemic risk

This brief history of our conference series suggests that with the structural changes in financial markets, systemic risk has revealed a few faces in actual crisis and therefore the nexus between them has to be more deeply explored. Conventional thinking or the narrowest coverage inextricably tied systemic risk to banks. Systemic disturbances that originate in a bank spill over to the banking system, which in turn adversely affects the real economy. Obviously, this bank-centred risk propagation still holds; in fact, much of the existing safety net is aimed at preventing a chain reaction within the banking system. However, it has now become evident that financial markets play a significant role as sources of the disturbances as well as channels propagating them originated in the banking system and the real economy.

The importance of the market and its dynamics is underscored by our recent experiences in Japan, the Asian region, Russia and the LTCM case. The novelty of the Russian and the LTCM crises lies in the fact that the largest capital market in the world "seized up" without entailing any banking crisis. It was often the case that the sudden deterioration in asset prices brought about turbulence in the financial system. To illustrate, the successive failures of major Japanese financial institutions in 1997 and 1998 were not directly triggered by a major default. Instead, their undoing was a rapid loss of confidence in the market. Typically, as the soundness of a bank was questioned in the market, prices of its stocks and credit ratings started to fall. The bank would then begin to experience funding difficulties, as its access to markets became problematic. In such a situation, the troubled bank had to resort to fire sales of assets, which in turn damaged its balance sheet and drove its stocks down even further. In this self-fulfilling spiral, several banks went out of business. At the same time, the deterioration in asset prices led to further difficulties in channelling funds to the corporate sector, a familiar credit crunch process.

A credit crunch is usually attributed to the dysfunction of the banking system - a correct observation of one aspect of such a phenomenon. A deeper look suggests that the process is more complex. We have witnessed that the borrowers blame the banks for tightening credit standards, while the lenders complain of the lack of credit demand. No doubt, an important feedback mechanism also runs from the real economy to the financial system via corporate balance sheets, asset prices, and banks' capital position, among others.

I am not attempting to draw definitive lessons from a specific episode of the past crisis, let alone from the unsolved problems of the Japanese economy. However, the experiences of the last several years show that disturbances are multifaceted. Systemic problems develop as market risk, liquidity risk and credit risk factors interact with each other in a complex manner. This means that any evaluation of systemic risk based on one isolated factor could only provide a fragmentary view. What is called for is the "triangular view of systemic risk" - comprehensive analysis covering the interrelations or nexus between the banking system, financial markets and the real economy. It is against this background that I think we need to devote at least as much attention to market microstructure as to sophisticated analyses of "fat tails" in loss distribution. A focus on market microstructure could shed light on the relations between various risk factors. Particularly important is to investigate how individual market participants under different budget and information constraints would behave rationally when faced with stressful events, and how such behaviour would affect the formation of asset prices.

Strategic interactions

Recent episodes of financial crises seem to defy explanation on the basis of conventional economic theory, which regards macroeconomic phenomena as a mere aggregate of independent decision-making by economic agents. As a reflection of such limitations of conventional theory, there is a growing body of work attempting to interpret financial crises from the viewpoint of "strategic interactions" among market participants. I would like to devote a few minutes to outlining why.

Strategic interaction can be defined as a process in which each market participant explores his/her optimal strategy by conjecturing the response of other participants. Some of the papers presented at this conference follow this path. Herd behaviour is one example. As you know, a large number of small investors tend to follow a small number of large investors. Once stressful events happen, such behaviour is likely to lead to one-sided market sentiment, which accelerates and propagates the stress within and across the markets. From the viewpoint of policymakers, herd behaviour as a phenomenon is hard to tackle. If we understand such a phenomenon as a consequence of strategic interactions among market participants, however, we might find a key to reducing the risk of triggering herd behaviour.

According to my reading of this line of research, the outbreak of systemic disturbances would heavily depend on how many market participants, when faced with systemic threats, expect disturbances to actually occur. In other words, a crisis is not necessarily an accident, but a consequence of market participants' expectations. Their expectations are formed from conjectural views of other market participants' responses to such threats. The magnitude of any crisis and the extent of contagion would critically depend on the feedback from market participants reacting collectively to systemic threats. Feedbacks could also accelerate any crisis. These explanations seem to offer a useful perspective on the mechanism of systemic disturbances and appropriate policy responses thereto.

The strategic interaction framework seems to offer us a roadmap for developing more stress-resistant markets. A possible approach would be to enhance the visibility of future stress. Let us suppose that there is a scenario consisting of a series of events leading to stress. If market participants have the view that such a scenario could result in a serious impact on a market in the future, they might take necessary actions to avoid losses which would materialise under the scenario. As long as market participants take necessary actions gradually and individually, the actual impact of events as they happen would be softened and stress would not materialise. In other words, a stress scenario would not remain a stress scenario once it is publicly recognised as such. In fact, we observe such episodes in financial markets. For example, proposed changes to accounting rules sometimes raise concerns initially, but only rarely would they result in severe impacts when they are implemented. Based on these experiences, I should say that an approach enhancing the visibility of stress appears more appealing.

“Macro stress-census”, an experiment conducted by the CGFS, might be one of the options for developing commonly recognised stress scenarios among market participants and central banks, although not a panacea.

Challenges to central banks

Before concluding my remarks, let me outline the challenges facing central banks with regard to systemic risk. In the six years since the first Systemic Risk Conference, we have learned considerably from our involvement in real-life crises and through intellectual interchanges at this conference and other venues. At the same time, one answer leads to new questions and there remain many unanswered questions. The same can be said for policy responses by central banks in times of financial crisis.

In relation to policy responses to systemic risk, we have generally recognised the importance of both pre-emptive actions, ie actions aimed at preventing systemic problems, and after-the-fact measures to contain an unfolding crisis. In this regard, I see a greater rationale than ever for views that stress the importance of preventive measures. This is because globalisation of financial markets and consolidation of financial institutions have considerably raised the possible costs of dealing with actual systemic disturbances. To this end, the strengthening of market discipline as well as supervision would be essential, and the international community has made serious joint efforts.

However, even the best of preventive measures may not be always successful in completely removing sources of systemic crises in an environment where financial intermediation keeps evolving at a speed beyond our wildest imagination. If there is the slightest chance of severe financial disturbances, the central bank must not lower its guard. In envisaging crisis management, the changing environment could compel us to rethink established doctrines.

For example, there is no doubt about stressing that we need to minimise moral hazard. Nevertheless, in the event that a systemic crisis is actually unfolding, we must not overlook the fact that there is an inherent, conflicting aspect in crisis management. In a sense, crisis management artificially creates moral hazard in order to avoid catastrophic consequences. In real-life policy responses, authorities inevitably face a trade-off between prevention of systemic crisis and minimisation of moral hazard. Another example concerns the traditional lender of last resort functions of the central bank. According to traditional thinking, this is aimed only at banks. But the contemporary reality, as I noted earlier, is that systemic problems could originate in financial markets and such markets are populated not only by banks but also by a large number of non-bank financial institutions and conglomerates. This may argue in favour of the view that the traditional principle should be augmented. A related issue is the conditions under which central banks would take certain policy actions, such as invoking their lender of last resort functions. Traditionally, “constructive ambiguity” was regarded as the golden rule in such cases, but the Bank of Japan distanced itself from this in dealing with the crisis in the late 1990s, with a view to precluding speculations and enhancing policy transparency and accountability. The issue of the practical significance of “constructive ambiguity” must be explored vigorously without leaving any ambiguity.

Conclusions

Today, I have offered my views on systemic risk, which might have raised more questions than answers for central banks. In concluding my remarks, I would like to stress that central banks must continue to pursue these issues to discharge their responsibilities. The responsibilities arise from the following facts. First, central banks are unique economic agents having relations with each corner of the systemic risk triangle - the banking system, financial markets and the real economy. Second, central banks are expected to confront almost every systemic crisis as entities that can readily provide liquidity. Fortunately, central banks have made progress in gaining insights through extensive research on market dynamics. Nevertheless, central banks must not be satisfied with what they have achieved so far. In order to answer the remaining questions, and refine views on established concepts, we are looking forward to continuously interacting with market participants, who have first-hand knowledge of the markets, and members of academia, who have been laying the groundwork. In this regard, I hope this series of conferences will remain a valuable venue that continues to inspire the central bank community.

Reflections on recent financial incidents

Luncheon speech given by Tommaso Padoa-Schioppa

Introduction

I am delighted to have this opportunity to meet this distinguished group of experts and former colleagues that has come together here in Basel to study issues related to financial stability. I would like to share with you a few thoughts, inspired by the two recent cases of Enron and Argentina, respectively the largest ever corporate and sovereign defaults we remember. Like many observers, I will try to identify whether the two cases raise any *fundamental questions* concerning the functioning of the financial system and the interplay between market forces and public authorities.

I started my career as a central banker some 35 years ago, when public intervention in the economy, and particularly in the financial sphere, was very pervasive. Almost all business activities by banks required a specific authorisation, and many actions were simply forbidden. Market participants had little room for free and innovative action. It has taken a long time for the *pendulum* of ideas, economic realities and policies to move towards market forces. Public intervention has been gradually scaled back, from having an excessively wide scope to a narrow one, carefully targeted at market failures. This long shift - which has accidentally coincided with the span of my professional life, and to which both my actions and my convictions have fully adhered - has produced extraordinarily large efficiency gains from which our economies have greatly benefited.

Enron and Argentina can undoubtedly be looked at from various angles, and only time will clarify the *lessons* we have learnt from the two cases. One question we can ask today is: do Enron and Argentina indicate that the pendulum may not be very far from swinging back again, between the two extremes of very pervasive public intervention and complete laissez-faire? Is it time to reconsider - with some historical perspective - what public intervention is needed to best support the orderly functioning of financial markets? Without pretending to provide full answers to these questions, let me just offer a few thoughts.

Risks in the financial sector

I start with two observations about the origin and propagation of risks. First, although the increasing complexity of the financial system renders it more and more difficult to identify the *origin* of risks, we should never forget that the threat to financial stability stems, fundamentally, from the *real sector*. It is in the real sector that events occur that ultimately cause gains and losses in the financial field. Such events may be the unexpected disruption of a particular market, a price shock, a sharp change in technology, the deterioration of macroeconomic conditions or the policy decisions of a government. Managing the risks associated with the uncertainty and risk of the real sector is at the core of financial intermediation.

Second, the *propagation* of risk. The way in which risk is spread within the financial system varies over time in relation to several factors, including market and regulatory developments. Enron and Argentina highlight once again the importance of two aspects that characterise risk propagation today: first, the growing use of complex financial instruments to assume and transfer risks and, second, the abrupt changes in international capital movements. As to the first, some evidence suggests that the markets for credit risk transfer instruments are quite concentrated, in terms of both dealers and ultimate risk-takers. As to the second, lack of data on many important players in the global financial system leaves us relatively uninformed about the possible sources of destabilising capital movements.

Enron: failure to deliver transparency

Let me move to the Enron case. There seems to be a broad consensus that this incident points not only to truly illegal actions and infringements of ethical codes of conduct, but also to ineffective *market discipline* exercised by Enron's equity and debt holders, due to lack of adequate transparency. Enron owed much of its initial success to deregulation, both in the gas and electricity sectors and in a variety of other areas. It was publicly perceived as a highly successful company. Only when the company was approaching bankruptcy did market analysts react and shareholders and creditors become aware of its vulnerabilities. Only then did attention focus on the risks entailed in its extensive off-balance sheet transactions. Inadequate accounting rules are partly responsible for the failure to uncover highly risky operations or for the inadequate disclosure of complex off-balance sheet transactions. The extensive and parallel consulting business with Enron that auditors entertained is also to blame.

We are in the process of drawing many lessons regarding the public policies required to ensure the smooth operation of market discipline, which is also of utmost importance for the functioning of the financial system. Of these lessons, three - in my opinion - stand out as crucial. First, it is timely to recall Paul Volcker's proposal of June last year calling for an international initiative to update *accounting standards* so as to adequately deal with the complexity of derivative financial instruments. We should do our utmost to ensure that the Enron affair serves as a powerful incentive to speed up efforts in this field.

Second, the case highlights the question of adequate *oversight of financial activities* undertaken by non-financial corporations. Despite being the main dealer, market-maker and liquidity provider in important areas of the energy and other derivatives markets, Enron was not required, by either regulators or market practice, to disclose information to its counterparties, or to set aside capital against its trading risks. The absence of such mechanisms prevented an early detection of the problem and might even have created incentives for imprudent risk-taking.

Finally, the case suggests that our system is not sufficiently alert to possible *conflicts of interest*. The combination of auditing and consulting in the Enron case is only one example. Such conflicts arise whenever a financial institution provides corporate finance and similar services to a specific client who issues securities in which the financial institution can invest its own funds or those of its clients.

All in all, these three issues give cause for concern and also deserve careful consideration by public authorities. My feeling here is that, if a player such as Enron is not under the control of regulators, it should be under tight market control exercised by analysts, accountants, shareholders and lending banks. If these *endogenous* controllers fail to be alert, they should be sanctioned in the form of monetary losses or regulatory constraints.

Argentina: hands-off approach coupled with official sector weakness

Let me turn to Argentina. Here, the lessons are at the international *macroeconomic* level. Not too long ago, Argentina was the focus of attention, though for very different reasons than now. In the early 1990s, "neo-liberal" economic reforms were implemented; hyperinflation was brought to a halt; the economy was progressively deregulated and privatised. As macroeconomic stability was achieved, foreign capital poured into the economy and growth quickly resumed.

In a continent that had just emerged from the debt crisis of the 1980s and with very few success stories to tell, Argentina's experience under this economic paradigm was *very positive* for much of the 1990s, growing at an average rate of nearly 5% from 1991 to 1998. This was a period marked by a series of external shocks, which Argentina's currency board successfully overcame, namely, the "Tequila" crisis in 1995, the East Asian crises in 1997, and the devaluation of the Russian rouble in 1998. But they were not cost-free: in the absence of using the exchange rate as a shock absorber, the burden of adjustment in the economy under a currency board agreement necessarily falls on wages and prices.

In the case of Argentina, the rigidity of the hard peg came to the forefront in the wake of a series of external shocks in early 1999 - notably the higher cost of financing to emerging markets, the sharp devaluation of the Brazilian real, the rapidly appreciating US dollar, and falling commodity prices. The straitjacket imposed by the currency board cast doubts on Argentina's medium-term economic performance, and concerns about its ability to service and refinance debt were further compounded by the relative fiscal laxity in previous years.

As the credibility of the currency board came under increasing pressure, the country required policy adjustments but also sustained *signs of support* from the official sector. Yet Argentina's misfortune is that, as its need for official financing was increasing, opinions about how multilateral agencies should act when faced with emerging market crises were changing - in particular, with regard to the need of engaging private creditors (particularly bondholders) in the resolution of debt problems. The last IMF package in support of Argentina (September 2001), for example, contained specific provisions to this end. In the subsequent months, Fund officials publicly encouraged the Argentine authorities to reach an agreement with private sector agents over debt exchange operations. And in December 2001, the Fund suspended its loan programmes with the country.

While *private sector involvement* in crisis resolution should be welcome, one may wonder if it should be the sole instrument to deal with such circumstances. Many feel that the official sector was rather unkind to Argentina. After all, its macroeconomic indicators - particularly its fiscal accounts, which were the main source of concern - were broadly the same as (or better than) other countries which had received large IMF funding in recent years, such as Turkey or Brazil. Argentina's central government debt in 2001 was less than 55% of GDP, and its government deficit (including the provinces) amounted to less than 6% of GDP in the same year. In contrast, Turkey posted a 57.4% debt-to-GDP ratio and a government deficit of 11.6% of GDP in 2000, right before its currency and banking crises. Brazil had a government deficit of 7.9% and external debt-to-GDP ratio of 30% in 1998.

The international community is relieved that economic and financial *contagion* has not spread from Argentina to other economies in the region, notably Mexico and Brazil. Yet I cannot but wonder how Mexico and Brazil would be doing today, had the same Argentine-style "hands-off" approach been followed back in 1995 and 1999 respectively.

Conclusion

Let me conclude. Are we addressing Enron and Argentina jointly just because the two events happened at the same time or because they have something else in common? There is no doubt that the two cases are quite different. Yet I see both of them as a reminder that we need to distinguish clearly between the *scope* of public intervention and its *effectiveness*. Where there is room for public action, a *minimum* scope of intervention should not be tantamount to *weak* or ineffective intervention. The important lesson that emerged from the past experience of overextended public intervention is about excessive scope and not about unnecessary strength.

Both events highlight weak responses by the authorities to a deteriorating situation. In the case of Enron, the signals provided by market authorities and policymakers were not strong enough to ensure adequate transparency and avoid conflicts of interest. While some initiatives to improve the situation were put forward over a relatively long period before the Enron incident, the prevailing pressure from the corporate sector prevented substantive achievements. Regulators and policymakers have something in common with policemen. A policeman has to be friendly and helpful to citizens - just as regulators need to be market-friendly - but a policeman always has to remember who he is.

Hence, the *main lesson* I would draw from the recent events is that strong public intervention is necessary on those occasions when markets fail to work properly. This should not be confused with a wide and pervasive intervention in the markets as public authorities used to do in the past. We who are responsible for the oversight of markets should signal our commitment to well-defined and effective intervention, when needed, and thus contribute to the stability of the financial system.

Part 2

Papers

Session 1

Banking stability

Liquidity, asset prices and systemic risk

Franklin Allen, Department of Finance, Wharton School,
University of Pennsylvania
Douglas Gale, Department of Economics, New York University[1]

Abstract

Central banks often intervene in the financial system to prevent crises. They frequently cite financial fragility or the contagion that might otherwise occur as a justification for their actions. This argument has traditionally been based on historical experience rather than a theoretical understanding of these phenomena. This paper discusses a theoretical framework for considering these issues and the role of central bank intervention.

1. Introduction

In August 1998 the Russian government defaulted on its domestic debt. Despite the fact that the amount of this debt was small relative to the total value of assets in the world, the event had a large effect on the global financial system. On the day the default was announced, three quarters of the stock markets in the world fell (Kaminsky and Schmukler (1999)). In the weeks that followed, there was considerable turbulence in financial markets. In October, the Federal Reserve Bank of New York facilitated a takeover of the hedge fund Long-Term Capital Management (LTCM) by a consortium of major banks. The unusual movements in asset prices following the Russian default had brought LTCM to the brink of bankruptcy. If LTCM had gone into bankruptcy, its assets would have been liquidated. The precise way in which this bankruptcy would have been handled was fraught with uncertainty. LTCM was incorporated in the Cayman Islands and there were few precedents for this type of event there (Allen and Herring (2001)).

One rationale for the New York Fed’s action was the argument that the financial markets in which LTCM traded were fragile and subject to contagion: rapid liquidation of such a substantial amount of assets would have overwhelmed the liquidity available in the markets, causing a significant drop in asset prices. This would have caused problems for other intermediaries that in turn might have been forced to liquidate assets, causing prices to fall even further. The cumulative effect of LTCM’s default might have been a global financial crisis. The New York Fed’s action pre-empted this possibility and markets soon stabilised.

Did the bankruptcy of LTCM really pose a systemic risk for the global financial system? Would asset prices have collapsed if LTCM had been forced to liquidate its assets in a short space of time? Many have doubted this and argued that the New York Fed acted inappropriately. Standard models of asset pricing suggest that a single liquidation of the size of LTCM will not lead to a meltdown in asset prices. According to these models, asset prices are determined by the discounted stream of cash flows generated by the assets. Changes in the supply of assets to the market does not affect their price provided that such changes do not signal information (Scholes (1972)). It seems unlikely that LTCM’s bankruptcy would have signalled very much about the future cash flows of corporations or discount rates in the global economy. So, according to this view, the New York Fed’s intervention was unnecessary.

In this paper we review recent theories of financial crises. In particular, we are interested in understanding the systemic risk associated with financial fragility and contagion and how central banks should respond. The rest of the paper is organised as follows. Section 2 reviews the early literature on

[1] Allenf@wharton.upenn.edu, Douglas.Gale@nyu.edu.

financial crises, based on the Diamond-Dybvig model of bank runs. Section 3 introduces a more recent class of models in which the source of financial crises is real business cycle shocks. This section focuses on the welfare economics of financial crises from the point of view of optimal risk-sharing. One of the important elements of this discussion is the relationship between market provision of liquidity and its effect on asset prices, which is further explored in Section 4. In Section 5 we return to the debate about whether financial crises result from real business cycle shocks or self-fulfilling expectations. Section 6 discusses models of contagion. Section 7 sums up our discussion of the policy implications of the research reviewed here.

2. Risk-sharing

First-generation models of financial crises were developed in the 1980s, beginning with seminal work on bank runs by Bryant (1980) and Diamond and Dybvig (1983) (hereafter DD). Important contributions were also made by Chari and Jagannathan (1988), Chari (1989), Champ et al (1996), Jacklin (1987), Jacklin and Bhattacharya (1988), Postlewaite and Vives (1987), Wallace (1988; 1990) and others. DD is at the core of most models in the literature on bank-centred financial crises. A typical DD-style model has three dates $t = 0, 1, 2$ and a large number of identical consumers. Each consumer is endowed with one unit of a homogeneous consumption good. At date 1, the consumers learn whether they are early consumers, who only value consumption at date 1, or late consumers, who only value consumption at date 2. Consumers' uncertainty creates a preference for liquidity.

An individual can invest in a combination of long-term (illiquid) investments, yielding high returns, and short-term (liquid) assets, yielding low returns. The short-term asset pays a return of one unit after one period and the long-term asset pays a return $r < 1$ after one period or $R > 1$ after two periods. The long asset has a higher return if held to maturity, but liquidating it in the middle period is costly, so it is not very useful for providing liquidity. An individual investor faces a difficult choice between return and liquidity. If he holds the long asset and turns out to be an early consumer, he will lack liquidity. If he holds the short asset and turns out to be a late consumer, his returns will be low. What he really wants is insurance against his uncertain demand for liquidity, but this he cannot provide by holding a mixture of the two assets.

Banks have a comparative advantage in providing liquidity insurance. The bank can offer each depositor a superior contract, promising a combination of liquidity and high returns that an individual investor cannot match using markets. For simplicity, assume that the fraction of early consumers is constant. Thus, there is no uncertainty about the aggregate demand for liquidity. There is only uncertainty about which individuals will demand liquidity at the intermediate date. At the first date, consumers deposit their endowments in the banks, which invest them on behalf of the depositors. In exchange, depositors are promised a fixed amount of consumption at each subsequent date, depending on when they choose to withdraw. The banking sector is perfectly competitive, so banks offer risk-sharing contracts that maximise depositors' ex ante expected utility, subject to a zero profit constraint.

DD provides a simple explanation of bank runs. The optimal insurance scheme requires the bank to promise depositors a fixed payment if they withdraw early. If too many depositors withdraw, the bank is unable to meet its promises without liquidating assets. Under some conditions, if most or all of the early depositors withdraw early, there will be nothing left for those who withdraw late. Thus, a bank run becomes a self-fulfilling prophecy: if a depositor believes that others will withdraw their deposits from the bank, it becomes optimal for the depositor to withdraw his deposits too.

There are two Nash equilibria of this "game", one in which only the early consumers (those who have received a liquidity shock) withdraw early and one in which everyone withdraws early. The former is incentive-efficient, the latter is not. What determines which equilibrium is observed? Market psychology? Animal spirits? Sunspots? We return to this point in Section 5 below.

Here we want to emphasise the importance of DD as a contribution to the microeconomic theory of intermediation. Apart from its usefulness as a model of bank runs, the DD model is remarkable for providing an account of what banks do and why they are needed. The insurance function (converting liquid liabilities into illiquid assets) is interesting in its own right, because it provides a theory of banking based on rational optimising behaviour and opens it up to microeconomic analysis. The same approach can be extended to the welfare analysis of monetary and banking policy.

3. Optimal financial crises

Gorton (1988) finds evidence from the United States during the National Banking Era which is consistent with the view that banking panics are related to the business cycle. The five worst recessions, as measured by the change in pig iron production, were accompanied by panics. In all, panics occurred in seven out of the 11 cycles. Using the liabilities of failed businesses as a leading economic indicator, Gorton finds that panics were systematic events: whenever this leading economic indicator reached a certain threshold, a panic ensued. The stylised facts uncovered by Gorton thus suggest banking panics are intimately related to the state of the business cycle. Calomiris and Gorton (1991) consider a broad range of evidence and reach similar conclusions.

A number of authors have developed models of banking panics caused by aggregate risk. Wallace (1988; 1990), Chari (1989) and Champ et al (1996) extend Diamond and Dybvig (1983) by assuming that the fraction of the population requiring liquidity is random. Chari and Jagannathan (1988), Jacklin and Bhattacharya (1988), Hellwig (1994) and Alonso (1996) introduce aggregate uncertainty which can be interpreted as business cycle risk. Chari and Jagannathan (1988) focus on a signal extraction problem where part of the population observes a signal about future returns. Others must then try to deduce from observed withdrawals whether an unfavourable signal was received by this group or whether liquidity needs happen to be high. Chari and Jagannathan are able to show panics occur not only when the outlook is poor but also when liquidity needs turn out to be high. Jacklin and Bhattacharya (1988) also consider a model where some depositors receive an interim signal about risk. They show that the optimality of bank deposits compared to equities depends on the characteristics of the risky investment. Hellwig (1994) considers a model where the reinvestment rate is random and shows that the risk should be borne both by early and late withdrawers. Alonso (1996) demonstrates using numerical examples that contracts where runs occur may be better than contracts which ensure runs do not occur because they improve risk-sharing.

Starting from this point of view, Allen and Gale (1998) (hereafter AG) construct a model in which financial crises are driven by fundamentals. An economic downturn will reduce the value of bank assets, raising the possibility of banks being unable to meet their commitments. If depositors receive information about an impending downturn, they will anticipate financial difficulties in the banking sector and try to withdraw their funds. This attempt will precipitate a crisis.

The main objective of AG is to analyse the welfare properties of the model and understand the role of central banks in dealing with panics. Bank runs are an inevitable consequence of the standard deposit contract in a world with aggregate uncertainty about asset returns. Furthermore, they play a useful role insofar as they allow the banking system to share these risks among depositors.

The basic assumptions about technology and preferences have become standard in the literature since the appearance of DD. AG retains many of the standard assumptions in the model but it differs from DD in important ways.

- There are aggregate shocks to asset returns. More precisely, banks hold long-term (illiquid) assets that earn a random return $\tilde{R}$ at date 2. Moreover, the returns to the risky assets are perfectly correlated across banks. Uncertainty about asset returns is intended to capture the impact of the business cycle on the value of bank assets. Information about returns becomes available before the returns are realised and when the information is bad it has the power to precipitate a crisis.
- The long-term asset is completely illiquid: none of the returns from this asset are available for sharing out among the early consumers. This is different from assuming that $r = 0$ in the DD model. Here the long asset cannot be touched in the short run. The “winding-up” of an insolvent bank must take time and, more importantly, there will be something left for late withdrawing depositors.
- AG does not impose the first come, first served assumption. (This assumption has been the subject of some debate in the literature as it is not an optimal arrangement in the basic DD model (see Wallace (1988) and Calomiris and Kahn (1991)). Instead, an insolvent bank shares its liquid assets equally among the early withdrawers. Those who do not withdraw early have to wait to obtain their funds and, again, they share the remaining assets equally.

In a number of countries and historical time periods, banks have had the right to delay payment for some time period on certain types of account. This is rather different from the first come, first served assumption. Sprague (1910) recounts how in the United States in the late 19th century people could obtain liquidity once a panic had started by using certified cheques. These cheques traded at a discount.

In the simplest version of the model, which serves as a benchmark for the rest of AG, there are no costs of early withdrawal, apart from the potential distortions that bank runs create for equilibrium risk-sharing and portfolio choice. In this context, a laissez faire banking system which is vulnerable to crises can actually achieve the first-best allocation of risk and investment. The first-best allocation can be identified with an optimal mechanism-design problem in which the allocation can be made contingent on a leading economic indicator (eg the return on the risky asset), but not on the depositors' types. By contrast, a standard deposit contract cannot be made contingent on the leading indicator. However, depositors can observe the leading indicator and make their withdrawal decisions contingent on it. When late-consuming depositors observe that returns are high, they can deduce that they will obtain more by waiting and are content to leave their funds in the bank until the last date. When the returns are low, they deduce that they are better off to withdraw their funds than leave them in, causing a bank run. The somewhat surprising result is that the optimal deposit contract produces the same portfolio and consumption allocation as the first-best allocation. The possibility of equilibrium bank runs allows banks to hold the first-best portfolio and produces just the right degree of contingency to provide first-best risk-sharing.

The idea that financial crises can be optimal is an important one. It is often taken as axiomatic among policymakers and in the literature that crises should be avoided at all costs. As the example in AG indicates, crises can perform a useful role in sharing risk. In fact, Allen and Gale (2000a) are able to show in the context of a more general model of banking and financial markets that crises can be constrained-efficient in a wide range of circumstances. They argue that policies towards crises need to be based on a careful understanding of the nature of the market failure that occurs. In the absence of market failure, intervention by the central bank may not be justified.

To provide a rationale for central bank intervention, some cost of early liquidation has to be introduced. AG considers a second version of the model in which the storage technology available to the banks is strictly more productive than the storage technology available to the late consumers who withdraw their deposits in a bank run. In these circumstances, where crises are costly, appropriate central bank intervention can avoid the unnecessary costs of bank runs while continuing to allow runs to fulfil their risk-sharing function. A bank run, by forcing the early liquidation of too much of the safe asset, actually reduces the amount of consumption available to depositors. In this case, laissez faire does not achieve the first-best allocation. This provides a rationale for central bank intervention. The central bank can intervene with a monetary injection that implements the first-best allocation. Suppose that a bank promises the depositor a fixed nominal amount and that, in the event of a run, the central bank makes an interest-free loan to the bank. The bank can meet its commitments by paying out cash, thus avoiding premature liquidation of the safe asset. Equilibrium adjustments of the price level at the two dates ensure that early and late consumers end up with the correct amount of consumption at each date and the bank ends up with the money it needs to repay its loan to the central bank. The first-best allocation is thus implemented by a combination of a standard deposit contract and bank runs.

Finally, AG considers the role of markets for the illiquid asset in providing liquidity for the banking system. The first two versions of the model have the very special feature that the risky, long-term asset is completely illiquid. There is no way of liquidating the risky asset to meet the claims of the early consumers, and this plays an important role in the "equilibration" of a bank run: the fact that some assets are always left over at the final date means that it can never be optimal in equilibrium for all the late consumers to join a run and withdraw early.

In the third version of the model, there is an asset market in which the risky asset can be traded and this provides a means of liquidating the long-term asset. Somewhat surprisingly, the introduction of asset markets leads to a Pareto *reduction* in welfare in the laissez faire case. The bankruptcy rules force the bank to liquidate as much of its assets as possible in an attempt to meet the claims of the depositors who withdraw early. Liquidation turns out to be self-defeating because the asset sales drive down the prices available on the market and the depositors are the losers. Once again, though, central bank intervention in the form of a monetary injection allows the financial system to share risks without incurring the costs of inefficient investment.

4. Market provision of liquidity

It is worth dwelling on the role of asset markets in some detail, since it has broad methodological implications for the analysis of crises. As the discussion of LTCM illustrates, the provision of liquidity to the market plays an important role in the analysis of financial fragility.

In both DD and AG, assets are represented by investment technologies. The short-term (liquid) asset is represented by a storage technology and the long-term (illiquid) asset is represented by a two-period investment technology. In the DD model, the possibility of premature liquidation of the long-term assets is also represented by a technology. If the long-term asset is liquidated prematurely, it yields a return of only $r<1$ per unit invested. The difference in returns, $R-r$, is the cost of liquidation. The costly liquidation technology reflects the assumption that, when financial institutions have to realise the value of their assets in a hurry, they are typically unable to realise the full value that they would receive if they could wait until maturity. This loss of value is one of the costs of financial distress. However, the use of a reduced-form "liquidation technology" obscures a number of interesting features that are highly relevant for understanding the welfare economics of financial crises.

It is easy to see why the introduction of an asset market might amplify the effects of a bank run. By making all assets liquid, the new market allows the bank to liquidate all its assets in an attempt to meet the claims of the early withdrawers. Now the banks may be forced to liquidate their previously illiquid assets in order to meet their deposit liabilities. However, by selling assets during a run, they force down the price and make the crisis worse. This destroys the equilibrating mechanism of the earlier versions of the model in which the returns to the illiquid asset were untouchable at date 1.

Liquidation is obviously self-defeating, in the sense that it transfers value from depositors to the speculators in the market. A transfer is not inefficient and it might be thought that, unlike in DD, the premature liquidation does not involve a deadweight loss. The welfare analysis of the market's impact is a bit more subtle, however. The deadweight loss associated with liquidations takes the form of sub-optimal risk-sharing, not a loss of value per se. Optimal insurance would provide depositors with a positive transfer in bad states, where asset returns are low, and impose a tax or negative transfer in good states, where returns are high. The asset market does the opposite. By making transfers in the worst states, the market provides depositors with negative insurance.

In this case, there is an incentive for the central bank to intervene to prevent a collapse of asset prices, but again the problem is not runs per se but the unnecessary liquidations they promote. Another solution, explored in Allen and Gale (2000a), is the provision of optimal liquidity insurance through the market. Liquidity insurance takes the form of Arrow securities in theory and of credit derivatives in practice. If insurance markets are complete, banks can insure against runs and crises and once again achieve optimal risk-sharing. This is not to say that complete insurance eliminates crises - it may be socially optimal to have crises because of the flexibility default introduces into risk-sharing contracts - but simply that the market will determine the optimal incidence of financial crises. Conversely, incompleteness of insurance markets may provide a rationale for central bank intervention.

The role of liquidity in determining asset prices is explored in a different context by Allen and Gale (1994). However, the same feature that assets have to be sold at a loss in some states occurs there. When a liquidity shock hits the market, forcing some investors to sell assets quickly, there are two possible regimes. If the amount of liquidity in the market, as measured by holdings of liquid assets, is high, then the asset price is determined by the expected future returns to the asset. On the other hand, if the amount of liquidity in the market is low, then the price is determined by the amount of "cash in the market". Of course, the amount of liquidity is itself endogenous, and results from prior decisions by investors. Liquidity providers need a profit to induce them to participate in the market for assets. Speculators have an incentive to hold liquid assets in order to buy up assets only if the price is low enough. So, in some states, the market has to be illiquid and there has to be "cash-in-the-market" pricing.

In summary, modelling the provision of liquidity by the market instead of assuming banks have a costly "liquidation technology" is a methodological innovation in several respects:

- First, the cost of liquidation is now endogenous. Whether there is a loss of value in selling assets in the intermediate period is determined by the liquidity of the market, that is, by the portfolio choices of the investors and institutions that make up the market.
- Ex post, there is no deadweight loss from selling assets. An asset sale involves a transfer, but the asset's returns are not affected by the sale. This is a major change from the DD

model and its successors, in which the returns of the liquidated asset are determined by the technology and assumed to be lower than the asset's returns at maturity.

- Ex ante, liquidation may impose a cost. While the seller's "loss" is the buyer's "gain" ex post, they are both losers ex ante. Liquidation is inefficient ex ante because it does not provide the bank with insurance against changes in the asset price. The bank obtains a good price in states where the demand for liquidity is low and a bad price in other states where the demand for liquidity is high.
- The market's provision of liquidity is necessarily inadequate. Because the return on holding the short asset is lower than the return on holding the long asset, investors require some additional incentive for providing liquidity. They obtain this incentive in the form of a capital gain if they can buy the long asset cheaply in the middle period and realise a high return in the last period. This will happen only if there is a distress sale from which they can profit. In other words, the market will be willing to provide liquidity to a distressed institution only if the terms are sufficiently profitable and this means that the assets have to be sold "at a loss". Thus, the amount of liquidity provided in equilibrium will never be adequate to support asset prices at a level that would give optimal risk-sharing for banks.

5. Sunspots

Theoretical research on speculative currency attacks, banking panics, and contagion have taken a number of approaches. One is built on the foundations provided by early research on bank runs (eg Allen and Gale (1998; 2000a-d), Chang and Velasco (2000; 2001)) and Peck and Shell (1999)). Other approaches include macroeconomic models of currency crises that developed from the insights of Krugman (1979), Obstfeld (1986) and Calvo (1988) (see, for example, Corsetti et al (1999) for a recent contribution and Flood and Marion (1999) for a survey), game theoretic models (see Morris and Shin (1998; 2000) and Morris (2000) for an overview), amplification mechanisms (eg Cole and Kehoe (2000) and Chari and Kehoe (2000)) and the borrowing of foreign currency by firms (eg Aghion et al (2001)).

Two main perspectives on financial crises can be discerned in this literature. One is that they are *random events*, unrelated to changes in the real economy. The classical form of this view suggests that crises are the result of "mob psychology" or "mass hysteria" (see, for example, Kindleberger (1978)). The modern version, developed by DD and others, is that bank runs are self-fulfilling prophecies. As we saw in Section 2, there are two equilibria in the DD model, one with runs and one without. Which of these two equilibria occurs depends on extraneous variables or "sunspots". Although sunspots have no effect on the real data of the economy, they affect depositors' beliefs in a way that turns out to be self-fulfilling. (Postlewaite and Vives (1987) have shown how runs can be generated in a model with a unique equilibrium.)

The alternative to the sunspot view, discussed in Section 3, is that financial crises are a natural outgrowth of the *business cycle*. An economic downturn will reduce the value of bank assets, raising the possibility that banks are unable to meet their commitments. If depositors receive information about an impending downturn in the cycle, they will anticipate financial difficulties in the banking sector and try to withdraw their funds. This attempt will precipitate the crisis. According to this interpretation, panics are not random events but a response to unfolding economic circumstances. Mitchell (1941), for example, writes (p 74):

> "when prosperity merges into crisis ... heavy failures are likely to occur, and no one can tell what enterprises will be crippled by them. The one certainty is that the banks holding the paper of bankrupt firms will suffer delay and perhaps a serious loss on collection."

In other words, panics are an integral part of the business cycle.

Whichever view one takes of the causes of financial crises, there is some consensus based on historical experience that financial systems can be *fragile*. The threat of a financial crisis lies in the possibility that it will propagate through the economic system, causing damage disproportionate to the original shock. This notion of financial fragility is most easily seen in the sunspot model: the impact of extraneous uncertainty is equivalent to financial fragility, since the shock that "causes" the crisis has no effect on the fundamentals of the economy. Financial fragility can also be captured in a real business cycle model, where crises result from exogenous shocks. In this context, financial fragility is interpreted as a situation in which very small shocks can tip the economy over the edge into a full blown crisis. In other words, financial fragility is an extreme case of excess sensitivity to small shocks.

In terms of causation, the difference between sunspots (sensitivity to extraneous uncertainty) and excess sensitivity (extreme sensitivity to real exogenous shocks) is not great. The first could be thought of as a limiting case of the second. However, there are important differences between the two approaches. The sunspot theory does not predict crises; it allows for the possibility of crises. Furthermore, the sunspot theory also depends on fundamentals. Weak fundamentals are not sufficient for a crisis, but in the presence of weak fundamentals, self-fulfilling expectations may be sufficient for a crisis.

An approach that spans both the real business cycle approach and the sunspot theory is represented by AG, who call a crisis *essential* if, for certain parameter values, every equilibrium of the model is characterised by a crisis. Restricting attention to situations in which crises are essential gives the theory greater predictive power. These models may allow for sunspot equilibria, but do not rely on them.

A related approach is represented by the work of Morris and Shin (1998; 2000) and Morris (2000), who study models with multiple equilibria but use equilibrium selection arguments based on small amounts of asymmetric information about parameter values to predict which equilibrium will be chosen.

Using a special case of the framework developed in Allen and Gale (2000), Allen and Gale (2001) investigate the connection between financial fragility and the existence of sunspot equilibria. The connection is close. Financial fragility can occur when the spillover effect from liquidation of assets by banks is channelled to other banks through the price of assets in the market. What is crucial for understanding this phenomenon is the fact that the system minimises liquidity to be the minimum needed for preventing a crisis in certain states. If the demand for liquidity rises above this level, there will be a sharp fall in the price of assets. This drop in asset prices may force other banks into insolvency and exacerbate the crisis. The pecuniary externalities, to use the technical term, from one set of agents forces another much larger set into bankruptcy. In other words, a small shock (to liquidity demand) can have a large effect. In the limit, when the initial shock that causes the crisis becomes vanishingly small, we have something that looks very much like a sunspot equilibrium. However, the approach is different, since it does not rely on multiple equilibria.

The reason for financial fragility is the necessity for providing incentives to hold liquidity. It seems possible that as in AG the central bank can eliminate the inefficiencies associated with crises by an appropriate injection of money. This is an important topic for future research.

6. Contagion

The AG framework has also been used to construct a model in which small shocks lead to large effects by means of contagion - more precisely, in which a shock within a single sector can spread to other sectors and lead to an economy-wide financial crisis. Allen and Gale (2000b) construct a model in which, under certain circumstances, contagion is unavoidable when the economy is subject to a small shock.

The economy consists of a number of regions. The number of early and late consumers in each region fluctuates randomly, but the aggregate demand for liquidity is constant. This allows for inter-regional insurance as regions with liquidity surpluses provide liquidity for regions with liquidity shortages. One way to organise the provision of insurance is through an interbank market in deposits. Suppose that region *A* has a large number of early consumers when region *B* has a low number of early consumers, and vice versa. Since regions *A* and *B* are otherwise identical, their deposits are perfect substitutes. The banks exchange deposits at the first date before they observe the liquidity shocks. If region *A* has a higher than average number of early consumers at date 1, then banks in region *A* can meet their obligations by liquidating some of their deposits in the banks of region *B*. Region *B* is happy to oblige, because it has an excess supply of liquidity in the form of the short asset. At the final date the process is reversed, as banks in region *B* liquidate the deposits they hold in region *A* to meet the above average demand from late consumers in region *B*.

Inter-regional cross-holdings of deposits work well as long as there is enough liquidity in the banking system as a whole. If there is an excess demand for liquidity, however, the financial linkages caused by these cross-holdings can turn out to be a disaster. While cross-holdings of deposits are useful for reallocating liquidity within the banking system, they cannot increase the total amount of liquidity. If the economy-wide demand from consumers is greater than the stock of the short asset, the only way to provide more consumption is to liquidate the long asset. This is very costly (see Shleifer and Vishny (1992) and Allen and Gale (1998) for a discussion of the costs of premature liquidation), so banks try to avoid liquidating the long asset whenever possible. In this case, they can avoid liquidating the long asset by liquidating their claims on other regions instead. This mutual liquidation of claims does not create any additional liquidity, however. It merely denies liquidity to the troubled region and bank runs and bankruptcy may be the result. What begins as a financial crisis in one region can then spread by contagion to other regions because of the cross-holdings of deposits.

The interbank market works quite differently from the retail market. In the latter case, runs occur because deposit contracts commit banks to a fixed payment and banks must begin liquidating the long asset when they cannot meet liquidity demand from the short asset. In the interbank market the initial problem is caused by the fact that banks with an excess demand for liquidity cannot obtain anything from banks in other regions. This is the opposite of the problem in the retail market and, unlike there, cannot be solved by making the contracts discretionary or contingent since whatever their form they cancel each other out. Instead of being caused by the nature of interbank claims, spillovers and contagion result just from the fall in the value of bank assets in adjacent regions.

Whether the financial crisis does spread depends crucially on the pattern of interconnectedness generated by the cross-holdings of deposits. If the interbank market is *complete* and each region is connected to all the other regions, the initial impact of a financial crisis in one region may be attenuated. On the other hand, if the interbank market is *incomplete*, each region is connected with a small number of other regions. The initial impact of the financial crisis may be felt very strongly in those neighbouring regions, with the result that they too succumb to a crisis. As each region is affected by the crisis, it prompts premature liquidation of the long asset, with a consequent loss of value, so that previously unaffected regions find that they too are affected because their claims on the region in crisis have fallen in value.

It is important to note the role of the free rider problem in explaining the difference between a complete and incomplete interbank market. There is a natural pecking order among different sources for liquidity. A bank will meet withdrawals first from the short asset, then from holdings in other regions, and only in the last resort will it choose to liquidate the long asset. Cross-holdings are useful for redistributing liquidity, but they do not create liquidity; so when there is a global shortage of liquidity (withdrawals exceed short assets), the only solution is to liquidate long assets. If every region takes a small hit (liquidates a small amount of the long asset), there may be no need for a global crisis. This is what happens with complete markets: banks in the troubled region have direct claims on banks in every other region and there is no way to avoid paying one's share. With incomplete markets, banks in the troubled region have a direct claim only on the banks in adjacent regions. The banks in other regions pursue their own interests and refuse to liquidate the long asset until they find themselves on the front line of the contagion.

The notion of a region is intended as a metaphor for categories of banks that may differ in several dimensions. For example, some banks may be better at raising funds while other banks are better at lending them. Or it might be that banks focus on lending to different industries or in different regions and as a result have lending opportunities that are not perfectly co-related with their deposit base. In either case, an interbank market plays an important role in redistributing the funds efficiently. However, the existence of claims between different categories of banks opens up the possibility of contagion when one category is hit by a sudden demand for liquidity.

The reason that contagion can occur here is the existence of incomplete markets. The central bank can play an important role here by completing markets. If it is linked to all the banks, then it can overcome the free rider problem and simply reallocate liquidity to prevent the contagion.

7. Policy conclusions

The citing of "financial fragility" and "contagion" is often the rationale for intervention in the financial system by central banks and governments. Traditionally, the justification for intervention was based on historical evidence. The memory of the Great Depression and earlier crises is still with us and it powerfully reinforces the belief that such intervention is worthwhile. Until recently, there has been little attempt to try and understand these phenomena at a theoretical level. Although the state of the theory is too underdeveloped to allow for strong policy conclusions, some simple lessons can be drawn from the work reviewed here.

- In the first place, a micro-based theoretical analysis allows us to address *normative* questions about financial crises - for example, when are they consistent with optimal risk-sharing? - in addition to *positive* questions about what causes crises and how they can be prevented. Once the focus is on the welfare economics, we are led to think about the *optimality* of financial crises rather than mere crisis avoidance.
- A second lesson is that, in these models, the cost of financial crises comes from inefficient asset liquidation rather than the crisis per se. This may be because there is a real loss of asset value, as in DD, or because liquidation is associated with inefficient risk-sharing, as in AG. In either case, the policy recommendation is to avoid inefficient liquidation rather than prevent crises at all costs.
- There are several ways of avoiding the costs associated with inefficient liquidation. One is to substitute money for real claims, as in AG; another is to provide complete insurance through the market, as in Allen and Gale (2000a); another is to provide liquidity through the lender of last resort (LOLR), as in Bhattacharya and Gale (1987).
- Finally, we have seen that, under certain conditions, the laissez faire outcome is incentive-efficient or constrained-efficient, in which case there is no role for the LOLR. On the other hand, various frictions and imperfections give rise to the possibility that efficiency requires some intervention by the LOLR. For example, if insurance markets are incomplete, there may be a role for the central bank as a substitute for incomplete markets.

Our discussion has focused on financial issues, narrowly defined, and in particular on optimal risk-sharing. But it also has to be recognised that disruption of the financial sector has implications for the "real" sector (cf Bernanke and Gertler (1989)). The concern about financial fragility arises precisely because of the fear that what begins as a purely financial disturbance may spill over into the rest of the economy and cause a period of slow growth or even a recession. We have not discussed these issues explicitly, but we have examined models in which small shocks can have far-reaching consequences in the financial sector. We presume that when these disturbances do impose costs on the rest of the economy, there is a rationale for central bank intervention to prevent asset price volatility and bank defaults before they wreak havoc elsewhere.

Again, we have seen that these crises arise from mispricing of liquidity and/or lack of liquidity. For example, in AG, provision of liquidity by the market requires price volatility. The low return on liquid assets means that there must be states of the world where these can be used to make a profit. In Allen and Gale (2000b), a small shock in one region or sector can spread by contagion and cause a meltdown in the financial system if markets are incomplete. The discontinuity associated with bankruptcy means that even a small shock can have a large effect if it cascades sequentially through the financial system. In each case, liquidity provision by the LOLR appears to be the key to prevention.

Intervention by the central bank to provide liquidity is not the only way to deal with crises. Bank regulation such as capital controls is another instrument that can potentially be used to intervene. Bankruptcy law is another type of policy that is potentially important in controlling the effects of crises. In the models discussed, the bankruptcy law is such that banks must liquidate their assets to meet their obligations. Alternative laws that do not have this requirement but delay claims may be helpful in eliminating financial fragility and contagion. Much work remains to be done in the area of public policy and crises.

References

Aghion, P, P Bacchetta and A Banerjee (2001): "Currency crises and monetary policy in an economy with credit constraints", *European Economic Review* 45, 1121-50.

Allen, F and D Gale (1994): "Limited market participation and volatility of asset prices", *American Economic Review* 84, 933-55.

—— (1998): "Optimal financial crises", *Journal of Finance* 53, 1245-84.

—— (2000a): "Banking and markets", Working Paper 00-44B, Wharton Financial Institutions Center, University of Pennsylvania.

—— (2000b): "Financial contagion", *Journal of Political Economy* 108, 1-33.

—— (2000c): "Optimal currency crises", *Carnegie Rochester Series on Public Policy* 53, 177-230.

—— (2000d): "Bubbles and crises", *The Economic Journal* 110 (2000), 236-56.

—— (2001): "Financial fragility", Working Paper 01-37, Wharton Financial Institutions Center, University of Pennsylvania.

Allen, F and R Herring (2001): "Banking regulation versus securities market regulation", Working Paper 01-29, Wharton Financial Institutions Center, University of Pennsylvania.

Alonso, I (1996): "On avoiding bank runs", *Journal of Monetary Economics* 37, 73-87.

Bhattacharya, S and D Gale (1987): "Preference shocks, liquidity and central bank policy", in W Barnett and K Singleton (eds), *New approaches to monetary economics*, Cambridge University Press, New York.

Bernanke, B and M Gertler (1989): "Agency costs, net worth, and business fluctuations", *American Economic Review* 79, 14-31.

Bryant, J (1980): "A model of reserves, bank runs, and deposit insurance", *Journal of Banking and Finance* 4, 335-44.

Calomiris, C and G Gorton (1991): "The origins of banking panics, models, facts, and bank regulation", in Glenn Hubbard (ed), *Financial markets and financial crises*, University of Chicago Press, Chicago, IL.

Calomiris, C and C Kahn (1991): "The role of demandable debt in structuring optimal banking arrangements", *American Economic Review* 81, 497-513.

Calvo, G (1988): "Servicing the public debt: the role of expectations", *American Economic Review* 78, 1411-28.

Champ, B, B Smith and S Williamson (1996): "Currency elasticity and banking panics: theory and evidence", *Canadian Journal of Economics* 29, 828-64.

Chang, R and A Velasco (2000): "Financial fragility and the exchange rate regime", *Journal of Economic Theory* 92, 1-34.

—— (2001): "A model of financial crises in emerging markets", *Quarterly Journal of Economics* 116, 489-517.

Chari, V (1989): "Banking without deposit insurance or bank panics: lessons from a model of the US national banking system", *Federal Reserve Bank of Minneapolis Quarterly Review* 13 (Summer), 3-19.

Chari, V and R Jagannathan (1988): "Banking panics, information, and rational expectations equilibrium", *Journal of Finance* 43, 749-60.

Chari, V and P Kehoe (2000): "Financial crises as herds", working paper, Federal Reserve Bank of Minneapolis.

Cole, H and T Kehoe (2000): "Self-fulfilling debt crises", *Review of Economic Studies* 67, 91-116.

Corsetti, G, P Pesenti and N Roubini (1999): "Paper tigers? A model of the Asian crisis", *European Economic Review* 43, 1211-36.

Diamond, D and P Dybvig (1983): "Bank runs, deposit insurance, and liquidity", *Journal of Political Economy* 91, 401-19.

Flood, R and N Marion (1999): "Perspectives on recent currency crises literature", *International Journal of Finance and Economics* 4, 1-26.

Gorton, G (1988): "Banking panics and business cycles", *Oxford Economic Papers* 40, 751-81.

Hellwig, M (1994): "Liquidity provision, banking, and the allocation of interest rate risk", *European Economic Review* 38, 1363-89.

Jacklin, C (1987): "Demand deposits, trading restrictions, and risk-sharing", in E Prescott and N Wallace (eds), *Contractual arrangements for intertemporal trade*, University of Minnesota Press, Minneapolis, MN.

Jacklin, C and S Bhattacharya (1988): "Distinguishing panics and information-based bank runs: welfare and policy implications", *Journal of Political Economy* 96, 568-92.

Kaminsky, G and S Schmukler (1999): "What triggers market jitters? A chronicle of the Asian crisis", *Journal of International Money and Finance* 18, 537-60.

Kindleberger, C (1978): *Manias, panics, and crashes: a history of financial crises*, Basic Books, New York, NY.

Krugman, P (1979): "A model of balance-of-payments crises", Journal of Money Credit and Banking 11, 311-25.

Mitchell, W (1941): *Business cycles and their causes*, University of California Press, Berkeley, CA.

Morris, S (2000): "Contagion", *Review of Economic Studies* 67, 57-78.

Morris, S and H S Shin (1998): "Unique equilibrium in a model of self-fulfilling currency attacks", *American Economic Review* 88, 587-97.

—— (2000): "Global games: theory and applications", working paper, Yale University.

Obstfeld, M (1986): "Rational and self-fulfilling balance of payments crises", *American Economic Review* 76, 72-81.

Peck, J and K Shell (1999): "Bank portfolio restrictions and equilibrium bank runs", working paper, Cornell University.

Postlewaite, A and X Vives (1987): "Bank runs as an equilibrium phenomenon", *Journal of Political Economy* 95, 485-91.

Scholes, M (1972): "The market for securities: substitution versus price pressure and the effects of information on share price", *Journal of Business* 45, 179-211.

Shleifer, A and R Vishny (1992): "Liquidation values and debt capacity: a market equilibrium approach", *Journal of Finance* 47: 1343-66.

Sprague, O (1910): *A history of crises under the national banking system*, US Government Printing Office, Washington DC.

Wallace, N (1988): "Another attempt to explain an illiquid banking system: the Diamond and Dybvig model with sequential service taken seriously", *Federal Reserve Bank of Minneapolis Quarterly Review* 12 (Fall), 3-16.

—— (1990): "A banking model in which partial suspension is best", *Federal Reserve Bank of Minneapolis Quarterly Review* 14 (Fall), 11-23.

Implications of the bank merger wave for competition and stability

Elena Carletti, Mannheim University
Philipp Hartmann, European Central Bank and CEPR
Giancarlo Spagnolo, Mannheim University and CEPR

Abstract

This paper discusses the effects of bank consolidation on competition and stability in the banking sector. Most empirical literature seems to point towards the standard adverse effects on prices of increased concentration in banking. A major issue is the still regional character of loan and deposit markets for households and small enterprises, which contrasts with the generally increasing globalisation of other financial services. In line with other recent papers, we challenge the view that market power - as may be created through banking consolidation - is unambiguously good for banking system stability. Various features of bank mergers may actually increase the scope for instability, in particular when they lead to a small number of large "national champions", monitoring problems, lower money market liquidity or organisational inefficiencies/lack of market discipline. Overall, we stress that competition considerations need to receive adequate attention, even in the special banking sector.

1. Introduction

The "merger movement" in banking has been widely documented and debated in policy reports and research papers (see eg Boyd and Graham (1991, 1996); Berger, Kashyap and Scalise (1995); Berger, Demsetz and Strahan (1999); Dermine (2000); ECB (2000); OECD (2000); Group of Ten (2001)). While significant consolidation also took place among other financial service providers, the phenomenon was particularly concentrated among banking firms. Bank consolidation accelerated during the last three years of the 1990s and most importantly the largest number of mergers and acquisitions in this sector occurred within national borders.[1] As a consequence, several industrial countries reached a situation of high banking sector concentration or faced a further deterioration of an already concentrated sector (eg Australia, Belgium, Canada, France, the Netherlands and Sweden), while the banking sectors of a few other countries remained relatively unconcentrated (this group includes for example Germany and the United States; see Group of Ten (2001) for details).

Apart from general management objectives, such as increasing profitability by diversification and exploitation of economies of scale, dominating markets and governing larger firms, the origins of this merger wave were found in technical progress (particularly in communication technology), deregulation, European economic and monetary union, general globalisation and the resulting competitive challenges for financial firms. Such an extensive concentration process is of interest for various policy areas, including competition policies to ensure market discipline and the efficient functioning of the financial sector, prudential policies to maintain its stability, and monetary policies, regarding both bank sector liquidity management in the implementation of monetary policy and the monetary transmission mechanism.

In the present paper we discuss the implications of bank mergers and banking sector concentration for both competition and stability. Section 2 focuses on the intensity of competition in the banking sector, while Section 3 addresses the link between this and bank stability/systemic risk.[2] We review the

[1] In this paper we will not address the differences between mergers and acquisitions and often refer to both as mergers.

[2] The competition-stability nexus has recently also been discussed by Canoy et al (2001), Carletti and Hartmann (forthcoming) and Vives (2001). In a direct policy context it was addressed by the Cruickshank (1999) interim report in the United Kingdom.

related empirical literature and derive the main conclusions at the end of each section. The last section presents the start of a new line of research that models the joint consequences of consolidation for bank competition and interbank market liquidity fluctuations. This research and further variations of it have the potential to provide input in the discussion on the implications of consolidation on monetary, competition and supervisory policies and their relations to each other.

One main conclusion from the present paper is that market power and competition need to be carefully addressed in the banking sector, despite or even because of its special character in relation to financial stability.

2. Competition effects of bank mergers

A good deal of the debate on competition effects from bank consolidation has been phrased in terms of the conflict between two competing hypotheses or paradigms. The structure-conduct-performance (SCP) paradigm, going back to Mason (1939) and Bain (1956), highlights reductions in competition and increases in market power through firm growth and concentration. In contrast, the efficient-structure (ES) paradigm, related to Demsetz (1973) and Peltzman (1977), rather emphasises that differences in market shares/concentration reflect superior efficiency of growing firms.

The SCP and ES paradigms are also reflected in the more recent theoretical literature on the effects of in-market mergers on prices and quantities under imperfect competition (see, for example, Perry and Porter (1985) and Farrell and Shapiro (1990)). The main idea is that a merger has two effects: first, it enlarges the market share of the merged firms (and thereby enhances their market power); second, it may lead to efficiency gains in terms of a reduction in the costs of the merged firms. The first effect leads to upward pressure on prices. Since each firm involved in the merger internalises the effect of a change in its price on the demand of all other merged firms, it charges a higher markup than before the merger. The second effect tends to reduce prices. If lower costs materialise, then the merged firms become more aggressive and reduce prices in order to enlarge their customer base. Thus, whether a merger leads to price increases (and consequently reductions in quantities) depends on the relative importance of the internalisation effect (increase in market power) and the potential efficiency gains.

These standard results in industrial organisations apply of course also to banking markets. Therefore, if the SCP effects of bank mergers dominated, then bank consolidation should be associated with increasing loan rates and/or decreasing deposit rates (together with decreasing supply), as firms try to exploit market power to increase their profits. If the ES effects dominated, then the opposite should happen, since expanding firms would pass efficiency gains on to customers.

Note that under the antitrust practice followed in most countries the two paradigms lead to opposite policy conclusions. Since competition authorities tend to focus on prices, they would control consolidation that goes beyond a certain point when SCP effects dominate. This would not be the case when ES effects dominate. Now, focusing on prices alone in competition reviews of mergers may be regarded as suboptimal, since it implies that only *consumer surplus* is maximised by the authorities and increases in profits that may lead to higher *total surplus* are ignored. However, Neven and Roeller (2000) recently provided a clear rationale behind the current practice. They show in a political economy framework that the merging firms (here banks) are typically in a better position than their dispersed customers (here depositors and borrowers) to lobby and influence the decision of the antitrust agency. An exclusive consumer surplus objective corrects this imbalance. Therefore in this paper we do not question standard antitrust practice and focus on bank loan rate increases or deposit rate decreases and - to a lesser extent - on quantity reductions as indicators of adverse effects from mergers on competition.

2.1 The effects of bank mergers on small business and consumer loan markets

Quantitative empirical research on the relationship between market structures and loan rates seems to go back to the 1960s, when provisions in the Bank Holding Company Act of 1956 and the Bank

Merger Acts of 1960 and 1966 for the first time required supervisory authorities in the United States to also preserve competition in banking.[3] This implied that they generally had to review bank mergers from a competition perspective. In response to these developments, in 1962 the Board of Governors of the Federal Reserve System launched a comprehensive research programme on bank market structures and competition.

In this environment a “banking competition controversy” unfolded, as witnessed for example by the two conflicting research papers by Edwards (1965) and Flechsig (1965) as well as numerous follow-up papers, including Kaufman (1966), Phillips (1967), Taylor (1968) or Bell and Murphy (1969). Some of the authors followed Edwards’ conclusion that concentration increased loan rates, while others followed Flechsig’s conclusion that this relationship was not robust. Excellent summaries of the early literature of the 1960s and 1970s are provided by Gilbert (1984) and Weiss (1989), who conclude more or less that most of the better executed studies point to some adverse effect of concentration (as measured in this early literature in deposit markets) on loan rates.[4]

2.1.1 *Recent evidence on loan rates*

Most studies for the United States show that loan market *concentration* increases small business and consumer lending rates, in line with increased market power of the lenders.[5] Hannan (1991), Berger and Hannan (1997) and Hannan (1997) show this for various cross sections of small secured and unsecured business loans. Kahn et al (2001) also find this for personal loans, but not for automobile loans (which are often collateralised). One European study confirms the market power hypothesis at least for customer and mortgage loans of euro area banks (Corvoisier and Gropp (2001)), whereas a Swiss study on mortgage loans yields inconclusive results (Egli and Rime (1999)).

As to the effects of bank *mergers*, Akhavein et al (1997) find only insignificant changes in loan income of banks involved in 57 US “megamergers”. Kahn et al (2001) detect personal loan rate increases but automobile loan rate decreases from US mergers. For Europe, Sapienza (2002) shows in a very careful study of the Italian banking sector (combining information about lenders with information about borrowers for the first time) that only the largest mergers increased credit line rates, whereas smaller ones were associated with cheaper credit lines (indicating that efficiency gains could offset market power effects in those cases). A study for Spain yields inconclusive results in the mortgage market (Fuentes and Sastre (1998)). The papers that have some dynamic dimension indicate that adverse competition effects of bank mergers take time to materialise, often half a year or more after the operation.

2.1.2 *Recent evidence on quantities lent to small businesses*

Apart from pricing considerations, the bank merger wave raised concerns in the United States that banking consolidation would reduce the amount of credit available to small businesses. This argument was based on the observation that small banks mainly make small loans (since they do not have large enough balance sheets for more sizeable loans often required by larger businesses), assumed to go to small firms, and that large banks tend to lend to large businesses (as the monitoring costs of many small companies would be too high for them).[6] Another concern could be that larger banks would exploit their greater market power to reduce lending (and increase loan rates). This, it was feared by some, would lead to inefficient credit supply, hurting particularly the emergence of small startup firms. However, reductions in lending could of course also be the consequence of the elimination of previously inefficient loans, ie those funding negative net present value projects.

---

3 Strictly speaking the application of competition laws to the banking sector in the United States was only made explicit with the Philadelphia National Bank case in 1963 and with the subsequent amendment of the Bank Merger Act in 1966.

4 In this paper we look at bank market concentration in general and at bank mergers specifically. As a caveat it should be kept in mind that concentration may also be caused by other developments, for example voluntary market exits or failures.

5 Concentration is most often measured with the Herfindahl-Hirschman index, which is defined as the sum of the squared market shares of all active banks in a given market. Occasionally, it is also measured as the joint market share of the three or five largest lenders.

6 See eg Berger et al (1995, Table A.10) for detailed data about the size distribution of loans by small, medium-sized and large banks.

This static view has been challenged from various perspectives. Some authors question the assumption that all *merged banks* lend less to small businesses. For example, Strahan and Weston (1996) find that when small US banks merged in the mid-1990s, their post-merger small business lending was actually higher than before. For mergers among larger banks changes were insignificant. In contrast, Peek and Rosengren (1996) document for a small cross section in the New England area during 1993-94 that when a large bank takes over a small one, the small business lending by the target is lower than before the merger (and only a small part of this effect is offset by new entrants in the local market).

In a later study with broader US data the same authors show a more complex relationship between bank mergers and small business lending (Peek and Rosengren (1998)). They find that the acquirers tend to partially recast the targets on their own image, causing the small business lending share of the merged institution to move towards the acquirers' previous share. Whereas the balance of post-merger portfolio adjustments seems to indicate a higher likelihood of somewhat reduced small business lending, they conclude that the initial concerns seem to have been overstated. Strahan and Weston (1998) point out that not combining different banks into the full holding company may result in biases because of intracompany transactions. For a data set that combines banks in such a way they find similar results to their earlier paper, in that for mergers involving small acquirers and targets small business lending actually increases, whereas the effects of consolidation tend to be insignificant when intermediate or large banks merge. Their interpretation of the results is that lending diversification is important for the smaller players, and organisational diseconomies less so.

In Europe the few available papers point to the traditional concern about reduced small business lending through consolidation. For Italy, Sapienza (2002) shows that merged banks are less likely to extend a credit line to a small business than before merging. And Karceski et al (2000), who use Norwegian data, argue that (mostly small) mergers increase bank relationship exit rates. Borrowers from merger targets also suffer from (weakly significant) negative abnormal stock market returns after the transaction.

Another group of authors argues that merged banks reduce small business lending, but that this effect is offset by incumbent rival banks expanding their loans or de novo entry in the same local market. For example, Berger et al (1998) detect in a large data set that US mergers significantly increase small business loans by competitor banks. Goldberg and White (1998) consider the fact that the late 1980s and early 1990s saw a large number of new bank charters, in parallel with the merger wave, and estimate that de novo banks have a significantly larger share of small business loans on their balance sheets than comparable incumbents. In another long and broad data set Berger et al (1999) combine these two facts and find that mergers in local markets significantly increase the likelihood of new entrants in that market and that the new players have a larger share of small loans in their portfolio than incumbent banks. (However, Seelig and Critchfield (1999) find exactly the opposite, for a shorter and narrower data set.)

In a new line of research, Berger et al (2001) argue that small business lending can be heavily influenced by *market size structure*. Surprisingly, their data show that in markets with a higher share of large banks small businesses have a higher likelihood of receiving a credit line, and even at lower interest rates, than in markets composed of smaller banks. (However, larger borrowers are still more likely to go to larger banks.) They explain (part of) the difference to the previous literature with the fact that they can directly observe the size of the borrower (in a way similar to Sapienza for Italy) and do not have to approximate it by the size of loans. Apart from the two papers mentioned above, we could not find any other research on the small business lending issue with European (or Japanese) data (see also Dermine (2000)).

2.1.3 Summary and conclusions

In sum, the available research literature seems to suggest that increasing bank market concentration and consolidation tend to drive loan rates up in many local markets. This finding is in line with the SCP paradigm, according to which concentration leads to market power. The fact that sometimes loan rate increases are not quantitatively large may either be explained by successful bank merger reviews, stopping or amending those that risk creating institutions with stronger market power, or by remaining efficiency gains from mergers (not controlled for in the estimations) that partly offset rate increases.

Regarding the effect of consolidation on quantities, available literature seems to indicate that early concerns about collapsing small business loan supply seem to have been overstated, since dynamic competitive forces lead at least in part to the replacement of lending lost.

However, it should also be noted that research outside the United States remains relatively limited and less clear-cut. For example, in Europe it is of utmost importance that euro area or even EU-wide bank and firm micro data on local loan (and deposit) markets be collected in a broad and systematic way, covering all countries. Such data would allow researchers to undertake homogeneous cross-country analyses of competitive conditions in EU banking markets, comparable to a long tradition in the United States. They would put various area-wide policy areas on a much safer information basis than has been the case so far. Overall, the evidence available to date makes a case in favour of the systematic application of competition reviews in the banking sector.

2.2 The effects of bank mergers on retail deposit markets

The issue of concentration in deposit markets has recently received considerable attention in Europe through a report by the Competition Commission (2002) in the United Kingdom. This voluminous report on "The supply of banking services by clearing banks to small and medium-sized enterprises" highlighted in particular the "significant market concentration … in the markets for liquidity management services, 90 per cent or more of such services being supplied by four clearing groups in each geographical market".[7] The report concluded that "the restriction and distortion in price competition ... has led to excessive prices and profits" and that the situation constituted "a complex monopoly situation". Although shying away from structural measures, such as the divestment of bank branches, it recommended some behavioural measures, including minimum interest rates to be paid by the banks in England and Wales.

Turning back to research results, studies of the effects of concentration and consolidation on bank retail deposit markets to a large extent mirror the broad results found for small business and consumer loan markets, although they seem to have started much later. A larger number of papers using different US data sets find a statistically significant negative relation between market *concentration* and various customer deposit rates (such as those for money market deposit accounts (MMDAs), short-term certificates of deposit (CDs) or negotiated order of withdrawal accounts (NOWs)). These papers include Berger and Hannan (1989a,b), Calem and Carlino (1991) and Neumark and Sharpe (1992). Berger and Hannan (1997) estimate that this relationship continues to hold when one controls for changes in cost efficiency.

There also seems to be some time variation in the statistical significance of the relationship, in that it sometimes becomes quite weak (see eg Berger and Hannan (1992), or Hannan (1997)). Radecki (1998) argues that more recently this may be related to the fact that the borders of US retail deposit markets have expanded from Metropolitan Statistical Areas (MSAs; normally used in previous studies as the relevant market) to States, due to deregulation and the (internal) reorganisation of bank holding companies. He detects stronger relationships between concentration and deposit rates at State level.[8]

For euro area countries, Corvoisier and Gropp (2001) confirm the inverse relation between concentration and deposit rates for time deposits, but not for demand deposits, where paradoxically it is reversed. As with Egli and Rime (1999) for Switzerland, they find only mixed results for euro area savings deposit markets.

Regarding the effects of *mergers* on deposit rates, the analyses by Akhavein et al (1997) and by Praeger and Hannan (1998) suggest that only the larger in-market mergers have statistically significant adverse effects on more local MMDA and NOW rates, but not on three-month CD rates.[9] However, Simons and Stavins (1998) for the United States and Focarelli and Panetta (2002) for Italy point out in two more dynamic analyses that the largest deposit rate reductions happen in the first years after the operation and that in later years the rates come up again. This is explained with the fact that the necessary restructurings of merged banks to achieve cost efficiency gains can often take

[7] The three geographical markets identified were (1) England and Wales, (2) Scotland and (3) Northern Ireland. Liquidity management services include business current accounts, overdraft facilities and short-term bank deposit accounts.

[8] Berger et al (1999) discuss whether the negative relationship between market concentration and deposit rates weakened in the 1990s as compared to the 1980s. However, the papers reviewed do not allow for a clear conclusion in this regard.

[9] The results from Fuentes and Sastre (1998) for Spain are inconclusive.

years. Both papers find that competitor banks of merger parties in the same market consistently reduce deposit rates though, even in the long run.

The *conclusion* for retail deposit markets is then quite similar to the one for small business and consumer loans. The ES hypothesis only receives occasional support. Since there is evidence that consolidation can lead to increased market power, vigorous antitrust reviews in banking seem highly advisable to avoid consumers and small businesses paying too high loan rates, receiving too low deposit rates or receiving unsatisfactory service. However, could the limitation of profits through controls of market power have adverse effects on banking system stability? We address this question in the next section. As a final note, it appears that also for deposit market analyses there is an urgent need for carefully raised cross-country micro data sets in the euro area or even the European Union.

3. Stability effects of bank mergers

It has been argued in the literature that the erosion of market power is a source of banking instability (see eg Marcus (1984)). These arguments would suggest a more cautionary approach in competition policy, to avoid conflicts with supervisory policy. Carletti and Hartmann (forthcoming, Section 3) show that all G7 countries and all EU countries give a strong role to supervisory authorities in the review of bank mergers. In some countries the authority in charge of prudential supervision has a much stronger responsibility than the regular antitrust authority, or in one or two even has all the competence.

In this light, it is somewhat surprising that the number of research papers explicitly addressing the link between bank consolidation and stability is still relatively limited. A good deal of the debate was kicked off by the empirical work by Keeley (1990), who argued that the erosion of bank market power (as measured by a decline in banks' market-to-book asset ratio, Tobin's *q*) led to a higher risk premium that banks had to pay on certificates of deposit and in lower capital-to-asset ratios in the United States during the 1980s. The implied trade-off between the intensity of competition in the banking sector and its safety became known under the term "charter value hypothesis".[10]

3.1 Bank mergers and risk diversification

Other studies addressed reverse causation, namely whether bank mergers - which as shown in Section 2 often cause some increases in market power in loan and deposit markets - were associated with lower bank risk. Craig and Santos (1997) find the risk reduction effect confirmed (as measured by the *z*-score statistic of default probability and by stock return volatility) and relate it to benefits from diversification.[11] Benston et al (1995) argue on the basis of pre-merger earnings volatility and target-acquirer correlation that the motivation for mergers in the first half of the 1980s must have been risk reduction through diversification, rather than the exploitation of the put option on deposit insurance funds.

In a similar vein, Hughes et al (1999) simulate different consolidation strategies from structural bank holding company relationships estimated with 1994 data. They find that interstate expansion in the United States should lead to insolvency risk reductions, in particular when diversifying macroeconomic risks. The more recent paper by Amihud et al (forthcoming) addresses the issue for cross-border mergers covering many countries. Their result is that international mergers between 1985 and 1998 had no systematic effects on acquiring banks' total relative or systematic stock price risk. One interpretation of this result is that diversification benefits are offset by particular monitoring problems associated with foreign operations. However, as a cautionary note it should be recalled that cross-border and interstate mergers (almost by definition out-of-market mergers) have less potential to restrict competition than the in-market mergers discussed in the previous section.

---

10 "Charter value" denotes the present value of future monopoly rents from holding a bank charter.

11 The *z*-score used in this paper is a statistic derived from historical profits, equity and asset stocks measuring the number of standard deviations below the mean that a bank's profits would have to fall before its equity became negative. See Goodhart et al (1998, p 90) for a brief summary of credit scoring techniques more generally.

3.2 Bank size and risk-taking

Yet another group of papers checks whether larger banks actually fail less often than smaller banks or whether they take on new risks after diversification. For example, Chong (1991) undertakes an event study and finds that US interstate consolidation increases bank stock return volatility. Boyd and Runkle (1993) point out that the reductions in stock price volatility in their data (related to potential diversification benefits) do not translate into significant reductions in the failure probability of large banks. They find only insignificantly lower *z*-scores.[12] On the basis of realised bank failure rates Boyd and Graham (1991, 1996) document that on average large banks in the United States failed more often than small banks during the 1970s and the first half of the 1980s but not during the late 1980s/early 1990s. They explain the fact that better diversification of larger banks does not reduce failure risk systematically with their greater tendency to leverage, potentially as a consequence of an implicit too-big-to-fail protection.

Demsetz and Strahan (1995, 1997) argue that in line with diversification larger banks have lower stock return volatility if their portfolios are held constant. But when, for example, loan portfolios are allowed to vary, risk is no longer reduced. In other words, large banks benefit from their better risk-return trade-off by expanding risky loans and reducing equity ratios. Similarly, Hughes et al (1996) and Hughes and Mester (1998) argue that increased risk-taking by growing banks may be a reflection of the efficient exploitation of scale economies. If size increases go hand in hand with better risk diversification, then the implied lower average and marginal costs of risk management will naturally lead them to take on more risk.

De Nicolo (2000) reasserts with similar estimations to Boyd and Runkle for more recent (1988-98) and broader data that *z*-score failure probabilities increase with size not only for US banks but also for European and Japanese banks. As additional explanations to the ones put forward above, he also finds that state ownership has a positive impact on failure risk of banks and discusses recent theoretical literature arguing that size-related diversification does not necessarily reduce bank insolvency risk (Hellwig, 1998).

Finally, a background paper to the Ferguson Report (Group of Ten (2001)) by de Nicolo and Kwast (2001) observes that the market share of large and complex banking organisations (LCBOs) in the United States increased during the 1990s and that the increases in market shares were highly correlated with similarly increasing LCBO stock return correlations. The authors argue that this may be an indication of heightened systemic risk in the banking sector. Note that similar to the bank size and risk literature this is inconsistent with the typical “charter value” prediction of an inverse relationship between market power/concentration and risk.

3.3 Summary and conclusions

In sum, on the basis of this literature one cannot ascertain a clear-cut relation between the effects of consolidation and bank or systemic risk. Some studies suggest that a more consolidated banking sector would be more stable (in particular if concentration creates market power that avoids incentives for excessive risk-taking and if size brings about diversification gains which are not offset by the adoption of new risks) and other studies suggest the opposite (in particular if consolidation worsens too-big-to-fail problems, complicates monitoring in agency problems, is related to organisational diseconomies and reduces the costs of risk management). More research is certainly necessary to understand under which conditions which sign of the relationship applies. The last section discusses one possible avenue for such work. In any case, the available empirical literature does not contain a strong argument in favour of generally constraining competition, encouraging in-market consolidation or discouraging out-of-market consolidation as means to foster the stability of the banking system. Hence, given the risks to market efficiency discussed in Section 2, the conclusion that thorough competition reviews of bank mergers are necessary remains valid.

[12] Note, however, that Boyd and Runkle (1993) also find that greater size (among US bank holdings) is associated with unchanged or lower “charter value”, as measured by Tobin’s *q*. So we cannot assume that size in this study is related to market structure or market power in an unambiguous way.

4. Further research avenues

A cornerstone of a stable banking system is a robust and liquid interbank money market. The money market is particularly important since it links large banks to each other, so that a problem in this market may have widespread consequences. Recent theoretical literature has modelled the scope for contagion (Rochet and Tirole (1996); Allen and Gale (2000); or Freixas et al (2000)) and adverse selection in interbank markets (Flannery (1996)). However, such efforts have not yet incorporated the implications of bank mergers for the functioning of the money market. Nor have they modelled the structure and competitive pressures of banking markets, which - as discussed in Section 3 - may influence the risk of bank activities. Hellman et al (2000), Matutes and Vives (2000) and Cordella and Yeyati (2002) analyse the link between competition for deposits and individual banks' incentives for risk-taking on the asset side, while Perotti and Suarez (2001) examine the effects of active merger policy and temporary entry restrictions for bank stability in a dynamic duopolistic model where banks compete in deposits. None of these papers, however, addresses how competition affects banks' liquidity management and the functioning of the interbank market.[13]

Work in this direction has been started by Carletti et al (2002). The model addresses the consequences of consolidation for loan rates, reserve holdings and interbank market liquidity fluctuations. Following traditional banking theory, the model features stochastic withdrawal shocks on deposits, which banks can finance either with reserves or by interbank market borrowing. Less traditionally, it features competition in a differentiated oligopolistic loan market. When liquidity shocks are uncorrelated across merging banks, a merger creates an internal money market, saving interbank borrowing costs for the two institutions. Surprisingly, for most parameter configurations this internalisation effect dominates the diversification of liquidity risk, so that merged banks increase reserve holdings. As a consequence of the internal money market, they also enjoy lower liquidity risk and expect lower liquidity needs than competitor banks. Hence, regarding individual bank liquidity risk the effect of consolidation goes in the same direction as the one derived by the risk diversification literature described at the start of Section 3, although for different reasons.

As to the loan market, merged banks gain market power but also enjoy cost advantages through lower refinancing costs and potentially also through efficiency gains. Loan rates increase when the market power effects are stronger. So the competition model can accommodate both the SCP hypothesis (when market power effects dominate) and the ES hypothesis (when cost saving effects dominate), as described in Section 2.

Finally, aggregate bank system liquidity improves through higher reserve holdings and deteriorates through an asymmetry in deposit bases induced by loan competition. Hence, with uncorrelated shocks the aggregate liquidity effects of a merger are ambiguous, whereas with correlated shocks they are unambiguously negative. The latter effect illustrates the possibility that significant bank consolidation can make liquidity fluctuations in the interbank money market more violent and therefore, ceteris paribus, impair financial stability.

This finding provides a theoretical foundation for the statement in the G10 Report on Consolidation in the Financial Sector that "... by internalising what had previously been interbank transactions, consolidation could reduce the liquidity of the market for central bank reserves, making it less efficient in reallocating balances across institutions and increasing market volatility" (Group of Ten (2001), p 20). Now, the confirmation that such an effect is possible is first of all of historical value. In the absence of a central bank the more violent liquidity fluctuations will occasionally lead to liquidity crises, since the amount of available reserves is limited in the short term, even for very high money market rates. However, in the Carletti et al model, as in modern central banking practice, any missing liquidity can be provided elastically by the central bank in order to prevent the money market rate from deviating from the policy interest rate or in an extreme situation to avoid a liquidity crisis.[14]

[13] For a more comprehensive survey of the small theoretical literature on bank market structure and risk, see Carletti and Hartmann (forthcoming, Section 4.1).

[14] The central banks contributing to the G10 report did not see any evidence so far that financial sector consolidation had led to impairing money market liquidity. However, they agreed that the situation should be monitored carefully.

Although nowadays central banks have the relevant instruments available to keep the liquidity situation in the money market stable, the model conveys two lessons: (1) If there was no central bank or if the central bank could not perfectly anticipate the right amount of liquidity needed, then it cannot be excluded that liquidity crises may sometimes occur in the money market. The model shows how their frequency may vary as a function of bank consolidation. (2) In the presence of a central bank, the model informs about how liquidity management may have to change with significant bank consolidation.[15] For example, in the case of correlated deposit shocks across merging banks the average amount of liquidity to be provided by the central bank to stabilise the money market rate or to prevent a liquidity crisis in case of a shortage is larger after consolidation than before. However, the model also shows that there are plausible situations (under uncorrelated deposit shocks) in which consolidation leads to an improvement of the liquidity situation in the money market (contrary to the concern raised in the Ferguson Report, Group of Ten (2001)).

As already mentioned, aggregate liquidity fluctuations in the money market can sometimes impair overall financial stability. Therefore, the paper also has something to say about the controversial relationship between competition and stability in banking. Concretely, it describes different scenarios for this relationship. In one scenario mergers lead to more market power in the loan market (SCP effects dominate ES effects) and to more violent liquidity fluctuations in the money market. In this case, the negative relationship between competition and stability in banking - as claimed by the "charter value" literature - does not hold. Both competition and stability have worsened. Moreover, it is interesting to observe that the adverse aggregate liquidity effects of the merger are a function of the competitiveness of the loan market before the merger. The larger the number of banks and the more substitutable loans are, the less severe the adverse liquidity effects of the merger. In other words, in this relatively plausible scenario more competition is actually good for interbank market stability.

In other scenarios consolidation causes improvements in competition (ES effects dominate market power effects) and either also improvements in money market liquidity or a deterioration of money market liquidity. However, the empirical evidence provided in Section 2 indicates that in practice this may be a less frequent set of cases. Finally, the scenario in which market power increases and liquidity improves is also possible under certain parameter configurations in the model. The multiplicity of possible scenarios is not too surprising, given the heterogeneous results found in the empirical literature discussed in Section 3.

The results are also instructive regarding the relationship between antitrust and supervisory authorities in the review of bank mergers. In the cases where competition worsens and interbank stability improves or where competition improves and interbank stability worsens a policy conflict can emerge between the two types of authorities. Solving the trade-off would require some coordination, either directly between the two authorities or through a third, potentially higher authority. The latter is, for example, the case in Canada, where the Minister of Finance decides on bank mergers on the basis of two reports, one from the competition authority and the other from the supervisor. In the United States, this task is fulfilled by the courts. There are also countries in which supervisors have the competence to decide on their own. (See Carletti and Hartmann, forthcoming, for descriptions of these arrangements in G7 countries and the European Union.)

Finally, from the perspective of monetary policy implementation careful monitoring of consolidation tendencies is justified as well. Changes in aggregate liquidity risk, as described by the model, may affect the aggregate liquidity management by the central bank. How important such effects can become is an empirical question, which will inter alia depend on the importance of bank consolidation, as compared to the size of the money market.

[15] In all bank theories in which there is only liquidity risk, ie shocks do not adversely affect asset values, the introduction of a central bank that can provide unlimited amounts of liquidity removes the occurrence of liquidity crises. This feature is not specific to the present model.

References

Akhavein, J D, A N Berger and D B Humphrey (1997): "The effects of megamergers on efficiency and prices: evidence from a bank profit function", *Review of Industrial Organization*, 12, 95-139.

Allen, F and D Gale (2000): "Financial contagion", *Journal of Political Economy*, 108(1), 1-29.

Allen, F, H Gersbach, J-P Krahnen and A M Santomero (2001): "Competition among banks: introduction and conference overview", *European Finance Review*, 5, 1-11.

Amihud, Y, G L de Long, A Saunders (forthcoming): "The effects of cross-border bank mergers on bank risk and value", forthcoming *Journal of International Money and Finance*.

Bain, J S (1956): "Relation of profit rate to industry concentration", *Quarterly Journal of Economics*, 65, 293-324.

Bell, F and N Murphy (1969): "Impact of market structure on the price of a commercial bank's service", *Review of Economics and Statistics*, 51, 210-13.

Benston, G J, W C Hunter and L D Wall (1995): "Motivations for bank mergers and acquisitions: enhancing the deposit insurance put option versus earnings diversification", *Journal of Money, Credit, and Banking*, 27(3), 777-88.

Berger, A N and T H Hannan (1989a): "The price-concentration relationship in banking", *Review of Economics and Statistics*, 71, 291-99.

——— (1989b): "Deposit interest rates and local market concentration", in L W Weiss (ed), *Concentration and price*, Cambridge, MA, MIT Press, 255-65.

——— (1992): "The price-concentration relationship in banking: a reply", *Review of Economics and Statistics*, 74, 376-79.

——— (1997): "Using efficiency measures to distinguish among alternative explanations of the structure-performance relationship in banking", *Managerial Finance*, 23(1), 6-31.

Berger, A N, A K Kashyap and J M Scalise (1995): "The transformation of the US banking industry: what a long, strange trip it's been", *Brookings Papers on Economic Activity*, 2, 55-201.

Berger, A N, A Saunders, J M Scalise and G F Udell (1998): "The effects of bank mergers and acquisitions on small business lending", *Journal of Financial Economics*, 50, 187-229.

Berger, A N, R S Demsetz and P E Strahan (1999): "The consolidation of the financial services industry: causes, consequences, and implications for the future", *Journal of Banking and Finance*, 23, 135-94.

Berger, A N, S D Bonime, L G Goldberg and L J White (1999): "The dynamics of market entry: the effects of mergers and acquisitions on de novo entry and small business lending in the banking industry", paper presented at the Center for Financial Studies/Wharton Financial Institutions Center conference on "Competition among banks: good or bad?", Frankfurt, 7-8 April 2000.

Berger, A N, R J Rosen and G F Udell (2001): "The effect of market size structure on competition: the case of small business lending", *FRB Chicago Working Paper*, no 2001-10, Federal Reserve Bank of Chicago, IL, October.

Boyd, J and S Graham (1991): "Investigating the banking consolidation trend", *Federal Reserve Bank of Minneapolis Quarterly Review*, Spring, 1-15.

——— (1996): "Consolidation in US banking: implications for efficiency and risk", Federal Reserve Bank of Minneapolis Working Paper, no 572, December.

Boyd, J H and D E Runkle (1993): "Size and performance of banking firms: testing the predictions of theory", *Journal of Monetary Economics*, 31(1), 47-67.

Calem, P S and G A Carlino (1991): "The concentration/conduct relationship in bank deposit markets", *Review of Economics and Statistics*, 73(2), 268-76.

Canoy, M, M van Dijk, J Lemmen, R de Mooij and J Weigand (2001): "Competition and stability in banking", CBP Document, no 015, Netherlands Bureau for Economic Policy Analysis, December.

Carletti, E (1998): “Competition, stability and regulation”, traduzione italiana in M Polo (ed), *Industria Bancaria e Concorrenza*, Bologna, Il Mulino, 67-136.

Carletti, E and P Hartmann (forthcoming): “Competition and stability: what’s special about banking?”, forthcoming in P Mizen (ed), *Monetary history, exchange rates and financial markets: essays in honour of Charles Goodhart Vol 2*, Cheltenham, Edward Elgar. (ECB Working Paper, no 146, May 2002.)

Carletti, E, P Hartmann and G Spagnolo (2002): “Bank mergers, competition and financial stability”, paper presented at the third joint central bank research conference on “Risk measurement and systemic risk”, Basel, 7-8 March.

Chong, B S (1991): “The effects of interstate banking on commercial banks’ risk and profitability”, *Review of Economics and Statistics*, 73, 78-84.

Competition Commission (2002): *The supply of banking services by clearing banks to small and medium-sized enterprises*, report presented to Parliament by the Secretary of State for Trade and Industry and the Chancellor of the Exchequer by Command of Her Majesty, March.

Cordella, T and E L Yeyati (2002): “Financial opening, deposit insurance and risk in a model of banking competition”, *European Economic Review*, 46(3), 471-85.

Corvoisier, S and R Gropp (2001): “Bank concentration and retail interest rates”, ECB Working Paper, no 72, July.

Craig, B and J C Santos (1997): “The risk effects of bank acquisitions”, *Federal Reserve Bank of Cleveland Economic Review*, 33, 25-35.

Cruickshank, D (1999): *Review of banking services in the UK; competition and regulation: an interim report to the Chancellor of the Exchequer*, The Stationery Office, London, July.

De Bandt, O and P Hartmann (2002): “Systemic risk in banking: a survey”, in C A E Goodhart and G Illing (eds), *Financial crises, contagion, and the lender of last resort - a reader*, London, Oxford University Press, 249-98.

Demsetz, H (1973): “Industry structure, market rivalry, and public policy”, *Journal of Law and Economics*, 16, 1-9.

Demsetz, R and P E Strahan (1995): “Historical patterns and recent changes in the relationship between bank holding company size and risk”, *Federal Reserve Bank of New York Economic Policy Review*, July, 13-26.

——— (1997): “Diversification, size, and risk at bank holding companies”, *Journal of Money, Credit, and Banking*, 29(3), 300-13.

De Nicolo, G (2000): “Size, charter value and risk in banking: an international perspective”, International Finance Discussion Paper, no 689, Board of Governors of the Federal Reserve System, December.

De Nicolo, G and M Kwast (2001): “Systemic risk and financial consolidation: are they related?”, Finance and Economics Discussion Paper, no 2001-33, Board of Governors of the Federal Reserve System, August.

Dermine, J (2000): “Bank mergers in Europe: the public policy issues”, *Journal of Common Market Studies*, 38(3), 409-25.

Edwards, F R (1964): “Concentration in banking and its effect on business loans”, *Review of Economics and Statistics*, 46, 294-300.

——— (1965): “The banking competition controversy”, *National Banking Review*, 3, September, 1-34.

Edwards, F R and F Mishkin (1995): “The decline of traditional banking: implications for financial stability and regulatory policy”, *Federal Reserve Bank of New York Economic Policy Review*, 1, 27-45.

Egli, D and B Rime (1999): “The UBS-SBC merger and competition in the Swiss retail banking sector”, *Studienzentrum Gerzensee Working Paper*, no 00/02, December.

European Central Bank (2000): *Mergers and acquisitions involving the EU banking industry - facts and implications*, Frankfurt, December.

Farrell, J and C Shapiro (1990): "Horizontal mergers: an equilibrium analysis", *American Economic Review*, 80(1), 107-26.

Flannery, M (1996): "Financial crises, payments system problems, and discount window lending", *Journal of Money, Credit, and Banking*, 28(4), 804-24.

Flechsig, T (1965): "The effect of concentration on bank loan rates", *Journal of Finance*, 20, 298-311.

Focarelli, D and F Panetta (2002): Are mergers beneficial to consumers? Evidence from the Italian market for bank deposits, paper presented at the Federal Reserve Bank of Chicago conference on "Bank structure and competition", Chicago, IL, 8-10 May.

Freixas, X, B Parigi and J-C Rochet (2000): "Systemic risk, interbank relations and liquidity provision by the central bank", *Journal of Money, Credit, and Banking*, 32(3), 611-38.

Fuentes, I and T Sastre (1998): "Implications of restructuring in the banking industry: the case of Spain", *BIS Conference Papers*, no 7, 98-120.

Gilbert, R A (1984): "Bank market structure and competition: a survey", *Journal of Money, Credit, and Banking*, 16(4), 617-45.

Goldberg, L G and L J White (1998): "De novo banks and lending to small businesses: an empirical analysis", *Journal of Banking and Finance*, 22, 851-67.

Goodhart, C, P Hartmann, D Llewellyn, L Rojas-Suarez and S Weisbrod (1998): *Financial regulation: why, how and where now?*, London, Routledge.

Group of Ten (2001): *Report on consolidation in the financial sector*, BIS, IMF, OECD.

Hannan, T H (1991): "Bank commercial loan markets and the role of market structure: evidence from surveys of commercial loans", *Journal of Banking and Finance*, 15, 133-49.

——— (1997): "Market share inequality, the number of competitors, and the HHI: an examination of bank pricing", *Review of Industrial Organization*, 12, 23-35.

Hellman, T F, K Murdock and J Stiglitz (2000): "Liberalization, moral hazard in banking and prudential regulation: are capital requirements enough?", *American Economic Review*, 90(1), 147-65.

Hellwig, M (1998): "Allowing for risk choices in Diamond's financial intermediation as delegated monitoring", University of Mannheim Working Paper (SFB 504), no 98-04.

Hughes, J P and L J Mester (1998): "Bank capitalization and cost: evidence of scale economies in risk management and signalling", *Review of Economics and Statistics*, 80, 314-25.

Hughes, J P, W W Lang, L J Mester, C-G Moon (1996): "Efficient banking under interstate branching", *Journal of Money, Credit, and Banking*, 28, 1045-71.

——— (1999): "The dollars and the senses of bank consolidation", *Journal of Banking and Finance*, 23(2-4), 291-324.

Jackson, W E (1992): "The price-concentration relationship in banking: a comment", *Review of Economics and Statistics*, 74(2), 373-76.

Kahn, C, G Pennacchi and B Sopranzetti (2001): "Bank consolidation and consumer loan interest rates", *Financial Institution Center Working Paper*, no 01-14, Wharton School, March.

Karceski, J, S Ongena and D C Smith (2000): "The impact of bank consolidation on commercial borrower welfare", International Finance Discussion Paper, no 679, Board of Governors of the Federal Reserve System, Washington, DC, August.

Kaufman, G G (1966): "Bank market structure and performance: the evidence from Iowa", *Southern Economic Journal*, 32, 429-39.

Keeley, M (1990): "Deposit insurance, risk and market power in banking", *American Economic Review*, 80, 1183-1200.

Marcus, A J (1984): "Deregulation and bank financial policy", *Journal of Banking and Finance*, 8, 557-65.

Mason, E S (1939): "Price and production policies of large-scale enterprise", *American Economic Review*, 29, 61-74.

Matutes, C and X Vives (2000): “Imperfect competition, risk taking and regulation in banking”, *European Economic Review*, 44(1), 1-34.

Neumark, D and S A Sharpe (1992): “Market structure and the nature of price rigidity: evidence from the market for consumer deposits”, *Quarterly Journal of Economics*, 107, 657-80.

Neven, D J and L H Roeller (2000): “Consumers surplus vs welfare standard in a political economy model of merger control”, *WZB Discussion Paper*, IV 00-15, Berlin, October.

OECD (2000): *Mergers in financial services*, DAFFE/CLP Report, (2000)17, Paris, September.

Paroush, J (1995): “The effects of mergers and acquisition activity on the safety and soundness of a banking system”, *Review of Industrial Organization*, 10, 53-67.

Peek, J and E S Rosengren (1996): “Small business credit availability: how important is size of lender?”, in A Saunders and I Walter (eds), *Universal banking: financial system design reconsidered*, Burr Ridge, IL, Irwin Publishing.

——— (1998): “Bank consolidation and small business lending: it’s not just bank size that matters”, *Journal of Banking and Finance*, 22, 799-819.

Peltzman, S (1977): “The gains and losses from industrial concentration”, *Journal of Law and Economics*, 20, 229-63.

Perotti, E and J Suarez (2001): “Last bank standing: what do I gain if you fail?”, *CEPR Discussion Paper*, no 2933, August.

Perry, M and R Porter (1985): “Oligopoly and the incentive for horizontal merger”, *American Economic Review*, 75(1), 219-27.

Phillips, A (1967): “Evidence on concentration in banking markets and interest rates”, *Federal Reserve Bulletin*, 6(June), 916-26.

Praeger, R A and T H Hannan (1998): “Do substantial horizontal mergers generate significant price effects? Evidence from the banking industry”, *Journal of Industrial Economics*, 46(4), 433-52.

Radecki, L J (1998): “The expanding geographic reach of retail banking markets”, *Federal Reserve Bank of New York Economic Policy Review*, June, 15-34.

Rochet, J-C and J Tirole (1996): “Interbank lending and systemic risk”, *Journal of Money, Credit, and Banking*, 28(4), 733-62.

Sapienza, P (2002): “The effects of bank mergers on loan contracts”, *Journal of Finance*, 57(1), 329-67.

Seelig, S and T Critchfield (1999): “Determinants of de novo entry in banking”, *FDIC Working Paper*, no 99-1, Federal Deposit Insurance Corporation, Washington, DC.

Simons, K and J Stavins (1998): “Has antitrust policy in banking become obsolete?”, *New England Economic Review*, March/April, 13-26.

Strahan, P E and J P Weston (1996): “Small business lending and bank consolidation: is there cause for concern?”, *Current Issues in Economics and Finance*, 2(3), 1-6.

——— (1998): “Small business lending and the changing structure of the banking industry”, *Journal of Banking and Finance*, 22, 821-45.

Taylor, C (1968): “Average interest charges, the loan mix, and measures of competition: sixth federal reserve district experience”, *Journal of Finance*, 23, 793-804.

Vives, X (2001): “Competition in the changing world of banking”, *Oxford Review of Economic Policy*, 17(4), 535-47.

Weiss, L W (1989): “Banking”, in L W Weiss (ed), *Concentration and price*, Cambridge, MA, MIT Press, 219-54.

Financial crises and incomplete information

Mariassunta Giannetti[1]

Abstract

This article presents a review of Giannetti's (2002a,b) models arguing that incomplete information may be more relevant than moral hazard in explaining banking crises and episodes of overlending. It argues that overlending problems are not necessarily due to investor moral hazard and guarantees on deposits. Instead, guarantees on deposits may even limit the losses accumulated by the banking system. In fact, if international investors have incomplete information on the average quality of the investment opportunities available in a country and firms are financed by a main bank, a soft budget constraint distortion arises, because of capital inflows. The model shows that in equilibrium international investors rationally do not require any risk premium until a substantial amount of losses has been accumulated, even if there are no guarantees on deposits. Bond market development, by increasing the number of lenders, can eliminate the soft budget constraint distortion and prevent banking crises.

1. Introduction

Financial crises are generally thought to be caused either by liquidity problems, due to coordination problems among depositors, or by moral hazard. According to a strand of the literature (Radelet and Sachs (1998)), banks would fund profitable but illiquid projects: if agents panic and withdraw their deposits before the projects are completed, banks default. In contrast, according to the theories based on moral hazard (McKinnon and Pill (1996), Krugman (1998), Corsetti et al (1999)), banks fund insolvent projects because of corruption, looting and connected lending. International investors, who generally channel their funds through the local banks, would not exert any discipline by not making deposits in insolvent banks, because they expect the value of their deposits to be guaranteed by the government.

Unfortunately, none of these explanations consider the specific nature of financial markets and bank-firm relationships in emerging economies. Moreover, a mantained assumption of both classes of models is that international investors have complete information on the growth prospects and the banking system of the economies where they invest. In particular, it is commonly assumed that they can observe the quality of the projects banks fund.

In fact, these assumptions are rarely satisfied. Investors are uncertain about the origins of growth of an economy that may grow because of excessive investment in low productivity projects as well as the availability of good investment opportunities. Since also for economists it is an arduous task to measure total factor productivity and the determinants of growth, it is sensible to assume that also international investors are imperfectly informed about the determinants of growth in a country and ultimately about the aggregate productivity of the projects funded by the banking system.

At the microeconomic level, the pervasive lack of transparency of financial systems based on close bank-firm relationships suggests that it is more realistic to assume that international investors who make deposits in domestic banks, and to a large extent also domestic depositors, are imperfectly informed about the solvency of individual banks in a country.

In a series of papers, Giannetti (2002a,b) has taken seriously the implications of investors' incomplete information to analyse the determinants and the dynamics of financial crises. In her models, capital

[1] Assistant Professor, Department of Finance and SITE, Stockholm School of Economics, Sveavägen 65, Box 6501, S-113 83, Stockholm, Sweden. E-mail: mariassunta.giannetti@hhs.se; phone: +46-8-7369607; fax: +46-8-316422

inflows are demand-determined and international investors channel their funds through local banks and have incomplete information about the quality of projects funded by the bank where they have made deposits. They have prior beliefs about the probability that some banks are insolvent and renewing bad loans, instead of declaring default. In the model, this can happen because of a soft budget constraint or outright looting. However, investors attribute a positive probability to the fact that the growth in bank loans is due to the availability of profitable projects.

In this setting, investors update their beliefs on the probability that banks will be able to repay deposits at the end of each period. The risk of not being able to withdraw the deposits remains low as long as banks have an incentive to renew loans to non-profitable projects, because deposits can be withdrawn at any moment and there are other investors willing to provide the bank with funds if the interest rate is high enough.

It is possible to show that banks have an incentive not to renew loans and to declare default only if the risk premium required by international investors is sufficiently high. In equilibrium, this happens only if the expected level of the aggregate losses accumulated by the banking system is large enough.

In this context, if investors are not perfectly informed, it is possible to explain overlending without moral hazard. Moreover, it is straightforward to explain contagion. Since investors cannot distinguish across banks they can suddenly demand a high risk premium also from banks that are perfectly solvent, but illiquid. The increase in the interest rate burden may also drive illiquid banks into insolvency and cause widespread banking crises.

Some characteristics of the financial system are key for explaining the propensity to banking crises of emerging economies. Obviously, in this context, banking crises and problems of excessive lending would not arise if the banking system were transparent: investors would not make deposits in insolvent banks. Moreover, they could distinguish between insolvent and illiquid banks and this would make it possible to avoid contagion. On the other hand, the dearth of funds to intermediate, due to limited domestic saving before the liberalisation of capital inflows, puts a constraint on credit expansion and ensures a precarious financial stability. After the liberalisation of capital inflows, instead, international investors can provide any amount of funds the banking system demands, as long as they receive the same expected return they would have on similar international assets. Consequently banks have the possibility to fund and renew any amount of loans they wish.

Most importantly, it is possible to show that close bank-firm relationships, with a main bank providing the bulk of funds for a project, are at the origin of banks' incentives to renew loans to projects that are unprofitable. Therefore, banking crises are expected to be less likely in financial systems where firms have a multiplicity of lenders.

In what follows, I describe a few situations of financial crises in emerging markets that can be easily explained by the nature of bank-firm relationships and the lack of transparency. I refer the reader to Giannetti (2002a,b) for details on the models. The concluding section elaborates some policy implications based on the theoretical analysis mentioned above.

2. Stylised facts

Several regularities observed in a number of banking crises suggest the importance of underdeveloped financial markets, incomplete information and the lack of a variety of lenders in explaining financial instability. In what follows, I analyse the experiences of Chile in 1982, the Nordic countries (Finland and Sweden) in 1991-92, and East Asian economies in 1997 to evidence these common features.

2.1 Main banks

The absence of a variety of financial markets and the shortage of lenders are common features of emerging markets. Corporations are highly dependent on borrowing from financial institutions and, as is common in countries with bank-based financial systems, rely heavily on debt financing. This aspect is very important for the financial stability of the banking system, as the solvability of highly indebted lenders is easily undermined by changes in the cost of funds and a reversal in capital flows. Furthermore, there are close relationships between banks and firms and loan exposures are highly concentrated. This in turn provides incentives also to renew loans to insolvent projects, if the availability of funds allows increasing credit. Although not efficient, the financial system appears stable before the liberalisation of capital movements. Most importantly, in banking systems based on close bank-firm relationships less information is generally available to outside investors.

The empirical evidence corroborates the assumptions of the model on bank-firm relationships.

For instance, in Chile before the 1982 crisis, the grupos (large financial and manufacturing conglomerates) were highly dependent on bank loans and very often the financing bank itself belonged to the conglomerates (Velasco (1991)). Dependence on bank loans was high in Nordic countries as well. In 1980, the debt/equity ratios were about 3 and 4 in Finland and Sweden, respectively, compared to less than 1/5 in the United Kingdom and 1/4 in the United States. Moreover, most commercial banks had highly concentrated loan exposures, mostly to connected non-financial corporations. Relationship banking was also dominant in East Asian economies. In South Korea, for instance, bank loans were the main source of credit and there was a particular form of bank-enterprise relationship that linked each large business group, the chaebol, to a main bank, the so-called principal transactions bank (Nam (1996)). Amazingly, just a few years ago, these relationship-based financial systems were extolled for allowing financiers to take a longer view on investment and they were credited with the remarkably good economic performance of the East Asian economies (Rajan and Zingales (1998)). Their weaknesses became clear in 1997.

2.2 Large availability of funds

Banking crises follow the lifting of restrictions on capital movements, which allows banks to acquire funds abroad. These new funding opportunities, made possible by large capital inflows, permit greater credit expansion than domestic retail deposits. As a consequence, non-profitable projects are financed. As the first signs of banks' fragility become evident, capital inflows revert and the banking crisis begins. Although the financial systems of the economies that experience financial crises seemed relatively stable when capital inflows were restricted, the lifting of these restrictions coincides with the beginning of a lending boom, backed by an accumulation of foreign liabilities by domestic banks and apparently irrational lending policies. This is due to the fact that, when capital inflows are restricted, the amount of domestic savings imposes a cap on the amount of loans the banking system can extend. This dearth of funds gives banks an incentive to be more selective and a credible commitment not to provide working capital to insolvent projects, just to postpone the official recognition of the losses. The large availability of funds before the banking crisis is also a well-documented empirical fact. The 1982 Chilean crisis followed the financial liberalisation of the late 1970s and was preceded by massive capital inflows mainly in the form of short-term bank liabilities (see Table 1 for the data). The expansion of bank liabilities had as a counterpart an increase in bank loans that may have in fact acted as a pull factor for capital inflows. As Velasco (1991) notes in analysing the origins of the crisis:

> "Perhaps, the single most important factor behind the growth of domestic indebtedness was the rolling over of credits and the capitalization of interest…Furthermore, the line between a performing and a nonperforming asset becomes fuzzy when rollovers and capitalization of interest are widely used to keep many problem loans on the books."

By 1982, this provoked a massive increase in non-performing assets and loan defaults that required government interventions. Due to the rapid expansion of net domestic credit to rescue financial institutions, the fixed exchange rate collapsed in June 1982. The events surrounding the 1994 crisis in Mexico were very similar; the crisis was preceded by a credit boom and a large increase in non-performing loans, as noted by Edwards and Végh (1997).

The origin of the banking and balance of payments crises in the Nordic countries also seems to rely on the accumulation of losses by the banking system; here, the lifting of restrictions on capital movements in the 1980s allowed banks to obtain funds abroad to finance their rapid credit expansion. As a consequence, the ratio of bank loans to nominal GDP increased to 90% in 1990 from 55% in 1984 in Finland, while it increased to 58% from 41% in Sweden. Banks' difficulties became evident in 1991, when several banks were bailed out by the government and the central bank had to provide liquidity.

The balance of payments crisis hit these economies the following year in conjunction with the EMS crisis.[2]

The experiences of Korea, Thailand and Indonesia during the 1997 Asian turmoil are the most recent examples of crises driven by an accumulation of bank losses. Consider once again South Korea. In the years preceding the 1997 crisis but following the opening of the financial markets in the second half of the 1980s, South Korea also experienced a pronounced increase in external borrowing by domestic banks, which in turn lent these funds to the private sector. The data in Table 2 show large growth rates of lending to the private sector, which averaged almost 17% annually in the 1990s; this is well in excess of the average growth rate of per capita GDP, which was about 7% annually. As a result, at the end of 1996 the ratio of short-term external liabilities of BIS reporting banks to foreign reserves was 213%. The structural weaknesses of the Korean banking system became increasingly apparent during 1997. In particular, the large exposures of banks to the highly leveraged conglomerates and the huge amount of impaired loans became evident when six chaebols failed. Moreover, investors discovered that the average debt/equity ratio of the top 30 conglomerates was over 500% and that most of the loans were in effect without collateral, since group firms used cross-payment guarantees to facilitate borrowing. In order to increase the confidence of international financial markets, the government announced guarantees on the foreign liabilities of Korean financial institutions. The Bank of Korea provided liquidity and, in December, it was forced to allow the won to float freely. Investors and lenders panicked when they learned that the country's short-term external debt was approximately $104 billion (rather than the $66 billion originally reported) and that usable reserves were lower than expected. As a consequence, the Korean banks' short-term external liabilities fell dramatically, because of capital outflows, and the currency depreciated by 39%.

The sequences of events were similar in Thailand and Indonesia, which also experienced lending booms fuelled by capital inflows in the years preceding the crises, as is evident from Tables 2 and 3.

In all these episodes banks appear to have renewed their loans to insolvent firms. Why are there incentives for banks to overlend after the liberalisation of financial markets? Giannetti (2002a,b) argues that the lifting of restrictions on capital movements causes a soft budget constraint problem because a massive amount of capital becomes available at low cost in the early phase of the financial liberalisation. The Ponzi scheme only ends when the cost of funds rises because of the incipient crisis.

2.3 Incomplete information and contagion

Financial systems dominated by banks are generally more illiquid and opaque. As a consequence, their solvency is easily undermined by variations in the cost of funds.

In fact, close bank-firm relationships are not necessarily bad: many banks may fund illiquid projects that are profitable and solve problems of temporary illiquidity for projects that other financiers would not fund because of imperfect information. However, because of the very nature of these bank-firm relationships, investors are not able to distinguish banks that are funding unprofitable projects from banks that are helping firms to solve liquidity problems. Therefore, when expectations on the losses accumulated by the banking system worsen, international investors demand a higher interest rate on their deposits with all the banks of a country: the increase in the interest rate burden may provoke a banking crisis, although many banks may have been only illiquid ex ante. Moreover, the increase in the cost of funds and the crisis may also involve countries that for some reason are considered "similar" by international investors, because, for instance, they belong to the same geographical area.

The experience of Malaysia in 1997 provides a striking example of this vulnerability of relationship banking to external variations in the cost of funds. In comparison to the other East Asian economies, the situation of Malaysia was different because its banking system was relatively strong in 1997, before the onset of the crisis (IMF (1998)). In fact, following the banking crisis of 1985-88, the asset quality of the Malaysian banking system had improved substantially. The ratio of non-performing loans to total lending fell from a peak of 35% in 1987 to 3.6% by mid-1997 (even though banks' total lending to the private sector had increased in Malaysia as well). However, at the onset of the crisis, investors did not appear to notice these differences: the cost of external funds increased, and banks and finance

[2] The Nordic countries did not belong formally to the EMS, but had their currencies pegged to the ecu.

companies experienced a significant deterioration in asset quality. The main source of vulnerability was the high leverage of the economy: the ratio of banks' claims on the private sector to GDP was over 140% in 1996. The Malaysian authorities responded by injecting liquidity into the banking system in order to keep interest rates low regardless of the negative impact on the currency. The consequences of the crisis in Malaysia were almost indistinguishable from those in South Korea.

The experience of Malaysia suggests that an illiquid and highly leveraged banking system may be an important channel of contagion, even if the banking system is not insolvent.

A very similar mechanism may explain Argentina's experience during the Tequila crisis. On the eve of the introduction of the Convertibility Plan in 1991, financial intermediation in Argentina had reached its lowest point. With the advent of macroeconomic stabilisation, though, the banking industry registered significant productivity improvements and credit to the private sector rose. This process was interrupted by the devaluation of the Mexican peso in December 1994, which led to a sharp increase in the perceived risk of bank liabilities. As a consequence, the interest rate on commercial banks' 30-day deposits jumped and deposits fell (Edwards (1998)). Since Argentina had a currency board, which did not allow the central bank to provide liquidity or bail out the banking system, the increase in the interest rate on deposits may be attributed either to an increase in the perceived probability of bank defaults or to the currency board's imperfect credibility. In either case, the run on deposits and the increase in the cost of funds provoked widespread bankruptcies, bank failures and a deep recession.

Giannetti (2002b) shows formally the mechanism through which illiquid banks are driven into insolvency when the cost of funds rises because international investors have incomplete information on the quality of the banks' assets.

2.4 Summary

A common element of the aforementioned episodes is the centrality of the banking system in the development of the crises. In a few cases, such as Chile and Korea, the crises seem to have been unavoidable outcomes of the banks' insolvency. On the other hand, Malaysia was probably driven into insolvency by an increase in the interest rate burden, which resulted from a loss of confidence in East Asian economies. However, the crisis was made possible by the high indebtedness of the economy and the illiquidity of its banking system.

Moreover, in all cases, financial liberalisation was followed by massive capital inflows and a rapid increase in bank lending. What is striking is that the financial systems appeared stable before the financial liberalisation. Why did capital inflows undermine financial stability? What is specific to the financial systems of these economies? I suggest that if there is shortage of lenders and the source of funds is one main bank, a soft budget constraint problem may arise when an economy is opened to capital inflows. Consequently, insolvent projects may be financed, driving an accumulation of losses by the banking system. Moreover, if international investors have incomplete information about the solvency of a country's banking system and if they attribute a positive probability to banks' default in countries that are only illiquid, then an increase in the interest rate burden may drive banks into insolvency, even if they would have been able to recover their loans in the long run.

3. Some policy implications

The previous section has described how several recent episodes of banking crises can be associated with financial underdevelopment, close bank-firm relationships and a lack of transparency. Under these conditions, it is possible to show that excessive lending to non-profitable projects and sudden stops of capital inflows emerge in equilibrium. This section discusses different institutional arrangements that can improve financial stability.

The imposition of capital controls can, of course, reduce the incentives for banks to renew loans also to projects that are not profitable, since it provides a credible commitment not to expand credit. However, capital controls also impede banks' funding of new investment opportunities that may arise in an economy and, for this reason, may not be the most desirable solution.

Guarantees on deposits are totally irrelevant to improvement of financial instability, if problems of coordination among investors can be reduced through bankruptcy laws or by international institutions. In Giannetti's models, in which the existence of coordination problems among investors is ruled out for simplicity, financial crises may emerge both with and without guarantees on deposits. Guarantees on deposits can only affect the timing of the crisis.

In contrast, financial market development can reduce dramatically the propensity to financial crisis: of course, an improvement in transparency, such as better accounting practices and more stringent disclosure requirements, would eliminate the problems arising from incomplete information, which are supposedly at the origin of banking crises. However, this may be difficult to achieve as very fine information must be provided in order to enable investors to distinguish between illiquid and insolvent banks.

It may be relatively easier to influence the structure of bank-firm relationships: if a firm has a multiplicity of lenders, either banks or bondholders, the incentive to renew loans to projects that turn out to be insolvent disappears. Therefore, the possibility of excess lending and of sudden stops is eliminated if the banking system becomes more competitive and firms no longer have a main bank providing most of the credit. Most importantly, if bond markets become more important, firms acquire many more lenders, who have no incentive to continue to provide working capital if the firm cannot repay previous loans. This provides a theoretical foundation and illustrates a mechanism that supports an often advocated policy tenet: a country should have appropriate financial structures in place before removing capital controls.

Tables

Table 1

Chile, 1982

	1975	**1976**	**1977**	**1978**	**1979**	**1980**	**1981**	**1982**
Outstanding short-term liabilities (as a percentage of GDP)	12.1	9.2	10.1	13.5	16.4	21.9	30.3	48.9
Loans of banking system to private sector (as a percentage of GDP)	6.4	8.9	14.8	20.3	28.2	40.2	54.9	61.7

Source: Velasco (1991).

Table 2

Lending boom in East Asian economies

Rate of growth of bank lending to the private sector

	1991	**1992**	**1993**	**1994**	**1995**	**1996**
Korea	20.78	12.55	12.94	20.08	15.45	20.01
Indonesia	17.82	12.29	25.48	22.97	22.57	21.45
Thailand	20.45	20.52	24.03	30.26	23.76	14.63
Malaysia	20.58	10.79	10.80	16.04	30.65	20.24

Source: Corsetti et al (1998).

Table 3

Financial fragility of East Asian economies

Short-term liabilities towards BIS banks
as a percentage of foreign reserves, end-1996

Korea	**Indonesia**	**Thailand**	**Malaysia**
213	181	169	47

Source: Corsetti et al (1998).

References

Corsetti, G, P Pesenti and N Roubini (1998): "What caused the Asian currency and financial crisis?", Banca d'Italia, *Temi di Discussione*, no 343.

——— (1999): "Paper tigers? A model of the Asian crisis", *European Economic Review*, vol 43, pp 1211-36.

Drees, B and C Pazarbaşioğlu (1995): "The Nordic banking crises: pitfalls in financial liberalization?", *IMF Working Paper*, no 95/61.

Edwards, S and C A Végh (1997): "Banks and macroeconomic disturbances under predetermined exchange rates", *Journal of Monetary Economics*, vol 40, pp 239-78.

Giannetti, M (2002a): Bank-firm relationships and contagious banking crises, *Journal of Money, Credit and Banking*, forthcoming.

——— (2002b): *On the causes of overlending: are guarantees on deposits the culprit*?, Stockholm School of Economics, mimeo.

International Monetary Fund: *International capital markets*, various issues.

Krugman, P (1998): *What happened to Asia?*, Massachusetts Institute of Technology, mimeo.

McKinnon, R and H Pill (1996): "Credible liberalization and international capital flows: the overborrowing sydrome", in T Ito and A O Krueger, eds, *Financial deregulation and integration in East Asia*, Chicago University Press, Chicago.

Nam, Sang-Woo (1996): "The principal transactions bank system in Korea and a search for a new bank-business relationship", in T Ito and A O Krueger, eds, *Financial deregulation and integration in East Asia*, Chicago University Press, Chicago.

Radelet, S and J Sachs (1998): "The East Asian financial crisis: diagnosis, remedies and prospects", *Brookings Papers on Economic Activity*, pp 1-90.

Rajan, R G and L Zingales (1998): "Which capitalism? Lessons from the East Asian crisis", *Journal of Applied Corporate Finance*, vol 11, pp 40-8.

Velasco, A (1991): "The Chilean financial system, 1975-85", in V Sundararajan and T J T Balino, eds, *Banking crises: cases and issues*, International Monetary Fund, Washington DC.

Session 2

Market contagion

The transmission of contagion in developed and developing international bond markets

Mardi Dungey, Renée Fry, Brenda González-Hermosillo and Vance Martin[1]

Abstract

The potential for contagion effects resulting from financial crises has become an important policy issue. The results presented in this paper quantify the impact of financial contagion in global bond markets from the Russian crisis and the LTCM near collapse during the latter part of 1998. Using a panel of bond spreads in 12 countries we find discernible contagion from these two crises to both developing and developed markets. The proportion of total volatility attributable to contagion varies widely across countries but it is not always the case that it is more substantial for developing countries. However, due to the absolutely higher level of volatility experienced in developing markets, the squared basis point contributions of contagion to volatility are relatively higher in those markets.

1. Introduction

The period 1994 to 2001 witnessed financial crises from diverse regions, including Mexico, East Asia, Russia, the near collapse of Long-Term Capital Management (LTCM) in the United States, Brazil, Turkey and Argentina. Although the origins of these crises can be geographically located, their effects were not necessarily isolated as shocks spilled over geographical boundaries causing financial turmoil in other, sometimes unrelated, financial markets. This was the case with the Russian crisis of August 1998, which precipitated sharp rises in bond spreads in a broad range of countries, which were followed in the next month by further movements in bond spreads arising from the near collapse of LTCM.

The Russian and LTCM crises are qualitatively different as the Russian crisis was a crisis of credit risk, whilst the LTCM crisis was a crisis of liquidity. The crises originated in very different markets as Russia is characterised as a developing market and the United States as developed, suggesting the impacts of the two crises on international bond markets should differ. CGFS (1999) claims that the Russian crisis affected only developing markets, while the LTCM crisis affected developed markets. A similar conclusion is put forward by Bae et al (2000), who find that for a range of international equity markets, developing markets are more susceptible to international financial crises than developed markets; see also Kaminsky and Reinhart (2002), who propose that developed markets act only as a conduit between regions of developing markets. However, these conclusions are not based on any formal study of the interactions of bond markets between countries. Part of the reason for the lack of emphasis on bond markets is the difficulty of constructing consistent data sets across both developed and developing bond markets. These data issues are addressed in this paper.

The aim of this paper is to investigate the transmission of the Russian and LTCM crises across the bond markets of nine developing and three developed countries. The countries are grouped into three regions: Argentina, Brazil and Mexico from Latin America; Indonesia, South Korea and Thailand from

---

[1] Australian National University, Queensland University of Technology, International Monetary Fund and University of Melbourne, respectively. This paper builds on the research undertaken in a companion paper "International contagion effects from the Russian crisis and the LTCM near collapse" which was presented at the Third Joint Central Bank Conference on Risk Measurement hosted by the BIS, Basel, 7-8 March 2002. Part of the research for this paper was undertaken when Mardi Dungey was a visiting scholar at the IMF Institute. Mardi Dungey and Vance Martin acknowledge funding from ARC grant no A00001350. The authors are grateful to Takishito Ito, Charles Goodhart, Leslie Hull, José Lopez, Reza Vaez-Zadeh, participants at the Third Joint Central Bank Research Conference, seminar participants at the IMF, the German Association of Investment Professionals and the Swiss Association of Investment Professionals for helpful comments. The views expressed in this paper are those of the authors and do not necessarily represent those of the International Monetary Fund.

Asia; Bulgaria, Poland and Russia from eastern Europe; and the Netherlands, the United Kingdom and the United States as the representative developed markets. Daily yields on sovereign and corporate bond issues are used over the period February to December 1998.

The financial market shocks transmitted across geographical boundaries are specified to occur through either anticipated or unanticipated channels. The anticipated spillovers include linkages which capture changes in market fundamentals and economic relationships between countries. The unanticipated spillovers are the shocks of interest in the present study. These shocks represent the possibility of significant linkages between countries that are not transparent. Unanticipated spillovers are defined here as contagion; see Masson (1999a,b,c), Favero and Giazvazzi (2000) and Forbes and Rigobon (2000, 2002).

In modelling the international linkages between markets, anticipated spillovers are specified as latent factors to overcome the ad hoc identification of market fundamentals from proxy variables; see also Dungey et al (2000). As the latent factor model does not rely on observable data on market fundamentals, this structure has the additional advantage of allowing for high-frequency transmission processes, an advantage when shocks are relatively short-lived and occur in close succession as in the Russian and LTCM crises.

The rest of the paper proceeds as follows. Some background to the Russian and LTCM crises is presented in Section 2. A model of contagion is specified in Section 3. The empirical results are presented in Section 4 with concluding comments given in Section 5. The key result of the empirical analysis is that contagion from the Russian and LTCM crises is spread across both developing and developed bond markets. However, the impact of contagion from the Russian crisis on the bond markets of the three developed countries investigated, in terms of squared basis points, is relatively small. For the United States, the effect of contagion from the Russian crisis is less than 1 squared basis point, whereas the effect of contagion on the Russian bond market emanating from the LTCM crisis is over 6,000 squared basis points.

2. Background

On 17 August 1998, Russia announced sweeping changes to its financial system, including the intention to restructure all official domestic currency debt obligations falling due to the end of 1999, imposed a 90-day moratorium on the repayment of private external debt, and effectively devalued its currency by widening the trading band of the rouble (see Kharas et al (2001) for a discussion of the Russian crisis). These events in Russia led to increased volatility in global bond markets, as credit and sovereign risks were reassessed by the global financial community.

On 23 September 1998, just weeks after the instability caused by the events in Russia, financial markets were hit with another shock. The New York Federal Reserve was compelled to orchestrate a rescue package to prevent the US-based hedge fund LTCM from collapsing. The investment strategies of LTCM had priced risk using "normal" volatilities and spreads between closely related securities, some of which seemed to have changed in the aftermath of the Russian crisis.

The Russian crisis and the near collapse of LTCM led to large jumps in spreads and risk premia. The impacts of these crises on the global bond markets are highlighted in Figure 1, which gives the daily spread of long-term debt over the appropriate risk-free yield for a range of developing and developed countries over the period May to December 1998 (see Appendix A for source descriptions and definitions). In the discussion that follows, the spread is referred to as the "premium" while recognising that it may reflect a myriad of factors.

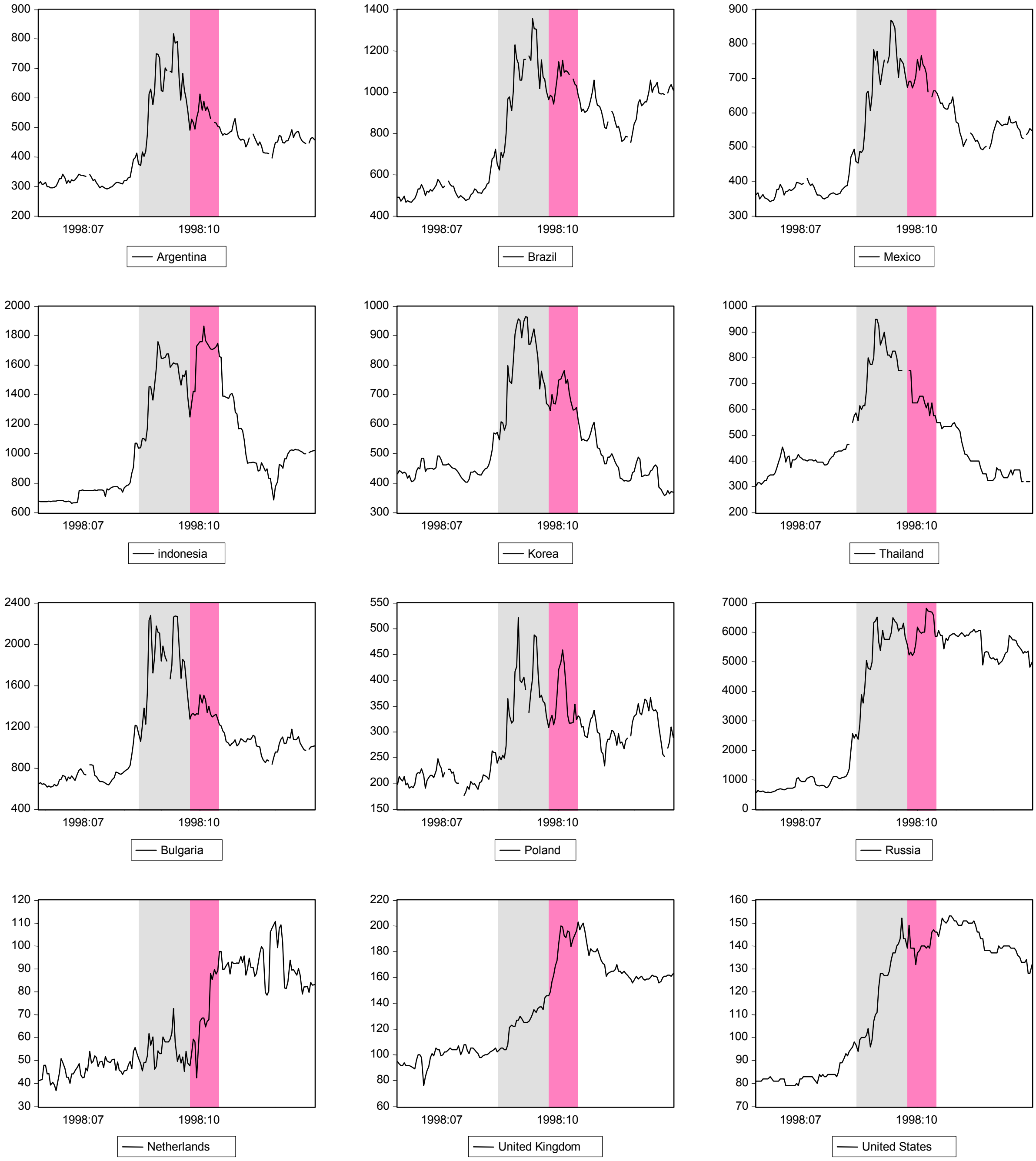
The extent and timing of the Russian and LTCM shocks on international bond markets are further highlighted in Figure 2, which gives daily changes in bond spreads in each country. One characteristic demonstrated in these figures is that both emerging and mature markets were affected by these unanticipated events.[2]

[2] The two crises had a much more dramatic impact on global bond spreads than other recent shocks, such as the Hong Kong speculative attack; see Dungey et al (2002).

Figure 1

Bond spreads, May-December 1998[1]

(basis points)



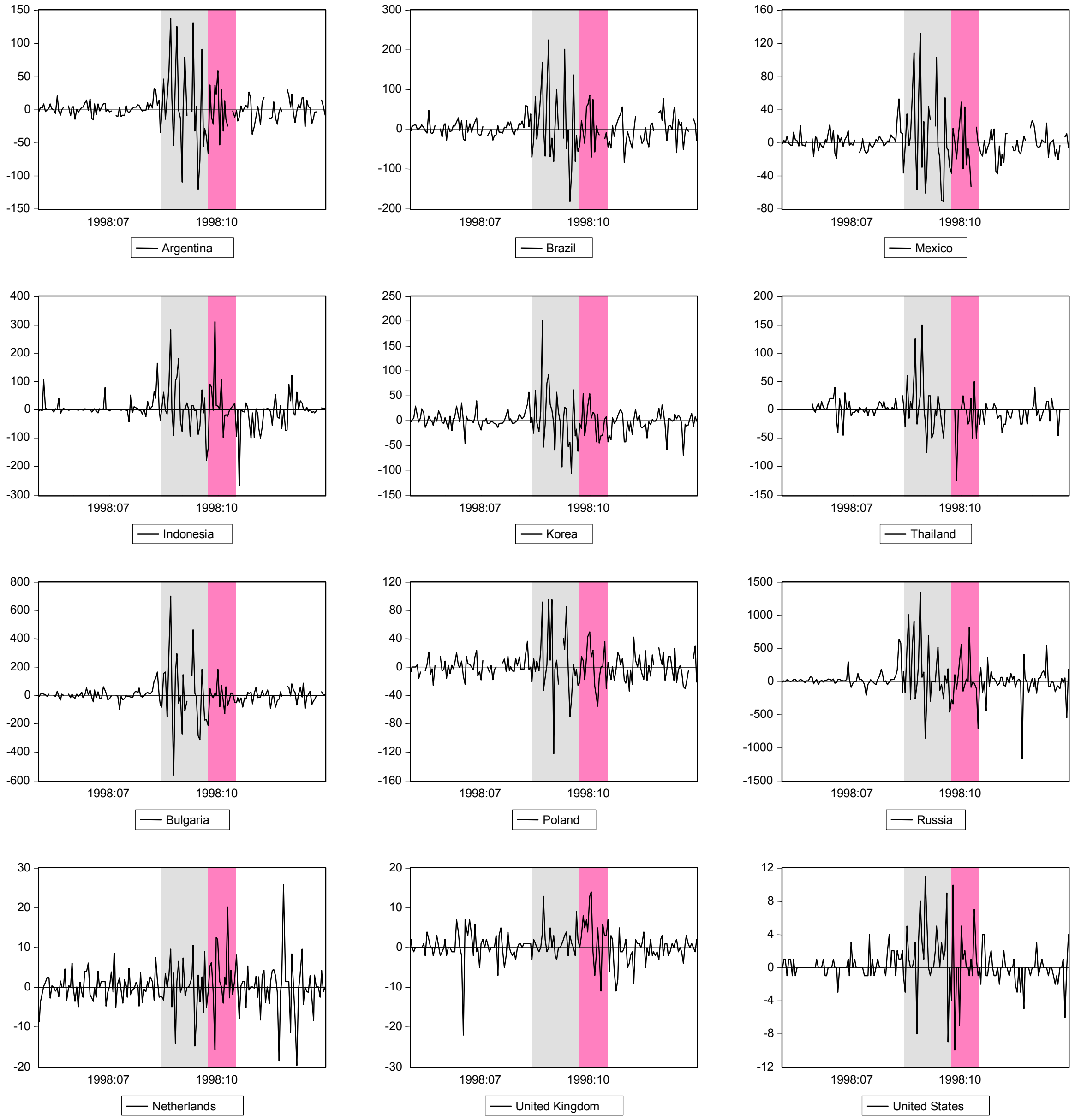
[1] The shaded areas refer to episodes of crisis in international bond markets during this period: the Russian bond default on 17 August 1998; the bailout of LTCM orchestrated by the New York Federal Reserve on 23 September 1998; and the inter-FOMC Fed interest rate cut on 15 October 1998 which signalled thebeginning of the "end" of the LTCM crisis.

Sources: US Federal Reserve; Bloomberg; Scotia Capital; Credit Suisse First Boston.

Figure 2

Bond spreads in first differences, May-December 1998[1]

(basis points)



[1] The shaded areas refer to episodes of crisis in international bond markets during this period: the Russian bond default on 17 August 1998; the bailout of LTCM orchestrated by the New York Federal Reserve on 23 September 1998; and the inter-FOMC Fed interest rate cut on 15 October 1998 which signalled the beginning of the "end" of the LTCM crisis.

Sources: US Federal Reserve; Bloomberg; Scotia Capital; Credit Suisse First Boston.

Figures 1 and 2 suggest that the aggressive easing of monetary policy in the United States was helpful in ending the LTCM crisis.[3] The interest rate cuts were in part motivated by concerns that the US economy was on the verge of experiencing a liquidity crash as bond spreads in the United States, and

[3] According to market participants surveyed by CGFS (1999), the surprise inter-FOMC meeting interest rate cut on 15 October 1998 signalled the beginning of the abatement of financial constraints.

many other countries, rose to exceptionally high levels. The Federal Reserve actions may have staved off an even more dramatic crisis. Based on interviews with market participants, CGFS (1999, p 40) noted (our italics) that:

"Only a very small number of market participants declined to characterise the 1998 crisis as "exceptional". Most interviewees mentioned that the events described […] led to *the worst crisis ever*."

Inspection of the bond spreads in the second half of 1998 (Figures 1 and 2) suggests that the Russian crisis had a substantial impact on all countries examined, both advanced economies and emerging markets. The LTCM shock also appears to have had an impact on all the countries, with a relatively smaller hump experienced by most emerging markets relative to the effect of the Russian shock. An inspection of the data suggests that the Russian and LTCM shocks were reinforcing in international financial markets as practically all markets experienced two jumps in their spreads: one following the Russian default (the first band in the figures) and another one following the announcement of LTCM's financial problems (the second band in the figures). Similarly, the fact that bond spreads began to rise in the United States following the Russian crisis and the Russian sovereign spread rose further in the aftermath of the LTCM crisis suggests that these two events were not totally independent.

The financial crises during August-September 1998 are interesting because the shocks during this period seem to have been transmitted across countries with little in common - including countries that do not fit traditional explanations of contagion based on trade links, competitive devaluation or regional effects (see for example Lowell et al (1998) and Goldstein (1998) for taxonomies of contagion). The crises of 1998 affected countries as diverse as Russia and Brazil (Baig and Goldfajn (2000) argue that the Russian crisis precipitated the Brazilian crisis), and emerging and advanced economies. Examining these crises is complicated by the fact that the two shocks occurred within weeks of each other.

3. A model of contagion

Existing empirical models of contagion generally proceed by both conditioning on a set of economic indicators to proxy market fundamentals and specifying the timing of contagious events. These choices tend to be based on an ex post evaluation of the data, and are often statistically insignificant (see, for example, Eichengreen et al (1995, 1996), Sachs et al (1996) and Glick and Rose (1999)). Latent factor models provide a desirable alternative circumventing the need to choose specific indicators to proxy economic fundamentals (see Dooley (2000) and Edwards (2000)). This type of model has been adopted previously for equity markets (Forbes and Rigobon (2002), King et al (1995)), currency markets (Dungey (1999), Mahieu and Schotman (1994), Diebold and Nerlove (1989)) and fixed interest markets (Gregory and Watts (1995), Dungey et al (2000)).

The premium of each of the 12 countries in Figures 1 and 2 is presumed to evolve in response to a number of alternative types of factors. These factors are classified as common to the entire set of countries, common to a regional grouping of countries, or idiosyncratic and related only to individual countries themselves. However, in contrast to many of the existing empirical models of contagion, the factors are not assumed to be observed directly, instead the revealed information in the data on premia is used to identify the factors. The structure of the factor model developed here has origins in the two factor models developed particularly in the equity market, where equity market returns are classified into common and country-specific shocks; see, for example, King et al (1995). In the case of the *N=12* countries investigated in this paper, it is natural to include also a further set of regional factors to capture shocks contained within specific geographical areas. Thus the premium $P_{i,t}$ on the bond in country *i* at time *t* is expressed as

$$P_{i,t} = \lambda_i W_t + \phi_i f_{i,t} + \gamma_i R_{k,t} \qquad i = 1,...,12,\ k = Latin, Asia, Europe, \tag{1}$$

where the premium is represented as the sum of a time-varying common factor, W_t, a time-varying country-specific factor, $f_{i,t}$, and a time-varying regional factor, $R_{k,t}$. The loadings on these factors vary across countries and are given by the parameters λ_i, ϕ_i and γ_i.

To identify the latent factors, and hence the parameters of the model, the common world factor W_t and the three regional factors $R_{k,t}$ are modelled to evolve as unit root processes

$$W_t = W_{t-1} + \eta_t, \tag{2}$$

$R_{k,t} = R_{k,t-1} + \nu_{k,t}$ *where k = Latin, Asia, Europe*, (3)

where η_t and $\nu_{k,t}$ are stationary and independent disturbance terms. This structure is motivated by the need to specify a model which captures the unit root properties in the premium variables (see Dungey et al (2002) for details of the unit root tests). The remaining factors, the country-specific factors $f_{i,t}$, are specified as stationary and independent disturbance terms. In addition, Figure 1, and in particular Figure 2, highlight the occurrence of volatility clustering especially during the two crisis periods. To capture this autocorrelation in the volatility of the bond spreads, the world common factor error term η_t is assumed to have a GARCH (1,1) conditional variance.[4]

Equation (1) provides an initial decomposition of the premia of each of the 12 countries in terms of common, country-specific and regional factor shocks. To capture the effects of contagion across country bond markets, equation (1) is expanded to include the effects of the transmission of unanticipated shocks across international bond markets. The focus of the empirical investigation is on identifying and measuring the relative size and impact of contagion from the Russian and LTCM crises. Equation (1) is expanded as

$$P_{i,t} = \lambda_i W_t + \phi_i f_{i,t} + \gamma_i R_{k,t} + \delta_i f_{Russia,t} + \rho_i f_{US,t} \qquad i = 1,...,12,\ k = Latin,\ Asia,\ Europe, \tag{4}$$

where δ_i, measures the impact of contagion from Russia, and ρ_i measures the impact of contagion from the LTCM crisis, which is proxied by the unanticipated shocks from the US bond market.

In measuring the relative size of the impact of contagion across international bond markets, the latent factor model can be used to decompose the relative contributions of each factor to the volatility in the bond premium of each market. In deriving this decomposition it is necessary to work with the change in the bond premia, as these variables are non-stationary. To achieve this, equation (4) is interpreted as a cointegrated system which is used to derive an error correction model in terms of $\Delta P_{i,t}$. Following Dungey et al (2002), the volatility decomposition is expressed as

$$Var\left(\Delta P_{i,t}\right) = \lambda_i^2 + 2\phi_i^2 + \gamma_i^2 + 2\delta_i^2 + 2\rho_i^2. \tag{5}$$

In turn, the total decomposition can be re-expressed as a proportion of the contribution of each factor to total volatility:

(i) contribution of the world factor $\dfrac{\lambda_i^2}{Var\left(\Delta P_i\right)}$

(ii) contribution of country-specific factor $\dfrac{2\phi_i^2}{Var\left(\Delta P_i\right)}$

(iii) contribution of the regional factor $\dfrac{\gamma_i^2}{Var\left(\Delta P_i\right)}$

(iv) contribution of contagion from Russia $\dfrac{2\delta_i^2}{Var\left(\Delta P_i\right)}$

(v) contribution of contagion from LTCM $\dfrac{2\rho_i^2}{Var\left(\Delta P_i\right)}$

These statistics are average measures over the sample of the proportion of volatility arising from shocks from each factor. It is also possible to calculate conditional decompositions, which give the proportionate contribution of each shock at each point in time over the sample period.

In the special case where the factors have autoregressive representations and homoskedastic error variances, a Kalman filter can be used to estimate the unknown parameters. However, the inclusion of

---

4 Univariate GARCH (1,1) tests on the individual country premium data confirm the presence of GARCH processes, with some common features. In earlier work we allowed the GARCH to apply to a greater variety of the factors, but found that this was generally insignificant. In line with Dungey et al (2000) and Kose et al (1999), GARCH on the common factor appears to capture the properties of the data.

conditional volatility in the factor variances precludes the use of the Kalman filter, as the parameter estimates are no longer consistent. To overcome this problem a simulation-based estimator is adopted following the approach of Gourieroux et al (1993) and Gallant and Tauchen (1996) (see also Gourieroux and Monfort (1994)). The approach consists of simulating the contagion model in equations (2) to (4) to generate a set of simulated bond spreads for the 12 countries in the sample. The simulated spreads are then calibrated with the actual bond spreads via a set of moment conditions derived from a set of VARs based on both the levels and squares of the bond spreads (the details of the estimation method are contained in Dungey et al (2002)).

4. Results

This section presents the results of estimating the latent factor contagion model for international bond markets. The sample period consists of daily bond yield spreads in 1998, beginning in February and ending in December, for the 12 countries shown in Figures 1 and 2.

Table 1 gives the volatility decompositions based on equation (5), expressed in percentage terms, whilst Figure 3 provides a graphical representation. Given the integration of international financial markets, volatility in bond premiums should exhibit strong co-movements. The contributions of the world factor confirm this, accounting for between 82% and 99% of total volatility. The regional and country-specific factors have little influence on volatility, with these factors accounting for less than 1% of total volatility, with the exceptions of the Netherlands and South Korea, where the effects are still relatively small.

Table 1

Volatility decomposition with contagion effects from Russia and the United States

(contribution to total volatility, in percentages)

	World	Country	Regional	Contagion		
				From Russia	From US	Total
Industrial						
US	91.080	0.050	0.000	8.870	–	8.870
UK	99.344	0.133	0.000	0.040	0.482	0.523
Netherlands	82.793	2.777	0.000	10.615	3.815	14.431
Eastern Europe						
Russia	89.145	0.222	0.086	–	10.547	10.547
Poland	88.963	0.050	0.514	1.279	9.194	10.473
Bulgaria	90.204	0.375	0.417	8.111	0.893	9.004
Asia						
Indonesia	99.213	0.268	0.254	0.217	0.048	0.265
South Korea	91.285	5.269	0.913	0.163	2.369	2.533
Thailand	91.174	0.786	0.547	1.521	5.973	7.493
Latin America						
Mexico	99.426	0.003	0.002	0.327	0.242	0.569
Argentina	83.436	0.028	0.007	0.022	16.508	16.529
Brazil	84.388	0.055	0.045	11.1047	4.407	15.511

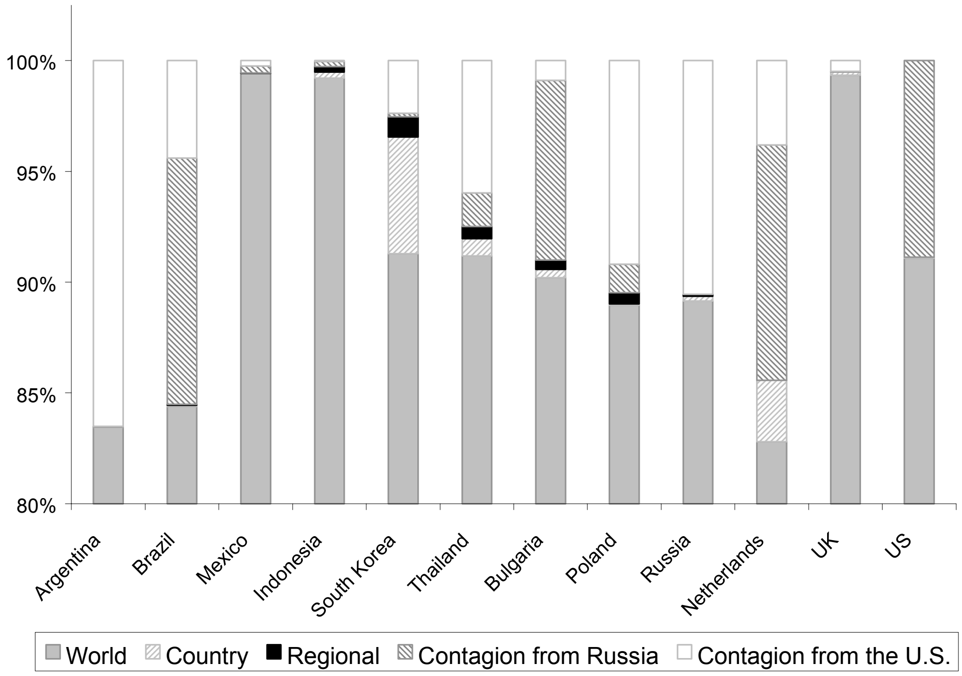
Table 1 shows that contagion is widespread, with five countries experiencing more than 10% of volatility due to contagion, three countries with contagion between 5 and 10% of total volatility, and four countries with contagion less than 5% of total volatility. The results also show that there is no clear association of contagion effects with either developed or developing countries. The three countries experiencing the largest proportion of volatility from contagion, at around 15%, are

Argentina, Brazil and the Netherlands. The three countries with the lowest total contribution from contagion are the United Kingdom, Indonesia and Mexico, with less than 1% in each. There is no evidence that contagion from Russia is confined to developing countries, or that contagion from LTCM mainly affects developed markets. In fact, the greatest impact of contagion from LTCM is felt in developing markets.

Figure 3

Volatility decomposition with contagion effects from Russia and the United States

(contribution to total volatility, in percentages)



The results in Table 1 also show that contagion is not generally contained within regions (the importance of regional effects has been studied by Goldstein (1998), Kaminsky and Reinhart (2002), Goldstein et al (2000), Eichengreen et al (1996) and Glick and Rose (1999)). For example, within the eastern Europe region, the Russian crisis has a substantial impact in Bulgaria, but not in Poland. Further, larger contagion effects are felt outside eastern Europe, in Brazil and the United States for example.

An alternative representation of the volatility decompositions is provided in Table 2 by expressing the volatility decompositions in squared basis points. This is achieved by multiplying the results in Table 1 by the sample variance of the daily change in the bond premia. Figures 4 and 5 provide a graphical representation of the relative size of contagion in terms of squared basis points.

The relatively higher overall level of volatility in developing markets means that the basis point effects of contagion are larger in developing markets than developed markets. In the United States and United Kingdom, contagion effects are less than 1 squared basis point. In the Netherlands, the effect is around 4 squared basis points. The only developing markets to have a single digit impact from contagion are Mexico, at 3 squared basis points, and Indonesia, at around 8 squared basis points. The remaining countries show contagion effects ranging from 21 squared basis points in South Korea to 6,200 in Russia.

The contribution of contagion to volatility in Russia and Poland is given as approximately 10% in Table 1, yet the scaled results in Table 2 show that contagion in Poland is only 55 squared basis points, in contrast with the 6,200 squared basis points for Russia. Similarly, the proportionate contribution of contagion from Russia to volatility in the United States and Bulgaria given in Table 1 is approximately the same, but translates to less than 1 squared basis point for the United States and about 811 squared basis points for Bulgaria.

Argentina and Brazil, the two developing countries most affected by contagion in percentage terms, were themselves to experience a financial collapse in January 1999 and 2001 respectively. However, the results in Table 1 show that the sources of contagion in these two countries are different. Almost all of the contagion to Argentina was sourced from the LTCM near collapse, with little impact from Russia. In Brazil approximately two thirds of the contagion effects were sourced from Russia, with the remaining third from the LTCM near collapse. This result is consistent with Baig and Goldfajn (2000), who emphasised the importance of contagion from Russia in explaining the financial crisis in Brazil in 1999. In basis point terms, volatility in Argentina and Brazil was substantial, with contagion from the crises contributing about 187 squared basis points to Argentina's premium, and 545 squared basis points to Brazil. These results may provide an interesting lead for future work in establishing at what point evidence of pre-crisis jitters are evident in financial markets.

Table 2

Volatility decomposition with contagion effects from Russia and the United States

(contribution to total volatility, in squared basis points)

	Total	Components			
		World	Country	Regional	Contagion
Industrial					
US	7.503	6.838	0.004	0.000	0.666
UK	13.895	13.822	0.019	0.000	0.073
Netherlands	29.052	24.053	0.807	0.000	4.192
Eastern Europe					
Russia	57872.003	52401.260	130.337	50.573	6199.837
Poland	527.622	469.385	0.263	2.715	55.259
Bulgaria	10006.001	9025.839	37.527	41.680	900.955
Asia					
Indonesia	3121.457	3096.893	8.359	7.941	8.264
South Korea	820.250	748.769	43.217	7.490	20.775
Thailand	499.870	455.843	3.928	2.735	37.465
Latin America					
Mexico	526.703	523.678	0.017	0.011	2.997
Argentina	1133.669	945.883	0.315	0.081	187.390
Brazil	3515.304	2966.509	1.939	1.585	545.270

The Indonesian results also raise interesting questions. In analyses of the East Asian crisis, Indonesia has been singled out as the country most affected by contagion (see Goldstein et al (2000) and Radelet and Sachs (1998)). However, the impact of contagion from both Russia and the LTCM near collapse in Indonesia is relatively small. This raises the question as to whether this is due to some structural change in Indonesia, or perhaps a heightened sensitivity to financial crises, moving the transmission mechanism to anticipated effects and hence away from contagion. A proposition worthy of investigation is whether a country can become immune to contagion, but nonetheless experience relatively high volatility from anticipated spillovers.

Figure 4

Contagion in squared basis points - the smaller contributions

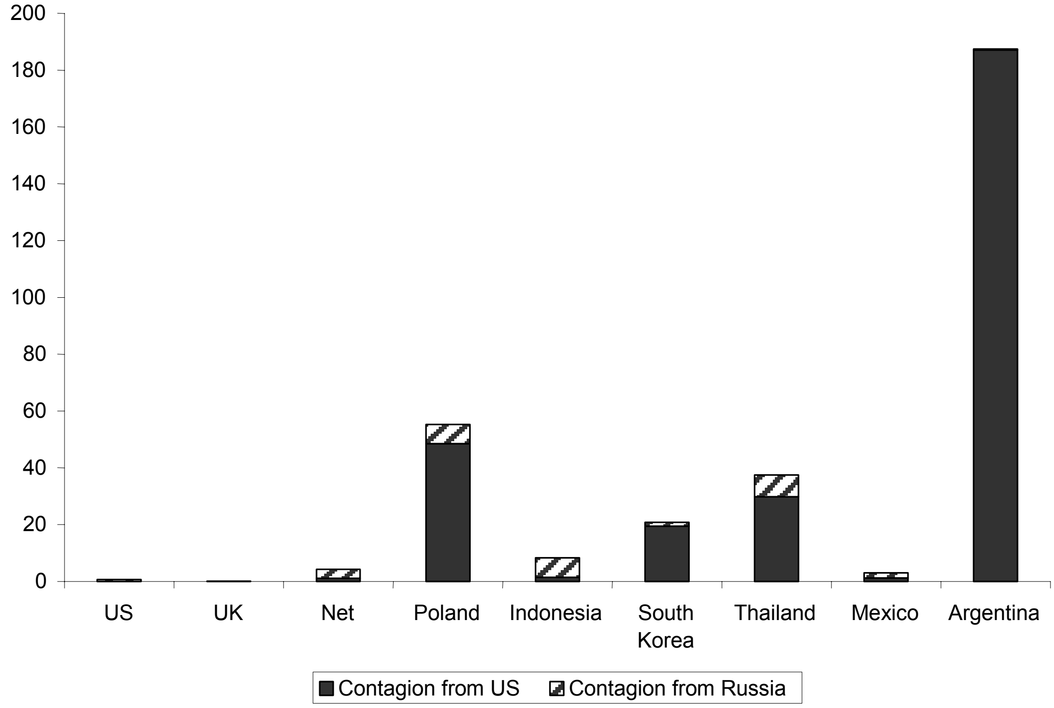
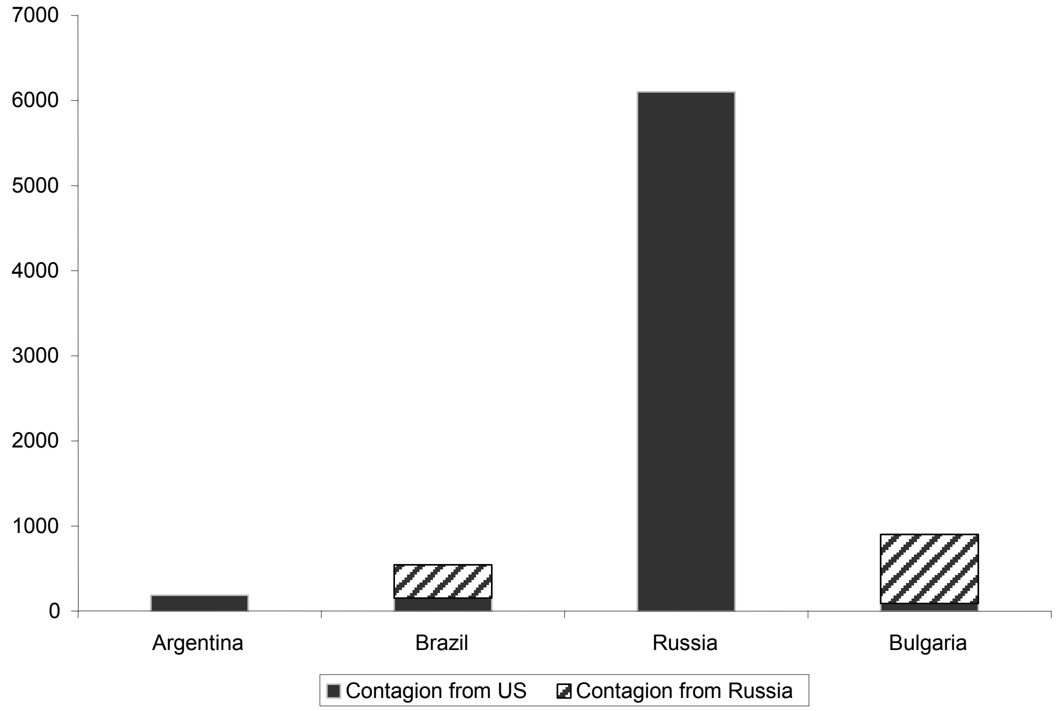


Figure 5

Contagion in squared basis points - the larger contributions



5. Conclusion

The Russian and LTCM crises have measurable contagion effects across a broad range of international bond markets in developing and developed countries. The markets examined were those of nine developing countries - Thailand, South Korea and Indonesia in Asia; Poland, Bulgaria and Russia in Eastern Europe; and Mexico, Argentina and Brazil in Latin America - and three developed markets - the US, UK and Netherlands markets. Contagion effects from both crises affected all countries and regions to differing degrees.

The results show that the Russian crisis produced contagion to both developing and developed markets, with the largest proportionate effects on Brazil, Bulgaria, Thailand and the Netherlands. The LTCM crisis effects were generally smaller, but were felt most in Argentina and Russia. The mix of developing and developed markets in the results belies the conclusions of CGFS (1999) that the effects of the Russian crisis were felt in developing markets and that the LTCM near collapse mainly impacted mature markets.

Contagion effects expressed as a proportion of total volatility did not provide clear evidence as to whether contagion had a greater effect on developing or developed markets. The greatest proportionate effects were felt in Brazil and the Netherlands, and the least in Mexico and the United Kingdom, that is in a developing and developed country in both cases. However, when the results were expressed in squared basis points, contagion effects were larger in developing markets as a result of the higher degree of volatility in these markets.

The results also showed that Brazil was affected by contagion prior to its currency crisis in January 1999. The relatively large contagion effects may be a reflection of the vulnerability of this country. This hypothesis provides scope for future work identifying the timing of financial crises through the identification of pre-crisis jitters in financial markets.

Contagion has been viewed in the literature as mainly a concern for developing countries. The evidence from the Russian and LTCM crises suggest this is not necessarily the case. The overall higher volatility in developing markets means that the effects of contagion in those markets are higher measured in squared basis point terms. However, in proportionate terms, contagion effects are widely distributed across both developed and developing markets. Contagion is not a phenomenon reserved for developing countries; developed markets are also affected.

Appendix A:
Data definitions and sources

The data represent the spread of long-term debt over the appropriate risk-free yield for each country. The choice of the risk-free rate was specific to each long-term bond, because it depends at least in part on the currency of denomination of the bond issue. In the case of the emerging market countries, sovereign bonds were issued in US dollars, rather than in domestic currency, and hence the spread is calculated against the comparable maturity-matched US Treasury bond rate. To the extent possible, the bonds selected for emerging markets were sovereign issues (rather than Brady) to reflect the true cost of new foreign capital. For the advanced markets, which are able to issue international bonds in domestic currency, benchmark BBB investment grade corporate bonds were used and compared to the corresponding risk-free Treasury bond in each country. Sources of the data are:

Argentina: Republic of Argentina bond spread over US Treasury. Source: US Federal Reserve.

Brazil: Republic of Brazil bond spread over US Treasury. Source: US Federal Reserve.

Mexico: JP Morgan eurobond index Mexico sovereign spread over US Treasury. Source: US Federal Reserve.

Indonesia: Indonesian yankee bond spread over US Treasury. Source: US Federal Reserve.

South Korea: Government of Korea 8 7/8% 4/2008 over US Treasury. Source: Bloomberg (50064FAB0).

Thailand: Kingdom of Thailand yankee bond spread over US Treasury. Source: US Federal Reserve.

Bulgaria: Bulgarian discount stripped Brady bond yield spread over US Treasury. Source: US Federal Reserve.

Poland: Poland par stripped Brady bond yield spread over US Treasury. Source: US Federal Reserve.

Russia: Government of Russia 9.25% 11/2001 over US Treasury. Source: Bloomberg (007149662).

Netherlands: Akzo Nobel NV 8% 12/2002 yield spread over NETHER 8.25% 6/2002. Source: US Federal Reserve.

United Kingdom: UK industrial BBB corporate 5-year bond spread over gilt. Source: Bloomberg (UKBF3B05).

United States: US industrial BBB1 corporate 10-year bond spread over US Treasury. Source: Bloomberg (IN10Y3B1).

References

Bae, K H, G A Karyoli and R M Stultz (2000): "A new approach to measuring financial contagion", *NBER Working Paper* 7913.

Baig, T and I Goldfajn (2000): "The Russian default and the contagion to Brazil", *IMF Working Paper* WP/00/160, October.

Committee on the Global Financial System (1999): *A review of financial market events in autumn 1998*, Bank for International Settlements, Basel, Switzerland, October.

Diebold, F X and M Nerlove (1989): "The dynamics of exchange rate volatility: a multivariate latent-factor arch model", *Journal of Applied Econometrics*, 4, 1-22.

Dooley, M P (2000): "A model of crises in emerging markets", *Economic Journal*, 110(46), 256-72.

Dungey, M (1999): "Decomposing exchange rate volatility around the pacific rim", *Journal of Asian Economics*, 10, 625-35.

Dungey, M, V L Martin and A R Pagan (2000): "A multivariate latent factor decomposition of international bond yield spreads", *Journal of Applied Econometrics*, 15(6), 697-715.

Dungey, M, R Fry, B Gonzalez-Hermosillo and V L Martin (2002): "International contagion effects from the Russian crisis and the LTCM near-collapse", *IMF Working Paper*, forthcoming.

Edwards, S (2000): "Interest rates, contagion and capital controls", *NBER Working Paper*, 7801.

Eichengreen, B, A K Rose and C Wyplosz (1995): "Exchange market mayhem: the antecedents and aftermath of speculative attacks", *Economic Policy*, 21, 249-312.

——— (1996): "Contagious currency crises", *NBER Working Paper*, 5681.

Favero, C A and F Giavazzi (2000): "Looking for contagion: evidence from the ERM", *NBER Working Paper*, 7797.

Forbes, K and R Rigobon (2000): *Measuring contagion, conceptual and empirical issues*, MIT, mimeo.

——— (2002): "No contagion, only interdependence: measuring stock market co-movements", *Journal of Finance,* forthcoming.

Gallant, A R and G Tauchen (1996): "Which moments to match?", *Econometric Theory*, 12 (4), 657-81.

Glick, R and A K Rose (1999): "Contagion and trade: why are currency crises regional?", *Journal of International Money and Finance*, 18(4), 603-17.

Goldstein, M (1998): "The Asian financial crisis: causes, cures and systemic implications", *Policy Analysis in International Economics*, 55, Institute for International Economics, Washington DC.

Goldstein, M, G L Kaminsky and C M Reinhart (2000): *Assessing financial vulnerability: an early warning system for emerging markets*, Institute for International Economics, Washington DC.

Gourieroux, C and A Monfort (1994): "Simulation based econometric methods", *CORE Discussion Paper*.

Gourieroux, C, A Monfort and E Renault (1993): "Indirect inference", *Journal of Applied Econometrics*, 8, S85-S118.

Gregory, A W and D G Watts (1995): "Sources of variation in international real interest rates", *Canadian Journal of Economics*, 18, S120-S140.

Kaminsky, G L and C M Reinhart, (2002): "The center and the periphery: the globalization of financial turmoil", paper presented at the Third Joint Central Bank Research Conference on Risk Measurement and Systemic Risk, 7-8 March, BIS, Basel, Switzerland.

Kharas, H, B Pinto and S Ulatov (2001): "An analysis of Russia's 1998 meltdown: fundamentals and market signals", *Brookings Papers on Economic Activity*, Brookings Institution, 1, 1-68.

King, M, E Sentana and S Wadhwani (1995): "Volatility and links between national stock markets", *Econometrica*, 62, 901-33.

Kose, M A, C Otrok and C Whiteman (1999): *International business cycles: world, region and country-specific factors*, University of Iowa, mimeo.

Lowell, J, C R Neu and D Tong, (1998): "Financial crises and contagion in emerging market countries", RAND, MR-962.

Mahieu, R and P Schotman (1994): "Neglected common factors in exchange rate volatility", *Journal of Empirical Finance*, 1, 279-311.

Masson, P (1999a): "Contagion: macroeconomic models with multiple equilibria", *Journal of International Money and Finance*, 18, 587-602.

——— (1999b): "Contagion: monsoonal effects, spillovers, and jumps between multiple equilibria", in Agenor, P R, M Miller, D Vines and A Weber: (eds), *The Asian financial crisis: causes, contagion and consequences*, Cambridge University Press, Cambridge, UK.

——— (1999c): "Multiple equilibria, contagion and the emerging market crises", *IMF Working Paper* 99/164.

Radelet, S and J Sachs (1998): *The onset of the East Asian financial crisis*, Harvard Institute for International Development, mimeo.

Sachs, J, A Tornell and A Velasco (1996): "Financial crises in emerging markets: the lessons from 1995", *Brookings Papers on Economic Activity*, 1, 146-215.

Financial turmoil: systemic or regional?[1]

Graciela L Kaminsky, George Washington University[2]
Carmen Reinhart, International Monetary Fund[3]

The crisis-prone 1990s and the crises of the new millennium have triggered an ever increasing interest in the globalisation of financial turbulences. Many argue that crises are of a regional nature and point to the debt crisis in 1982, which mostly engulfed Latin American countries, and to the so-called Asian flu, which spread from Thailand to Indonesia, Malaysia and the Philippines in a matter of days but left emerging markets in other regions mostly unscathed.[4] Still, the Russian crisis in August 1998 and the debacle of financial markets around the world in autumn 1998 challenge this view and raise the question why some crises are only transmitted regionally while others affect countries around the globe.

Kaminsky and Reinhart (2002) examine whether the degree of globalisation of financial turmoil depends on the origin of the shocks. In particular, they ask whether the extent of the spillover effects depends on whether the shock originates in the periphery or in the centre. For example, were the regional consequences of the Thai crisis so severe owing to Thailand's direct links with other countries in the region or because that shock affected the region's largest economy - Japan? Was the paralysis of the bond markets in many parts of the globe and the persistent equity market volatility due to the Russian default or to concerns that there might be more LTCMs in the making in the financial centres of the world?

There may be various patterns in the propagation of shocks. First, there is the transmission of shocks from one periphery country to another periphery country, which can take place if the two countries are directly linked through bilateral trade or finance. Recent examples of this type of transmission mechanism include the adverse impact of the 1997-98 Asian crisis on Chilean exports and the contractionary consequences for Argentina of the Brazilian devaluation in early 1999. Second, the transmission of shocks from one periphery country to another periphery country may occur through a centre country. There are several prominent examples of this type of transmission mechanism. Corsetti et al (1998) model trade competition among the periphery countries in a common third "centre" market. For instance, Malaysia exports many of the same goods as does Thailand to Japan, Hong Kong and Singapore. Hence, when Thailand devalued in mid-1997, Malaysia lost its competitive edge in the common third markets. Another example of this channel of transmission is analysed in Kaminsky and Reinhart (2000), who focus on the role of commercial banks as lenders in the centre country. For example, US banks had extensive exposure to Mexico in the early 1980s, much in the way that Japanese banks did to Thailand in 1997. The behaviour of the foreign banks can both exacerbate the original crisis, by calling loans and drying up credit lines, and propagate crises by calling loans elsewhere. The need to rebalance the overall risk of the bank's asset portfolio and to recapitalise following the initial losses can lead to a marked reversal in commercial bank credit across markets where the bank has exposure. Third, shocks may be transmitted symmetrically from the centre country to the periphery. This is the type of shock stressed in Calvo et al (1996), who analyse how changes in US interest rates influenced capital flows to Latin America in the early part of the 1990s.

To examine the characteristics of international spillovers, we analysed the daily behaviour of stock markets for 35 emerging-to-mature market countries[5] from January 1997 to August 1999 and

---

[1] This chapter summarises some of the findings in Kaminsky and Reinhart (2002).

[2] Department of Economics, George Washington University Washington, DC 20052, e-mail: graciela@gwu.edu, http://home.gwu.edu/~graciela.

[3] Research Department, International Monetary Fund , Washington, DC 20431, e-mail: creinhart@imf.org.

[4] See, for example, Glick and Rose (1998) and Kaminsky and Reinhart (2000).

[5] The 35 countries in our sample can be classified in five somewhat arbitrary groups: the G7 countries, namely Canada, France, Germany, Italy, Japan, United Kingdom and the United States; transition economies, comprising

examined the degree of globalisation of extreme returns, which were defined as those returns in the 5th and 95th percentile of the distribution of returns. Since we were interested in the centre and the periphery, we examined what happens in stock markets around the world on days of turmoil in financial centres (Germany, Japan and the United States) and on days of turmoil in crisis-prone emerging economies (Brazil, Russia and Thailand).

Our results indicate that turmoil in financial centres is an essential ingredient for systemic turbulences. For example, when there are market jitters in the United States, about 60% of emerging and mature markets worldwide also suffer market jitters. We also find that turmoil in crisis-prone emerging markets spills over into other countries when this turmoil affects financial centres (about 75% of mature and emerging markets worldwide are affected by market jitters in Brazil or Russia when either Germany or the United States are affected by those turbulences). But turbulences in crisis-prone emerging markets such as Brazil or Russia that do not affect financial centres do not have spillover effects worldwide (less than 15 of the countries are affected); they only spill over to other countries in the same region, with about 80% of the countries in Latin America being affected by financial turbulences in Brazil and about 40% of transition economies being affected by turmoil in Russia. That is, for worldwide globalisation of turmoil, financial centres have to be affected. Regional spillovers are different: trade links and wake-up calls may also have a contributing role.

Finally, our research also examines what type of news triggers worldwide turbulences. We find that financial concerns from bankruptcies of large banks or adverse shocks in one particular financial market seem to be at the core of high worldwide globalisation (76% of the episodes). Only 19% of the days of high spillovers seem to be driven by economic news. While financial worries are also at the core of high regional globalisation, their importance is moderate. Only 49% of the episodes of high regional globalisation are driven by financial concerns, with economic and monetary news explaining 37% of the episodes.

While an analysis of more episodes is a clear necessity, one of the preliminary conclusions we draw from this exercise is that to understand "systemic" problems - be these defined at the global or regional level - we have to understand how a shock to the periphery spreads to the periphery (or to other financial centres), via its impact on a financial centre. If the shock never reaches the centre, it is doubtful it can become systemic, irrespective of the definition of systemic that is used. Because financial market participants at the centre countries were largely positioned for the collapse of Ecuador's currency, banking system, economy and political system - not to mention its default on international obligations - these events were more of a ripple in global capital markets than a tidal wave.

References

Calvo, Guillermo A, Leonardo Leiderman and Carmen M Reinhart (1996): "Capital flows to developing countries in the 1990s: causes and effects", *Journal of Economic Perspectives*, vol 10, spring, 123-39.

Corsetti, Giancarlo, Paolo Pesenti, Nouriel Roubini and Cedric Tille (1998): *Structural links and contagion effects in the Asian crisis: a welfare based approach*, Yale University, mimeo.

Glick, Reuven and Andrew Rose (1998): "Contagion and trade: why are currency crises regional?", *NBER Working Paper* no 6806.

Kaminsky, Graciela L and Carmen M Reinhart (2000): "On crises, contagion, and confusion", *Journal of International Economics*.

——— (2002): "The center and the periphery: the globalization of financial turmoil", International Monetary Fund, *Working Paper*.

the Czech Republic, Estonia, Hungary, Poland, Russia, Slovakia and Ukraine; the Asian cluster, consisting of Hong Kong, Indonesia, Malaysia, the Philippines, Singapore, South Korea and Thailand; the "other European" group, which excludes those countries that are part of the G7 and comprises Finland, Greece, the Netherlands, Norway, Spain, Sweden and Turkey; and the Latin American group, which consists of the larger economies in the region: Argentina, Brazil, Chile, Colombia, Mexico, Peru and Venezuela.

Social learning and financial crises

Marco Cipriani and Antonio Guarino, NYU

Introduction

The 1990s witnessed a series of major international financial crises, for example in Mexico in 1995, Southeast Asia in 1997-8, Russia in 1998 and Brazil in 1998-9. These episodes have revived interest among economists in the study of financial system fragility. Theoretical research has analysed various problems, such as bank runs, currency attacks and international contagion. Although many approaches have been taken, two main perspectives have emerged. One part of the literature has emphasised the relation between financial crises and weak fundamentals of the economy. Another has stressed that crises may just be due to random events and self-fulfilling prophecies, with variables unrelated to the real economy acting as "sunspots". Early macroeconomic models of currency crises such as Krugman (1979) (for a survey, see Flood and Marion (1999)) are an example of the first perspective. Microeconomic models of bank runs, such as Diamond and Dybvig (1985), are an example of the second. Rather than alternative explanations, these two views are now considered complementary: financial crises do not occur only in the presence of weak fundamentals, but weak fundamentals can trigger bank run psychology and this in turn can have disproportionately bad effects on the real economy.

To see how difficult it is to reconcile some crisis episodes with a fundamental analysis, let us consider Figure 1, taken from Kaminsky (1999). The chart refers to the case of Malaysia. The solid line is the probability of a crisis estimated using Malaysia's macroeconomic fundamentals. The figure shows that fundamental variables may not be sufficient to forecast the occurrence of a crisis. For instance, the index failed to forecast the crises of the second half of the 1970s and of the second half of the 1990s. In contrast, it forecast a crisis in the mid-1980s, which failed to materialise. This finding is common in much of the empirical work on financial crises. Fundamentals do help to predict when a crisis will occur; nonetheless, crises may occur when the fundamentals seem sound or may not occur when fundamentals are weak.

A possible explanation of why sound fundamentals may not be reflected in asset prices is that information about these fundamentals is spread among economic agents (ie investors) and prices may fail to aggregate it. In particular, this would happen if investors, instead of acting according to their own private information, simply decided to follow the actions of previous traders, ie if they herded. Herd behaviour may, therefore, be a reason why we can observe a misalignment between prices and asset values.

Herd behaviour and information revelation

Several recent models of social learning have shown that herding is not necessarily an irrational phenomenon.[1] The argument was originally made in two seminal papers by Banerjee (1992) and Bikhchandani et al (1992).[2] These papers show that, if people act in sequence and observe the action of their predecessors, the information contained in the history of actions may overwhelm private information. When this happens, agents will disregard their own private information and follow the actions of their predecessors, thus joining a herd.

---

1 We limit our analysis to information-based herding in financial markets and do not discuss herd behaviour due to reputation or compensation schemes. A comprehensive survey of herd behaviour in financial markets is offered by Bikhchandani and Sharma (2001).

2 After these papers, much effort has been dedicated to the topic. See, among others, Chamley and Gale (1994), Chari and Kehoe (2000) and Smith and Sørensen (2000). For a critical assessment, see Gale (1996).

The mechanism can be illustrated with a simple example. Let us consider an economy in which agents can trade a good (ie a financial asset) that can take two values, $0 or $100, with equal probability. In this economy, agents do not trade among themselves, but with a market-maker, who sets the price at which traders can buy or sell the asset (by going short). Let us assume that the market-maker sets a price equal to $50, the expected value of the asset. This price is kept fixed, ie the market-maker does not change it after observing a buy or a sell. Each agent, before making his trading decision, receives some private information (a binary signal) on the value of the asset. This signal has a 70% chance of being correct. Suppose that the value of the asset is $100, but the first two agents arriving in the market receive the wrong (ie low) signal and therefore sell the good. What will the third agent do? Even if his signal is high, he realises that the two previous agents (who sold) had low signals. The negative information contained in the first two sell orders overwhelms his private information. Therefore, he will also sell and will not reveal his (positive) information on the asset value. All the following agents will be in the same position as the third, since they realise that the third agent's action did not depend on his private information. Therefore they will all join the sell herd. Although the value of the asset was $100, everyone will sell and the true state of the world will never be revealed (as it would be, by the law of large numbers, if all agents traded according to their own private information). The actions of the first two agents have a disproportionate and pathological effect on the history of trades.

One of the characteristics of the previous example is that the price does not adjust to the order flow. Indeed, we have assumed that even after a series of buy orders the price is fixed at the level of $50. This is a perfectly reasonable assumption in many economic contexts. For instance, Bikhchandani et al (1992) refer to the choice of adoption of a new technology whose cost is fixed.

In financial markets, however, prices are certainly not fixed. Avery and Zemsky (1998) have shown that, in this case, the argument of Banerjee (1992) and Bikhchandani et al (1992) no longer holds.[3] The presence of a flexible price induces people to follow their own private information since the price adjusts in order to factor in the information contained in the past history of trades.[4]

Let us repeat the example in the previous paragraph. Let us assume, however, that the price is not fixed at $50, but is set equal to the expected value of the asset given the past history of trades. After the first two traders sell, the market-maker will lower the price from $50 to $15.50[5] to take into account that the first two sells came from agents with a low signal. The third agent knows that the two previous traders had a negative signal. If his signal is high, his expected value of the asset will be $30. Given that he faces a price of $15.50, he will buy, ie he will follow his own private information. By the same argument all traders will always follow their private information. Since the signal that they receive is correct 70% of the time, over time the price will converge to the fundamental value of the asset, thus aggregating the private information dispersed among traders. Therefore, when prices are set efficiently, agents will follow their own private information and the price will aggregate the information spread among traders. Consequently, we should not observe misalignments of the price with respect to the fundamentals.

---

3 Avery and Zemsky base their analysis on the Glosten and Milgrom (1985) model of a specialist market.

4 Avery and Zemsky argue that herd behaviour arises in their model when there are multiple dimensions of uncertainty, ie when agents are uncertain not only on the asset realisation, but also on whether an informational event has occurred. Their definition of herd behaviour, however, is different from the one that is standard in the literature and refers more to "swings" of the traders' beliefs. They say that there is herd behaviour when an agent who is originally more pessimistic (optimistic) than the market on the asset value becomes more optimistic (pessimistic) after seeing a sequence of buy (sell) orders. Whereas multiple dimensions of uncertainty make this type of "swing" possible, they do not make informational herding (which the authors call an "informational cascade") possible (see Proposition 2 on page 728 of Avery and Zemsky's paper).

5 The value of $15.50 is obtained by using Bayes's formula.

Explaining rational herds in financial markets

The point made in the previous example is a powerful one. Flexible prices seem to rule out situations in which rational traders choose to disregard their own private information. Given this result, can we still relate the observed price misalignments to rational herding?

In the example, for prices to be able to destroy herds, the traders and the market-maker must value the asset in the same way. In this case, traders find it convenient to use their informational advantage (their private signal) and never herd. The expected utility gained from buying or selling a financial asset, however, may be different across different classes of traders, or between the traders and the market-maker. In other words, a wedge can exist between the way the traders and the market-maker value an asset after observing the same history of trades. When this is the case, traders may decide to disregard their own private information and herd. In the remainder of the section we will briefly describe two papers of ours, in which we analyse an economy where the expectations of traders and market-makers diverge and, as a result, herds arise.

In Cipriani and Guarino (2001b) we show that a divergence between the trader's and the market-maker's valuations can arise when there is uncertainty on the degree of informativeness in the economy (for example, on the proportion of traders who act for informational reasons). Because of these different valuations, even a trader with a negative (positive) signal may decide to buy (sell) because he believes that the asset is undervalued (overvalued). Therefore, there may be situations in which all traders buy or sell independently of the information they have and the price is unable to aggregate the information dispersed among traders. Consequently, the price remains far away from its fundamental value for a period of time longer than if agents always followed their own private information. Eventually, however, the uncertainty on the degree of information in the economy will be resolved (ie it will be learned by the traders and the market-maker) and people will resume trading according to their private information. Therefore, the mechanism outlined in this paper can account for misalignments of the price with respect to the fundamentals, but these misalignments are not long-lasting. There are only a limited number of periods in which people disregard their own private information.

In Cipriani and Guarino (2001a) we consider another source of asymmetry between the traders' and the market-maker's valuations. Different valuations can arise because traders themselves are heterogeneous, ie they may have different degrees of risk aversion or different propensities to save or consume. Different valuations can also be the result of different hedging needs that make some traders more willing to buy an asset, and others more willing to sell it. Differences in the preferences of economic agents are a fundamental feature of markets, which is usually overlooked in the financial market microstructure literature. In many market microstructure papers, agents are assumed to trade only for information reasons (ie because they have a signal about the value of an asset). What characterises markets, however, is that agents are heterogeneous and there are gains from trade. Trade is not driven simply by asymmetric information.

When preferences are homogenous across agents, the price that the market-maker sets is equal to the expected utility that all agents enjoy from the asset. In contrast, when preferences differ across traders and between the trader and the market-maker, the price cannot be equal to the expected utility of each trader. At a given price, some agents will find the asset overvalued, and some will find it undervalued. This wedge between the market-maker's and traders' valuations implies that, for instance, even traders with good information on the value of the asset may decide to sell because the price is simply too high according to the utility that they expect from the asset.

In other words, in a market where traders' preferences are not identical, agents trade not only because they have different information, but also because the asset gives them different utilities. It may happen that this second reason becomes more important than the informational one and traders simply decide to disregard their own private information. In this case, a trade does not reveal anything about the traders' private information, which is therefore not aggregated by the market price. The price remains far away from the fundamental value forever.

In Figure 2, we consider an asset that can take two values, 1 and 2, with equal probability. We show a simulated price path for this asset. Although the realised value of the fundamental is 2, the price converges to a value close to 1. The prevalence of sell orders at the beginning of trading induces all following informed traders to neglect their private information. Given that the market-maker realises that trades are independent of private information, he does not revise the price, which remains stuck at a low value. This price behaviour may explain why, as we discussed in the introduction, models based on the fundamentals may fail to predict the occurrence of a financial crisis.[6]

Financial contagion

Financial crises often spill over from one country to the other, even when these countries are not closely linked. Consequently, asset prices show a correlation in excess of that between the fundamentals. This phenomenon, labelled as financial contagion, is of great relevance, as an economy with sound fundamentals might be affected by a financial crisis just because another economy has been hit. In the 1990s, episodes of contagion were the "tequila" effects of the Mexican currency crisis of 1994-5, the "yellow fever" during the Asian crisis of 1997-8, the asset market crises following the Russian default in 1998 and the Brazilian devaluation in 1999.

We believe that herding can explain why we observe co-movements in asset prices that cannot be accounted for by the fundamentals. In Cipriani and Guarino (2001a) we show that sell orders in one market can affect the price path of another and make its price settle to a lower value. Of course, informational spillovers are to be expected in asset markets, as long as there is some degree of correlation between the fundamentals. We show, however, that these informational spillovers can have pathological consequences. Sell orders in country A not only depress the price of financial assets in country B, where fundamentals are good, but can also cause herd behaviour to arise in this country. Given that herding prevents the revelation of private information, asset prices in country B can remain below the fundamentals even in the long run.

Some evidence

During the 1990s, parallel to the development of the theoretical analysis of herding, many scholars made a significant effort to capture the presence of herd behaviour in empirical data. Starting with the seminal work of Lakonishok et al (1992a), several studies have tried to understand whether different categories of fund managers cluster their decisions (for a survey, see Bikhchandani and Sharma (2001)). These empirical investigations, however, do not estimate any theoretical model of herding, but test whether fund managers cluster their decisions significantly more than one would expect if they acted independently. The reason why there have been no attempts to test a model of informational herding is quite clear. There are no data on the information available to individual traders, and, therefore, it is difficult to gauge whether they disregard their private information when they trade.

An alternative route to test herding models is to gather experimental data. Experimental analysis allows us to test herd behaviour directly because we can control the information available to economic agents. For this purpose we have constructed a laboratory financial market to test whether people tend to imitate their predecessors (Cipriani and Guarino (2002)). In our study, experimental subjects traded in sequence an asset that could take values of $0 or $100. In a situation where all agents traded only for informational reasons and the price adjusted in a Bayesian fashion to the order flow, most experimental subjects did follow their own private information. This seems to show that prices destroyed the incentive of agents to herd. As a result, the price converged to its fundamental value.

When, however, there was a wedge between the expectations of the traders and of the market-maker (for instance, because of non-informational reasons to trade or trade costs), the behaviour of the experimental subjects changed substantially. After the first agents had traded, the following ones stopped trading according to their own private information. Consequently, the price did not always converge to the correct fundamental value. Figure 3 shows the histogram of the last prices (ie the

[6] Note, however, that the price will converge close to 1 more frequently when the fundamental is 1 than when the fundamental is 2. Therefore, when fundamentals are bad, a crisis is more likely to happen. The analysis does not show that the fundamentals are not useful in predicting a financial crisis; rather that some financial crises cannot be predicted looking at the fundamentals.

price after all agents had traded) for the runs of the experiment in which the realised value was 100. The histogram shows that in 10% of cases the price settled on a value far below the fundamental one.

Therefore both the chart in Figure 2 and the histogram in Figure 3 show a behaviour qualitatively similar to the one experienced by Malaysia in the 1970s and 1990s (see Figure 1). That is, in both the theoretical simulation and the experiment we have a financial crisis (ie a large number of periods in which the price is low) that cannot be justified by the fundamentals, but is only due to the inability of the price mechanism to aggregate private information.

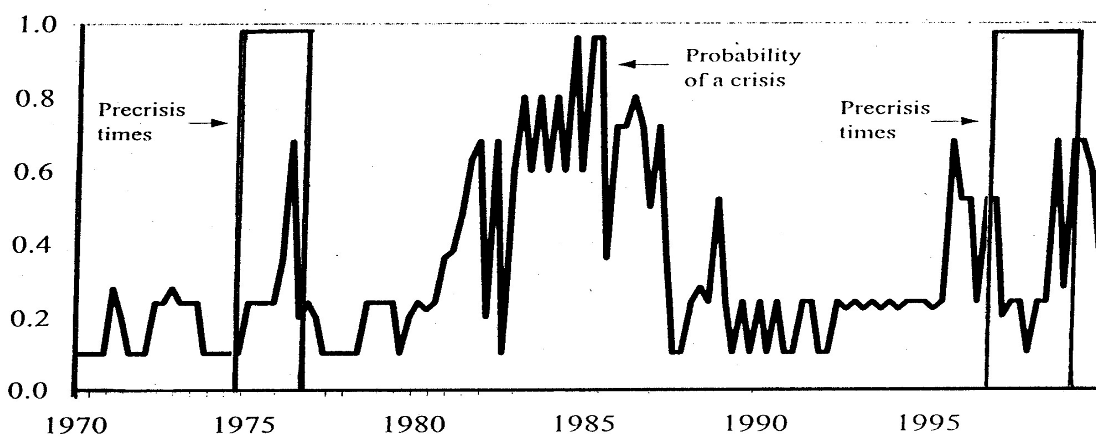
It is the mechanism itself by which prices are formed in financial markets that can explain why we sometimes observe a financial crisis when the fundamentals are good. Even when prices are flexible, rational traders may find it convenient to disregard their own private information. When this happens, the market price may fail to aggregate the information dispersed among traders and long-lasting misalignments may occur.

Conclusions

Our theoretical analysis shows that, in a financial market, the mechanism of price formation may lead traders to disregard their own private information and herd. When this happens, the price does not aggregate traders' information and does not reflect the fundamental value of the asset. Consequently, a financial crisis may occur even when the fundamentals of the economy are sound. Experimental data show that this phenomenon is observed in a laboratory financial market, where experimental subjects choose to disregard their own private information and herd.

Figure 1

Probability of a currency crisis in Malaysia according to fundamentals.

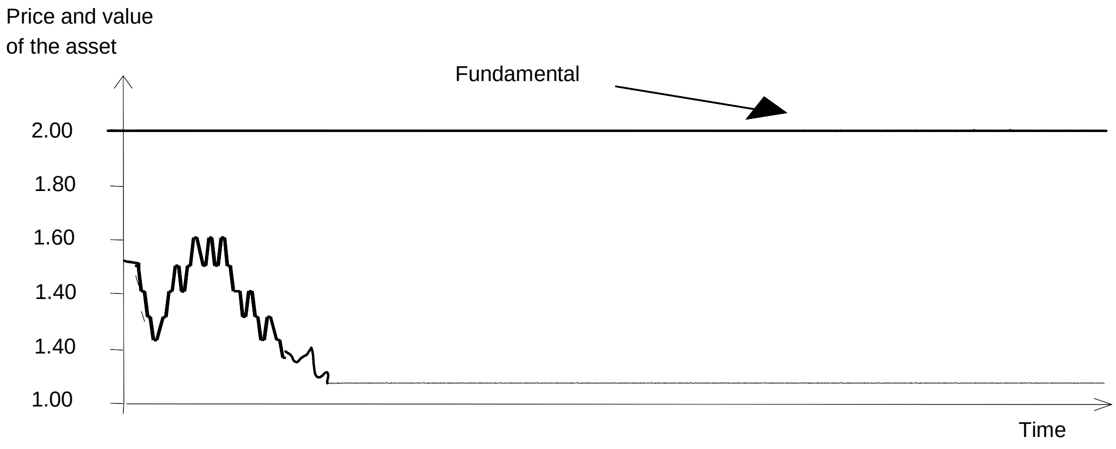


Note: The solid line represents the estimated probability of a financial crisis in Malaysia computed using fundamental variables. The rectangles represent 24-month windows before the occurrence of crises.

Source: Kaminsky (1999).

Figure 2

A simulated path for the asset price

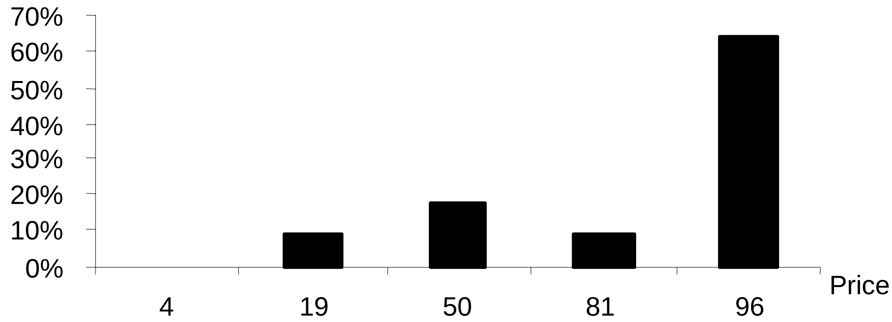


Note: The figure shows the simulated price of a security with a realised value of 2. The price starts from the unconditional expected value of 1.5 and then, after a predominance of sell orders at the beginning of trading, it decreases to a value close to 1. When traders start herding, the price does not change and fails to converge to the fundamental value.

Source: Cipriani and Guarino (2001a).

Figure 3

The histogram of prices in the experiment, V = 100



Note: The figure shows the histogram of last prices in the experimental study for all runs in which the asset value was 100. The last price is the price recorded after all experimental subjects had traded. In 10% of cases, the price converged to a level far from the fundamental value.

Source: Cipriani and Guarino (2002).

References

Avery, C and P Zemsky (1998): "Multidimensional uncertainty and herd behavior in financial markets", *American Economic Review* 88, 724-48.

Banerjee, A (1992): "A simple model of herd behavior", *Quarterly Journal of Economics* 107, 787-818.

Bikhchandani, S, D Hirshleifer and I Welch (1992): "A theory of fads, fashion, custom and cultural change as informational cascades", *Journal of Political Economy* 100, 992-1027.

Bikhchandani, S and S Sharma (2001): "Herd behavior in financial markets", *IMF Staff Paper*, vol 47, no 3.

Chamley, C and D Gale (1994): "Information revelation and strategic delay in a model of investment", *Econometrica* 62, 1065-86.

Chari, V and P Kehoe (2000): "Financial crises as herds", mimeo, University of Minnesota.

Cipriani, M and A Guarino (2001a): "Herd behavior and contagion in financial markets", mimeo, NYU.

—— (2001b): "Herd behavior and informational uncertainty: a note", mimeo, NYU.

—— (2002): "Herding and price convergence in a laboratory financial market", mimeo, NYU.

Diamond, D and P Dybvig (1983): "Bank runs, deposit insurance, and liquidity", *Journal of Political Economy* 91, 401-19.

Flood, R and N Marion (1999): "Perspectives on recent currency crises literature", *International Journal of Finance and Economics* 4, 1-26.

Gale, D (1996): "What have we learned from social learning", *European Economic Review* 40, 617-28.

Glosten, L and P Milgrom (1985): "Bid, ask and transaction prices in a specialist market with heterogeneously informed traders", *Journal of Financial Economics* 14, 71-100.

Kaminsky, G L (1999): "Currency and banking crises: the early warnings of distress", mimeo.

Krugman, P (1979): "A model of balance of payments crises", *Journal of Money Credit and Banking* 11, 311-25.

Lakonishok, J, A Shleifer and R W Vishny (1992a): "The impact of institutional trading on stock prices", *Journal of Financial Economics* 32(1), 23-43.

Milgrom, P and N Stokey (1982): "Information, trade and common knowledge", *Journal of Economic Theory* 26, 17-27.

Smith, L and P Sørensen (2000): "Pathological outcomes of observational learning", *Econometrica* 68, 371-98.

Welch, I (1992): "Sequential sales, learning, and cascades", *Journal of Finance* 47, 695-732.

Session 3

Liquidity I

Positive feedback trading under stress: evidence from the US Treasury securities market

Benjamin H Cohen and Hyun Song Shin,[1]
Bank for International Settlements, London School of Economics

Abstract

A vector autoregression is estimated on tick-by-tick data for quote changes and signed trades of two-year, five-year and 10-year on-the-run US Treasury notes. Confirming the results found by Hasbrouck (1991) and others for the stock market, signed order flow tends to exert a strong effect on prices. More interestingly, however, there is often a strong effect in the opposite direction, particularly at times of volatile trading. Price declines elicit sales and price increases elicit purchases. An examination of tick-by-tick trading on an especially volatile day confirms this finding. At least in the US Treasury market, trades and price movements appear likely to exhibit positive feedback at short horizons, particularly during periods of market stress. This suggests that the standard analytical approach to the microstructure of financial markets, which focuses on the ways in which the information possessed by informed traders becomes incorporated into market prices through order flow, should be complemented by an account of how price changes affect trading decisions.

Introduction

A principal conclusion of the theoretical literature on market microstructure holds that order flow - the sequential arrival of the buy and sell decisions of active traders - plays a vital role in price discovery. In the most influential papers, such as Glosten and Milgrom (1985) and Kyle (1985), order flow plays this role because of the presence of information asymmetries across traders, resulting in adverse selection effects. In Glosten and Milgrom (1985), for example, market-makers do not know whether an incoming order is from an informed or an uninformed trader, and quoted bid and ask prices reflect a trade-off between losses to trading with informed traders and profits to trading with uninformed traders.

By means of a vector autoregression (VAR) analysis of the time series properties of equity price changes and order flows, Hasbrouck (1991) documents a number of apparently robust empirical findings that support the adverse selection approach. Notably, order flow influences prices in the way predicted by the theory. Buy orders raise prices and sell orders lower prices, and there is a component of the price change that may be regarded as the permanent price impact of a trade that remains even after time has elapsed to smooth away transitory effects. Evans and Lyons (2002) document similar findings for the foreign exchange market.

Another robust finding in Hasbrouck's study, however - and one which is relevant for our paper - is that there is also a strong relationship in the opposite direction: from price changes to order flows. Specifically, Hasbrouck finds a strongly *negative* relationship between current order flow and past price changes. In other words, price increases are followed by sales, and price falls are followed by purchases. Given the strong positive effect of past order flow on prices, this relationship between prices and subsequent order flow therefore has a mildly dampening effect on price behaviour.

One of the goals of the present paper is to examine how well the intuitions and models motivated by the stock market and the associated empirical findings translate into another important class of assets:

[1] We are grateful to Marvin Barth, Jon Danielsson, Michael Fleming, Craig Furfine, Richard Payne and Eli Remolona, as well as to seminar participants at the BIS, the LSE, and the 2002 Central Bank Research Conference on Risk Management and Systemic Risk in Basel, for comments and discussions on earlier drafts. We are also grateful to Gert Schnabel for research assistance. All errors, and any opinions that we might express, are our own.

that of fixed income government securities. The market for government securities is important in its own right given its size and benchmark status in the financial market, but we believe that it may also offer some valuable lessons in our understanding of market dynamics that differ from those of the stock market. It is likely that the models motivated by the stock market would fit in less well in those markets, such as for foreign exchange or government bonds, where it is less clear how the theoretical categories can be mapped onto real world variables. The analogue of the "fundamentals" for stocks in the case of treasury securities corresponds to broadly macroeconomic considerations, and it seems less easy to tell a plausible story of a subset of (private sector) traders having strictly better information about these fundamentals than the others.

To a significantly greater extent than for equities, the fixed income (and foreign exchange) pages of the financial press as well as the commentary from traders themselves abound in strategic trading terms such as overhangs of leveraged positions, short covering and the like. This suggests that these strategic interactions between traders may result in market dynamics that differ from those in markets, such as equities, that conform to the adverse selection-based models of market microstructure.

Our objective in this paper is to investigate whether this intuition can be substantiated from the market data. We take the VAR methods used by Hasbrouck (1991) and apply them to the US Treasury securities market. Our conclusions point to some interesting and revealing differences from Hasbrouck's original results for the stock market. To anticipate our main findings, we find that:

– under tranquil market conditions, when trading is orderly and trading frequency is low, most of the qualitative conclusions found for the stock market are replicated. The key difference is that, whereas Hasbrouck found that past price changes generally have a negative effect on order flow, we find this only to be the case for the 10-year note. For the two- and five-year notes, the effect is significant and positive;

– however, during periods of high price volatility and active trading, there appears to be a structural shift in the market dynamics. In such periods, the positive effects of past order flow on current prices, *and vice versa*, are reinforced. In other words, not only do buy orders elicit higher prices, but price increases in turn elicit more buy orders. As a result, price movements become more positively autocorrelated (or less negatively autocorrelated) at short horizons. This is the case even though signed trades tend to become slightly *less* positively autocorrelated during such periods.

The structural shift in market dynamics to positive feedback trading is detectable even during a single day's trading, and coincides with bursts of intense trading activity. The onset of frenetic trading is accompanied by rapid price changes and a heavily one-sided order flow. We illustrate this effect by examining in some detail the particularly volatile trading on 3 February 2000, when markets were unsettled following the US Treasury's announcements on debt management policy and rumours about large losses at certain institutions.

Positive feedback trading is consistent with the market adage that one should not try to "catch a falling knife" - that is, one should not trade against a strong trend in price. Some recent empirical studies are also consistent with such behaviour. Hasbrouck (2000) finds that a flow of new market orders for a stock is accompanied by the withdrawal of limit orders on the opposite side. Danielsson and Payne's (2001) study of foreign exchange trading on the Reuters 2000 trading system shows how the demand or supply curve disappears from the market when the price is moving against it, only to reappear when the market has regained composure.

One way of understanding these patterns of trading is in terms of the constraints on traders that shorten their decision horizons and thereby encourage mutually reinforcing behaviour. Among these constraints might be position limits, risk management rules, or margined positions. For any of these reasons, a trader might be obliged to liquidate her position when prices move against her. If some traders believe that others will be faced by such constraints, they may attempt to anticipate the results of a sharp price move or magnify the trading profit of riding short-term price trends by selling into a falling market or buying into a rising one.

The next section describes the data set used and applies a VAR specification to intraday trading in on-the-run US Treasury notes over the period 1999-2000. Section II examines trading on an especially volatile day in some detail, as a means of illustrating how price and transaction behaviour can shift suddenly in volatile trading conditions in ways that cannot be fully explained by an approach based on adverse selection and order flow.

Providing a theoretical basis for an explanation of this kind of positive feedback trading is an important unresolved task. It is not our objective in this paper to tackle this important issue, but we will identify the possible ingredients of such a theory in Section III. We suggest an alternative (and to some degree complementary) theoretical approach that relies on the strategic interactions among traders. Section IV concludes.

I. Testing for strategic interaction among traders

A. The data

The data are provided by GovPX, Inc. GovPX provides subscribers with real-time quotes and transaction data on US Treasury and agency securities and related instruments compiled by a group of inter-dealer brokers, including all but one of the major brokers in this market. For each issue, GovPX records the best bid and offer quotes submitted by primary bond dealers, the associated quote sizes, the price and size of the most recent trade, whether the trade was buyer- or seller-initiated, the aggregate volume traded in a given issue during the day, and a time stamp. Dealers are committed to execute the desired trade at the price and size that they have quoted to the brokers. However, counterparties can often negotiate a larger trade size than the quoted one through a "work-up" process. Fleming (2001), who provides an extensive description of this data set, estimates that the trades recorded by GovPX covered about 42% of daily market volume in the first quarter of 2000.

We examine quotes and trades in two-year, five-year and 10-year on-the-run (ie recently issued) Treasury notes over the period from January 1999 to December 2000. Although GovPX provides round-the-clock data, we restrict the series to quotes and trades that take place between 7 am and 5 pm, when trading is most frequent. The quotes used are the midpoint of the prevailing bid and ask quotes. When a new issue becomes "on the run", the GovPX code indicating on-the-run status switches to the new issue starting at 6 pm; this means that a given set of *intraday* quotes and trades will always refer to the same issue. Trade volumes are calculated as the difference in the aggregate daily volume recorded for the corresponding security. Because these figures are provided in chronological order, the result is an ordered data set in which each observation is either a quote change, a trade or both.

Table 1a summarises the data used for the three securities. Our observations are in "event time" rather than chronological time. One issue is whether the tick-by-tick returns should be normalised so that they are comparable to calendar returns over a fixed time interval. Our main qualitative results turn out to be insensitive to whether we normalise or not. For the results to be reported below, returns (r_t) are defined as the difference in the log of the quoted price (more precisely, the midpoint between the prevailing bid and ask quotes) at event times t and t-1.

The number of observations increases with maturity, while the number and size of transactions fall. In other words, the data set includes more quote changes and fewer transactions as maturity rises. During the sample period, an average of $4.6 billion of trades are recorded daily on GovPX for the two-year note, more than the five-year ($2.5 billion) and 10-year ($1.6 billion) combined, reflecting both more trades and a greater volume per trade. As suggested by Fleming (2001), this may reflect differences in coverage by GovPX rather than differences in the actual relative liquidity of two-, five- and 10-year issues, since the excluded broker (Cantor Fitzgerald) is relatively more active in longer-term issues. The mean absolute value of the return from one observation to the next rises with maturity.[2] The same is true for daily returns.

Table 1a also gives the average duration (the time between observations) for the full sample of each bond and for four subsamples. This is about one minute for the two-year note, and about 45 seconds for the five- and 10-year notes. For the 50 trading days where average duration is highest, the time gap

---

[2] In terms of 32nds, which are the usual quote convention for Treasury notes, and assuming a price close to 100, the mean absolute returns shown correspond to price changes of 0.09 32nds for the two-year, 0.17 32nds for the five-year, and 0.32 32nds for the 10-year note.

is slightly less than two minutes for all three notes, while for the 50 trading days with the lowest average duration, this gap is about 40 seconds for the two-year note and 30 seconds for the five- and 10-year notes. This suggests that, while there clearly are more active and less active trading days in the sample, divergences in the frequency with which quotes and/or trades are observed are not great.

Average durations are also presented for the 50 days where the difference between the daily high and low price (the daily trading range) for the specified bond is highest, and for the 50 days where this difference is lowest. We would expect days in the former sample to correspond to relatively volatile trading conditions, while days in the latter are relatively quiet. Again, a clear difference between the two samples in terms of average duration can be observed. Days with wide price swings tended to see more frequent trades and/or quote changes, with observations coming in every 40 to 45 seconds, than quieter days, when the time between observations averaged 92 seconds for the two-year note and 56 seconds for the 10-year. Duration is also longer for low-volatility days (measured by the standard deviation of the tick-by-tick returns) than for high-volatility days.

Confirmation of the relationship between the frequency of trading and various volatility measures is presented in Table 1b for the two-year note. The average duration on a given day tends to be negatively correlated with the range (high-low) of prices observed during the day, and the standard deviation of tick-by-tick returns during the day, while the price range and volatility display a strong positive correlation. None of these variables seems to have a strong correlation with the change (open-close) in prices that occurred during the day, suggesting that trading conditions were about as volatile on days when bond prices rose as on days when prices fell.

B. Testing for the cross-effects of trades and quote revisions

B.1 What might the data tell us?

GovPX records the pricing and trading decisions of bond dealers, rather than those of speculative traders or long-term investors. A reasonable assumption is that the dealers participating in the system attempt to minimise their open exposures to bond yields as far as possible, and do not attempt to take a "view" on likely yield movements.[3]

Under this assumption, when a dealer accepts a bid or offer that has been posted on the system, he could be following one of two possible behavioural rules. One is that, whenever the dealer executes a trade with a customer, either by selling her a bond out of inventory or by buying a bond from her, the dealer immediately submits a countervailing trade to an inter-dealer broker in order to remain balanced. The other is that the dealer only rebalances his exposure periodically. Under the first rule, a transaction observed in the GovPX data closely tracks the transaction decision of a position-taker in the market. Under the second, an observed transaction primarily reflects inventory control operations and not a position-taking decision, except in the sense that a series of position changes should eventually (after several minutes or a few hours) lead to a corresponding inventory adjustment transaction. To the extent that both of these motivations are in action, the dealer-submitted transactions compiled by GovPX are likely to reflect a combination of the speculative strategies of traders and the inventory control strategies of dealers.

The quotes posted on the system are also likely to reflect a combination of speculative and inventory control motives. At certain times, a dealer may adjust his posted bid and ask quotes because of the information that he has gleaned from customer order flow. At other times, he may "shade" posted bid and ask quotes in order to induce a sufficient number of buy or sell orders to bring inventory back into line with its desired level. Both categories of motives are likely to influence the posted quotes that we observe on GovPX.

A primary aim of the analysis of intraday financial market data is to understand how the microstructure characteristics of a given market affect the time series characteristics of price quotes, signed

[3] Some dealers, however, execute trades on behalf of proprietary trading desks under the umbrella of the same financial institution. For the purposes of this discussion, a proprietary trading desk would be thought of as a "customer" of its affiliated dealer. During the time period covered by this study, January 1999 to December 2000, many of the major government bond dealers had either closed or seriously curtailed their proprietary trading operations.

transactions, and the interactions between them. If the dealers whose quotes and trades are recorded by GovPX are primarily mimicking customer orders, then this would allow us to test for the informational interaction between prices and trades. Specifically, we could test the result in the theoretical literature on market microstructure noted above, namely that signed order flow should have a measurable impact on price formation. We could also test whether, for reasons that will be discussed in more detail in Section III, lagged price movements have an impact on trading under certain conditions.

Further, there are reasons to believe that the time series of both order flow and returns themselves exhibit serial dependence. Among the factors that might produce such dependence are inventory control motives, lagged adjustment to incoming information, and minimum tick sizes. Some of these factors would result if dealers followed a customer-driven rule, while others would imply the primacy of inventory adjustment in short-run dealer behaviour.

At a short enough time horizon - data observed in intervals of minutes and seconds, rather than days or months - one might expect these factors to exert an impact on observed quotes and trades that can be measured statistically, even if at longer time horizons price changes are thought of as being driven more or less exclusively by the arrival of new information. Examining prices and trades over very short intervals of time could thus enable us to determine which rules are being followed by the dealers in the market and, if we think the mimicking of customer orders is important, to learn more about customer behaviour as well.

B.2 A two-variable VAR of signed trades and returns

The following vector autoregression (VAR) should capture many of these short-horizon effects:

$$
\begin{aligned}
r_t &= \sum_{i=1}^{10} \alpha_i r_{t-i} + \sum_{i=0}^{10} \beta_i \, trade_{t-i} + \varepsilon_{1,t} \\
trade_t &= \sum_{i=1}^{10} \gamma_i r_{t-i} + \sum_{i=1}^{10} \delta_i \, trade_{t-i} + \varepsilon_{2,t}
\end{aligned}
\qquad (1)
$$

Here r_t is the return variable cited above, while $trade_t$ is a signed trade variable. Two variables are used for $trade_t$:

x_t, an indicator variable equalling 1 for a buyer-initiated transaction, -1 for a seller-initiated transaction, and 0 where there is a change in the price quote without a transaction; and

v_t, the size of the trade in millions of dollars, multiplied by 1 for a buyer-initiated transaction and -1 for a seller-initiated transaction.

The version using x_t is essentially identical to the VAR computed by Hasbrouck (1991). Like Hasbrouck we estimate the contemporaneous impact of trades on prices. That is, we include a term $\beta_0 trade_t$ on the right-hand side of the first equation. This allows for the possibility that trades are "observed" slightly before quote revisions, for example through the work-up process.[4] Although the estimate of β_0 is positive and significant in all versions of the VAR that we examine, excluding the contemporaneous $trade_t$ from the estimation of the first equation produces qualitatively similar results.

Results from the estimation of equation (1) on the full two-year sample are presented in Table 2 for $trade_t = x_t$, and in Table 3 for $trade_t = v_t$. For each trading day, the calculation of the VAR starts with the 11th observation of the day as the dependent variable. This eliminates the above-mentioned effect of the switch from one on-the-run issue to the next, the influence of overnight price changes and the inclusion of the effects of the last few observations in one day on the first few observations in the next.

For three of the four "quadrants" of coefficients - the effects of lags of r_t on r_t; the effects of contemporaneous and lagged $trade_t$ on r_t; and the effects of lags of $trade_t$ on $trade_t$ - there is a remarkable degree of consistency across the three maturities (two-year, five-year and 10-year) and

[4] In January 2000, the average length of the work-up process was 20.97 seconds for the on-the-run two-year note, 16.12 seconds for the five-year note and 17.86 seconds for the 10-year. These are all less than the average tick lengths, which were 59, 46 and 44 seconds respectively. Boni and Leach (2001) describe and analyse the work-up process in the US Treasury market.

across the two trade variables (x_t and v_t). The results for all three quadrants conform to those found by Hasbrouck (1991) for the US equity market.

- Lagged returns tend to exert a negative effect on present returns, though this effect is partially reversed in later lags. In other words, returns are negatively autocorrelated at very short time intervals. Although we use quote midpoints to calculate r_t, even for observations where the new line of data represents a new transaction (that is, we use the prevailing quotes rather than the transaction price), it is possible that the negative autocorrelation reflects a "bid-ask bounce" effect as described by Roll (1984). Engle and Patton's (2000) study for NYSE stocks shows that the price impact of an order falls asymmetrically on the bid and ask quotes. Buyer-initiated trades primarily move the ask price while seller-initiated trades move the bid price. When one side of the quote is updated more quickly than the other in response to an order, the midquote would exhibit behaviour similar to the bid-ask bounce.
- Current and lagged trades tend to exert a positive effect on present returns. In other words, price movements follow order flow. Besides Hasbrouck's findings for the equity market, similar effects have been found for the treasury market by Fleming (2001) and for the foreign exchange market by Evans and Lyons (2002).
- Lagged trades tend to exert a small but significantly positive effect on current trades. In other words, trades are positively autocorrelated. This may suggest that traders tend to adjust their positions in a series of trades, rather than all at once, or that some traders respond to new information with a lag.

It is in the "upper right" quadrant - the effect of lagged returns on current signed trades - where the consistency breaks down somewhat across maturities, and where the results are generally different from Hasbrouck's. For the two-year and five-year notes in the VARs using x_t, and for all three maturities in the VARs using v_t, the coefficients on lagged returns (sometimes with the exception of the first lag) tend to be *positive* for current trades. In other words, price increases tend to be followed by buy orders, at short horizons, while price decreases are followed by sell orders. Only for 10-year notes in the VARs using x_t are the coefficients generally negative, corresponding to Hasbrouck's results for the equity market. This set of effects will be the focus of Sections II and III of the paper.

B.3 Estimating cumulative effects

A standard tool for analysing the results of VARs is the impulse response function. In the present case, however, we are interested not in the usual impulse response function - the effect on the level of one of the variables at some future point from a shock to a variable in the system - but in the *cumulative* effects of shocks to the included variables. Thus, for example, we want to know the impact of a new buy order on the overall return over the next several minutes, rather than on the level of the observed return at a specific point in the future. Similarly, we want to know the total number of net buys or sells that happen in the aftermath of a new buy or sell.

To do this, we can cumulate the output of the usual impulse response function, taking account of the presence of the contemporaneous signed trade as an explanatory variable in the return equation. To construct the orthogonalised shocks to signed trades and returns, we need to make a prior assumption about the direction of causality between the variables. In this case, we assume that signed trades "cause" returns.

Graphs 1 to 4 show the cumulative effects of a one-unit increase in returns and buys (the x_t variable) on the cumulative return and number of net buys over the following 20 periods for the two-, five- and 10-year Treasury notes.

The graphs largely confirm the results identified in our earlier review of the signs of the respective raw coefficients. Roughly 77% of a given shock to the return of the five-year note is still contained in the price level 20 periods later; this proportion falls to 69% for the two-year and 68% for the 10-year note (Graph 1). A buy order has a strong positive effect on returns in the short term; a buy causes a cumulative positive return of about 0.27 hundredths of a percent for the two-year note, 0.63 hundredths of a percent for the five-year note, and 1.05 hundredths of a percent for the 10-year note (Graph 2). In the 20 observations after a net buy order is recorded, a further 0.74 net buys result for the two-year note, 0.60 net buys for the five-year, and 0.38 for the 10-year (Graph 4).

As maturity increases, there seems to be a greater impact of trades on returns and less positive autocorrelation of trades. One possible explanation of this is the relatively lower fraction of the market covered by the data at higher maturities. It is likely that returns are influenced not only by the trades executed by the brokers participating in GovPX, but also by those executed by the excluded broker; hence the impact of a trade on the observed return is overestimated when one looks only at GovPX trades. Similarly, the autocorrelation of trades is underestimated, because one is looking only at a fraction of the actual trades in any given period of time. There do not appear to be strong differences across maturities in the pattern of autocorrelation in returns.

The cumulative impact of returns on trades, which as already noted differs strikingly from Hasbrouck's results, is illustrated in Graph 3. The graph shows the impact of a one-unit increase in the return. When one considers the typical size of these returns, it becomes clear that the magnitude of the effect is not large. For the two-year note, for example, an increase of one standard deviation in the return (a return of 4.46×10^{-5} from one tick to the next, or about 0.4 hundredths of a percentage point) leads to the occurrence of 3.7% more net buys than would otherwise take place over the subsequent 20 periods, or roughly 19.6 minutes.[5] For the five-year note, there are 3.5% more net buys when the return rises by one standard deviation. However, the fact that the coefficients from the underlying VARs are significant suggests that this is more than a statistical artefact. For the 10-year note, the cumulative effect on x_t is negative, with net buys falling by 1.5%.

C. Estimation results for duration-based subsamples

More interesting than the size of these effects is the way they change over different subsamples. The lines in Table 4 labelled "Low duration" show the effects estimated from a VAR similar to that in equation (1) for the days on which the average adjusted duration is unusually low. These should be the days of relatively hectic trading (and indeed, as already noted, price volatility and the differential between the daily high and low tend to be highest on these days). Similarly, the "High duration" lines show the estimated cumulative effects on days when average adjusted duration was unusually high. These should be days when trading and changes in quotes are relatively slow, suggesting quiet trading conditions.

More precisely, the tables show the sums of different combinations of coefficients from the following VAR:[6]

$$
\begin{aligned}
r_t &= \sum_{i=1}^{10}\left(\alpha_i + \alpha_i^L d_{t-i}^L + \alpha_i^H d_{t-i}^H\right) r_{t-i} + \sum_{i=0}^{10}\left(\beta_i + \beta_i^L d_{t-i}^L + \beta_i^H d_{t-i}^H\right) x_{t-i} + \varepsilon_{1,t} \\
x_t &= \sum_{i=1}^{10}\left(\gamma_i + \gamma_i^L d_{t-i}^L + \gamma_i^H d_{t-i}^H\right) r_{t-i} + \sum_{i=0}^{10}\left(\delta_i + \delta_i^L d_{t-i}^L + \delta_i^H d_{t-i}^H\right) x_{t-i} + \varepsilon_{2,t}
\end{aligned}
\qquad (2)
$$

The dummy variable d_t^L takes the value of 1 when an observation occurred on one of the 50 days (10% of the sample) when duration, adjusted for time-of-day, seasonal, and time trend factors, was at its lowest, while d_t^H is 1 for observations on the 50 days when adjusted duration was highest. Table 4 also gives the significance levels for different combinations of variables, using a Wald test for the hypothesis that this sum is different from 0.

The duration-based subsamples are determined using an adjusted measure of duration. This adjusted duration equals the ratio of the actual duration to the fitted values from a model that estimates duration using time-of-day, time-of-year, and trend effects. The model closely resembles the linear spline model with "nodes" at the top of each hour developed in Engle (2000). We include a time trend in the estimation in order to account for the fact that the number of observations tends to decline throughout the sample period, reflecting the steadily declining share of US Treasury market trading that is covered by the data. We also add dummy variables for observations in November and December, two months when these markets are less active. The result is a series of fitted duration estimates for each Treasury note studied. The values of these fitted estimates, when graphed over the trading day, exhibit a distinct "U"-shape (Graph 5). Activity is very slow between 7 and 8 am, then speeds up dramatically

---

5 More precisely, the fraction of total transactions in the next 20 periods that are buys is 0.037 higher than it otherwise would have been.

6 To save space, the coefficients from this and the other VARs in the remainder of the paper are not given. Coefficients from these VARs are available from the authors.

between 8 and 9 am, when the most closely watched economic statistics tend to be released. The market then slows somewhat, but remains active until 3 pm, after which transactions and quote changes dwindle. Adjusting duration by dividing it by these fitted values results in a time series of duration "surprises".

For all three maturities, the effects of trades on returns tend to be higher on the low-duration days than on the high-duration days or on the days when duration was neither unusually high nor unusually low. These effects do not change in a significant way, however, when one compares unusually high-duration days to "normal" days. This suggests that the structural change may be non-linear: low-duration days stand out but high-duration days do not.

Effects in the opposite direction - from returns to subsequent trading behaviour - also shift on high- and low-duration days relative to the rest of the sample. For the two-year note, these effects are more strongly positive on low-duration days than in normal times (that is, they lead to more net buys), though the Wald test does not support the hypothesis that this change in the variables is significant. On high-duration days, however, the effects become insignificant in a statistical sense, and a Wald test supports structural change at an 8% significance level. For the five-year note, the results are qualitatively similar: there is no statistical difference between effects on low-duration and "normal" days, while the effects become insignificant on high-duration days. For the 10-year note, it will be recalled that positive price movements cause an increase in net selling in the sample as a whole. These effects, as well, become insignificant on high-duration days.

Impulse response functions for the different subsamples are illustrated for the two-year note in Graphs 6a-6d. For the cross-effects of signed trades on returns and returns on signed trades, these confirm what was observed from looking at the raw coefficients in Table 4. Whereas a new buy leads to an increase of 0.27 hundredths of a percent in the cumulative return after 20 periods in the sample as a whole, on low-duration days the impact rises to 0.40 hundredths of a percent, while on high-duration days it falls to 0.23 hundredths of a percent (Graph 6b). Effects in the opposite direction grow stronger as well. For the sample as a whole, it will be recalled that an additional standard deviation return results in an increase of 3.7% in the number of buy orders in the next 20 periods. On low-duration days, this rises to 5.3%, while on high-duration days net buys decline by 0.7% (Graph 6c).

This increase in the mutual impact of trades and returns on one another results in an increase in the persistence of shocks to returns. For the full sample, 69% of a shock to the quote midpoint remains in the price after 20 periods. On low-duration days, this proportion rises to 86%, while on high-duration days it falls to 62% (Graph 6a). However, the impact of a new trade on the direction of trading does not change appreciably across the different subsamples (Graph 6d).

II. A case study: 3 February 2000

The results in Section I suggest that, on days of relatively rapid trading activity, traders tend to reinforce price movements (at least at short time horizons) rather than dampening them. This section explores the dynamics of this shift on a very volatile trading day that occurred during the sample period.

A. Events of 3 February

3 February 2000 witnessed the sixth highest daily trading range for the on-the-run two-year note in the sample period (Graph 7). The price quoted on GovPX (using the average of the prevailing bid and ask quotes) for the two-year note opened at 99.551 at 7.04 am, reached a low of 99.523 at 10.03 am, rose to a high of nearly 99.977 at 12.36 pm, and finished at 99.727 at 5 pm. The range of the price from its lowest to its highest point, 0.45% of par, is very large in comparison with the sample median daily price range of 0.12%, the mean absolute value of the daily price change (open to close), 0.07%, and the standard deviation of the daily price change, 0.09%. This price range corresponds to 85 basis points in yield, in comparison with a median daily yield range of 23 basis points.

News accounts of the trading on 3 February, a Thursday, do not point to a specific new piece of macroeconomic information being digested by the market. The market was reported to be unsettled by the US Treasury's plans to change its auction practices and repurchase selected issues as part of a broader policy of using budget surpluses to reduce the debt held by the public. A key piece of public information relevant to that policy had been released on 2 February, when the Treasury outlined plans to reduce the amounts of specific maturities to be issued in future auctions, including the popular 30-year bond. This announcement came during trading hours on the 2nd, so it was no longer fresh news to the market on the 3rd. Nevertheless, market commentary relating to trading on the 3rd focused on the uncertain environment created by the previous day's announcement. In its daily report on the US Treasury market, the Associated Press emphasised the uncertain implications of the new Treasury programme on the liquidity of the 30-year bond, and the effects this uncertainty had had on market trading. According to one fund manager:

> Folks are kind of shocked. Treasuries have become a scarce commodity. … It's "wild, wild stuff", as Johnny Carson used to say. It's definitely a new environment for everybody. We're all trying to figure out what this means for the future. (AP Online, 2000)

In the same article, the Associated Press noted another series of events which may have influenced trading on 3 February:

> Adding to Thursday's mayhem was a widespread rumor that the dramatic decline in bond yields had wiped out a large unnamed financial institution and that a rescue meeting was being held at the Federal Reserve Bank of New York. The rumor prompted a statement from the New York Fed denying there was a meeting to discuss market volatility. (AP Online, 2000)

An item released on the Market News International Wire at 12.14 pm on that day reads in its entirety:

> NEW YORK (MktNews) - A spokesman for the Federal Reserve Bank of New York Thursday declined all comment on a rumor widespread in financial markets that there would be an emergency meeting at the Fed to address big losses at a financial firm.
>
> The spokesman said it is Fed policy not to comment on such rumors.
>
> The completely unsubstantiated rumor circulated all morning Thursday, and appeared related to the market dislocations triggered by the Treasury's plans to cut back on supply of long-term securities. That has resulted in an inversion in the Treasury yield curve in recent days and a huge rally in Treasury long bonds Wednesday and Thursday.[7]

3 February thus seems to offer an excellent opportunity for a case study of patterns of trading in the US Treasury market under conditions of uncertainty. With the exception of the Fed's announcement denying the rumour, there was no occasion when a piece of price-relevant information simultaneously became known to all participants. Instead, there was uncertainty as to how markets themselves would be expected to behave in the new environment of shrinking supply. The rumours of an institution in trouble added to the uncertainty, but undoubtedly, as tends to happen in these situations, the main area of uncertainty for market participants was the nature and extent of the knowledge possessed by *other* participants.

Examination of Graph 7 suggests that the day can be divided into four periods in terms of trading behaviour. Characteristics of these periods, and comparable figures for the full two-year sample, are presented in Table 5. From 7 to 11 am, prices were flat or slightly higher, bid-ask spreads were wider than usual but steady, duration was somewhat shorter than usual, and there was a roughly even balance between buys and sells. From 11 am to 12.15 pm, prices rise sharply, accompanied by an imbalance of buys over sells and a shortening of duration. This is presumably the time when rumours about a troubled institution dominate market trading, with prices at first bid up on the expectation that the institution would have to close out a large short position. From 12.15 to 2 pm, prices fall about as sharply, with sells outnumbering buys and duration remaining very low. This followed the New York Fed announcement. In both the second and third periods, quoted bid-ask spreads are wide and volatile, and occasionally negative.[8] Finally, from 2 to 5 pm, prices rise gradually amid relatively calm conditions, with duration close to normal levels, though bid-ask spreads remain elevated.

---

[7] We are grateful to Michael Fleming for calling our attention to this news story.

[8] Both the very wide and the negative bid-ask spreads are probably the result of "stale" quotes that dealers did not have time to update.

Two points are worth noting with regard to Table 5, both of which suggest that the bond market on 3 February behaved in a more complex way than would be implied by a simple adverse selection model in which information is incorporated in order flow.

First, while it is clear that an imbalance of buy orders over sell orders was associated with rising prices and vice versa, it is interesting that a virtually identical share of buys (66%) led to a sharp price increase between 11 am and 12.15 pm, but to only a relatively mild price increase between 2 and 5 pm.

Second, the bid-ask spread was at its highest between 12.15 and 2 pm - even though, as noted above, the Fed announcement was probably the day's most influential piece of *public* information. If wide bid-ask spreads indicate a high degree of information asymmetry, as the adverse selection model would predict, one would expect that when an important item of news, with a direct and immediate bearing on market prices, becomes known simultaneously to all market participants, this would contribute to a significant *narrowing* of bid-ask spreads.

B. Price movements and order arrival: a closer look

A closer examination of trading patterns throughout the day presents further puzzles (Graphs 8a-8d). It is worthwhile, first, to consider what the different theoretical frameworks used in market microstructure would predict about the patterns of price movements and orders. A pure neoclassical view would suggest that the price moves automatically to adjust to new information, and that buys and sells should be essentially balanced whatever the price level is and in whatever direction it is moving. If orders primarily reflect inventory adjustment, then groups of buys and sells should alternate, with a large number of buys leading to price increases (as dealers rebuild inventory) and sells leading to price decreases (as they lay off inventory) in an essentially predictable rhythm. According to an adverse selection-based view, we would expect to see an exogenous build-up of purchases to be followed more or less immediately by information-driven price increases, and a build-up of sales to be followed by price declines.

During the 7 to 11.30 am period (Graph 8a), buys and sells appear to be balanced over the period as a whole, but do not seem to follow any of these predictions closely. Rising prices are associated with buys (eg just after 10.04 am) and declines with sells (eg just before 8.18 am). But the order flows and price movements appear to be simultaneous; the price graph does not wait for a build-up of orders before it starts moving. And periods of persistent one-sidedness in the market (eg the buying activity from 10.17 until around 10.40 am) are not followed by price movements that would be sustained enough to return inventories to balance; instead, on this occasion, the price hovers for a while, then turns downwards - and only then (around 10.44 am) do we see clusters of sales.

As the rumours of a troubled institution begin to take hold (Graph 8b), the price rises amid heavy buying. But sometimes the price rises with little or no buying, as in the phase just after 11.46 am, and again around 12.12 pm. At the very top of the market, from around 12.15 pm onwards, traders appear to be buying at peaks, and selling at valleys. Again, neither the neoclassical, nor the inventory adjustment, nor the adverse selection view appears to explain the interaction between price and order behaviour.

The period after the Fed announcement (Graph 8c) is virtually the mirror image of the hour or so that preceded it - this despite the very different nature of the information that was driving the market in the two periods, with rumours replaced by credibly stated facts. Prices sometimes fall without any order flows, other times associated with heavy selling. Prices seem to stabilise around 1.05 pm, even though traders continue to sell. A cluster of buys eventually emerges just before 1.16 pm, but the market seems happy with its new level - even when the buys are followed by further sales.

During the last three hours of the trading day, the market rises slowly and without much volatility (Graph 8d). A heavy series of buy orders does not do much to move the price. These may derive from traders covering short positions entered into during the previous phase, or they may represent the rebuilding of inventory by dealers (though an examination of *cumulative* order flow, not shown here, would cast doubt on this).

For an example of an alternative kind of price volatility, consider the trading pattern for the two-year note on the morning of 28 January 2000 (Graph 9). In this case new information - an unexpectedly strong non-farm payroll figure - became instantaneously available to virtually all market participants when the data were released at 8.30 am. Trading appears to have reflected first the anticipation of, then the accommodation to, this new information, while virtually no trades took place when the announcement was being made. While some position-taking in anticipation of the announcement moved the price somewhat, in the aftermath of the announcement trades tend to have little or no impact on the price, perhaps because participants understand that this represented the squaring of speculative positions and the rebalancing of portfolios. Trading volume is much higher after the announcement than before, as can be seen in the shorter time intervals between the times indicated on the x-axis (which are spaced 50 ticks apart). This pattern of the adjustment of Treasury prices to information releases conforms to similar findings by Fleming and Remolona (1999a) and Huang et al (2001).

C. VAR analysis

Graphs 10a-10d illustrate estimations of the cumulative effects of returns and signed trades on one another, and of returns on subsequent returns, when the VAR in model (1) is applied to prices and trades recorded for the two-year note on 3 February 2000. Because there are fewer data points, five lags are used in each equation instead of 10. As before, the impulse response graphs assume that causation runs from trades to returns. Sums of coefficients for the different time periods for the two-, five- and 10-year notes are provided in Table 6. In what follows, we will focus on the results for the two-year note.

Cross-effects between trades and returns seem to have been stronger on 3 February than they were during the full two-year sample period. The impact of trades on returns is about twice as strong on 3 February as during the full sample, with a new buy order leading, on average, to an increase of 0.53 hundredths of a percentage point in the return (Graph 10b). The effect of returns on trades is also substantially higher than normal on 3 February: a one standard deviation positive return now leads to a 5.2% increase in the likelihood of a purchase after 10 periods, more than 50% higher than the effect estimated for the sample as a whole (Graph 10c). The persistence of shocks to returns is also stronger. Ten periods after a positive shock to the return, 77% of the increase remains in the bond price, compared with 69% for the sample as a whole (Graph 10a). The autocorrelation of trading behaviour is weaker, however. A new buy order is followed by an additional 0.56 of a net buy over the subsequent ten periods, in contrast to the effect in the broad sample, which was estimated to be 0.72 (Graph 10d).

These patterns shifted in the course of the day, in ways analogous to the shifts across the different subsamples studied in model (2). During the most turbulent period, 11 am to 2 pm, when duration was at its shortest, trades had a relatively stronger effect on returns and were relatively more autocorrelated than was the case either before 7 am or after 2 pm. In the 7 to 11 am and 11 am to 2 pm periods, returns had strong positive effects on the direction of trades, while after 2 pm this relationship became negative. The persistence of shocks to returns was much higher between 11 am and 2 pm, while before and after this time it was about the same as that estimated for the full sample.

D. Trading in volatile conditions: a summary

Combining the evidence from the duration-based subsamples and from 3 February 2000, it appears that the interactions between price movements and trade behaviour change in at least two ways at times when trading is volatile and uncertainty is high. First, the impact of trades on price movements (the conventional adverse selection effect) is stronger. Second, however, effects in the other direction - from price movements to trades - become stronger as well. It is also clear that markets can sometimes shift suddenly from one regime to another in terms of the absolute and relative strengths of these different effects. In the case of 3 February 2000, for example, it appears that positive feedback effects diminished substantially as price movements stabilised in the afternoon, and information-driven price dynamics were replaced with a greater role for inventory adjustments.

III. Discussion

The results presented in Sections I and II suggest that the traditional approach to market microstructure, which is focused on the ways in which information is incorporated into market prices through order flow, needs to be augmented by a deeper understanding of the strategic interactions among market participants.

When market participants pursue their individual goals in the face of uncertainty in the market, there are several ways in which they may affect each other's interests. As well as the direct interaction between the two counterparties to a transaction, there are other indirect interactions that occur through the impact of trades on price and other characteristics of the market. These interactions affect the incentives of market participants, and may also have a direct bearing on the performance of their portfolios, and hence their conduct in the market.

Take the example of a market in which two traders face a market-maker who attempts to smooth his inventory position across trades. When the market-maker receives a sell order from one of the traders, he may subsequently set a price that is relatively low in order to attract a buy order from the other trader. The trader who then purchases at this low price has benefited from the sell order from the first trader, even though the interaction is indirect, through the market-maker. This example is one where the actions of the two traders are *offsetting* in the sense that a sale by the first leads to a purchase by the second. The larger the sale, the greater the incentive to buy, and vice versa. When viewed over the two trading periods, the actions of the two traders can be seen as *strategic substitutes*, in which the greater incidence of one action leads to a greater incentive (via prices) to adopt the reverse action. In terms of price dynamics, the payoff interactions between the two traders have a stabilising effect in which any deviation of price from its fair value elicits a trade that dampens this deviation.

We may contrast this with modes of interaction where traders' actions are mutually reinforcing, and short-term fluctuations are amplified. For instance, let us modify the above example so that both traders are portfolio managers whose respective mandates dictate that they engage in portfolio insurance by using trading techniques that replicate a synthetic call option through delta-hedging. This entails selling the asset when its price falls and buying it when its price rises. In this scenario, when the price of the asset falls because of an exogenous shock, both traders will attempt to sell it to the market-maker. But if the market-maker then marks down the price because of inventory reasons, the rigid trading rule of both traders dictates a further round of selling, which may feed into even lower prices. This is an instance where the strategic interaction between the traders is *mutually reinforcing*, rather than offsetting. The greater the sale by one trader, the greater the sale by the other trader. In other words, the actions of the traders are *strategic complements*.

The example of strict portfolio insurance is admittedly extreme, although accounts of the 1987 stock market crash attribute some blame to such practices (see Gennotte and Leland (1990)). More generally, however, mutually reinforcing interactions are characteristic of markets where traders have short decision horizons, or where they operate under external constraints on their decisions. The short horizon may be due to internally imposed trading limits that arise as a response to agency problems within an organisation, or when traders operate under a risk management system which circumscribes their actions. In those markets where traders are highly leveraged, the short horizon can be attributed to bankruptcy constraints, which may require positions to be sold for cash when net asset values are low or when a margin call dictates liquidation of trading positions.

The distinction between stabilising and amplifying interactions between traders suggests an important dimension along which we can classify the interaction between market participants. Mutually reinforcing actions are a distinctive characteristic of markets under stress. We have had several occasions to witness their disruptive effects in the recent episodes of market distress following the Asian crisis of 1997 and the Russian/LTCM crisis of 1998. Financial commentators, central bankers and other regulators have consequently devoted a great deal of attention to understanding the nature of positive feedback trading and its implications for supervision and policy execution.

In contrast to the concerns expressed by central bankers and other regulators about the effects of feedback trading, the literature on market microstructure has placed relatively little weight on the possible payoff interaction between traders through mutually reinforcing actions.[9] In part, this is explained by the prevailing theoretical approach to microstructure issues, which emphasises the adverse selection problem confronted by a market-maker who faces possibly better informed traders. The task of the market-maker is to anticipate her losses to better informed insiders. This is typically done by quoting prices that incorporate an actuarially fair safety margin so that losses to insiders are

[9] Among the few exceptions is the literature on momentum trading in the stock market. See DeLong et al (1990), Grinblatt et al (1995) and Jegadeesh and Titman (1993).

compensated by gains from uninformed traders. The direction of causality runs from order flows to price changes.

In such an environment, the intensity of trading is related to the arrival rate of new information, although the theory admits a wide variety of empirical manifestations of this process. Easley and O'Hara (1992) propose a framework in which trading activity is positively related to the arrival rate of new information. When information flow is slow, trading activity itself is slow, while when information flow is fast, this is reflected in high trading activity. In this view, a burst of market activity is due to the exogenous arrival of new information. Easley and O'Hara coined the term "event uncertainty" to describe the fluctuations in the arrival rate of new information. The term refers to the uncertainty concerning this exogenous process. In contrast, Lyons (1996) proposes an alternative "hot potato" hypothesis for the foreign exchange market in which dealer inventory adjustment takes centre stage, and hence higher levels of trading activity are associated with lower arrival rates of new information. In both cases, however, the direction of causality runs from order flows to price changes.

In Sections I and II above it was shown that, while the order flow effect on prices is undoubtedly present and important in the US government securities market, under certain circumstances the causality runs in both directions, so that price changes influence order flow. The effect seems particularly strong in situations where trading is rapid and volatile.

These features are reminiscent of economic models where agents' actions are mutually reinforcing, such as during currency attacks or bank runs. Such contexts are usually fertile territory for multiple equilibria, where there is more than one set of self-fulfilling beliefs. For instance, in the currency attack context, when the agents believe that a currency peg will fail, their actions in anticipation of this precipitate the crisis itself, while if they believe that a currency is not in danger of imminent attack, their inaction spares the currency from attack, thereby vindicating their initial beliefs. The global game method advocated by Morris and Shin (2000) may be one way to introduce elements of concerted shifts in trading positions as a function of the underlying fundamental. Consider the following sketch of a model of short-term traders who operate in a market with limited liquidity. Traders face the choice of taking a long position in an asset, or taking a short position (both up to some fixed bound). They are assumed to have short horizons, so that their payoffs are determined by the price of the asset at the next date. The traders operate in a market with limited liquidity, in the following sense. When the net demand for the asset among the traders is non-zero, the market clears by means of a residual demand/supply function which is imperfectly elastic. The greater the net demand from the set of traders, the higher the market clearing price. Conversely, the greater the net supply, the lower the market clearing price.

This framework gives rise to strategic complementarities in which the actions of the traders are mutually reinforcing. If a large proportion of the traders decide to switch from being short to taking a long position, the market clearing price is raised accordingly, and hence the incentive for any individual trader to take a long position is increased. Conversely, the larger the proportion of the traders who switch to a short position, the lower the market clearing price, and hence the greater the incentive for an individual trader to take a short position. Notice the importance of the short horizon assumption here, and the absence of players with deep pockets that stand ready to provide an infinitely elastic demand/supply function. The uncertainty in the return from date t-1 to date t thus has two components. As well as any exogenous uncertainty in the fundamental value of the asset, there is the endogenous price response arising from the trading decisions of the traders themselves and the imperfectly elastic residual demand/supply function. When each trader has a noisy signal concerning the exogenous uncertainty, the traders follow a switching strategy around a threshold point for the signal realisation, in which a trader goes long if his signal lies above this threshold, but goes short if it lies below it.

One consequence of this equilibrium is that the short-run demand curve for the asset is upward-sloping. The traders buy the asset when the fundamentals are good, which is precisely when the fundamental value of the asset is high. But the traders' actions exacerbate the price response, sending the price higher. This price response validates the action to buy. In terms of the observables, this equilibrium entails that the traders tend to buy the asset (or keep to a long position) precisely when the price of the asset is high. Conversely, if the fundamentals are bad, the traders as a group tend to sell the asset, which brings about a low price for the asset. The demand curve for the group as a whole is therefore upward-sloping.

Since the degree of strategic interaction depends on the initial holdings of the traders, so will the return density. The price response seen for 3 February 2000 may be better understood by reference to the fact that many active traders had short positions on US Treasury securities before the Federal Reserve's announcement.

The price pattern for the trading on 3 February 2000 is suggestive of the following scenario. An initial frenzy of buying is triggered when traders who are caught short in a rising market close out their positions, and/or the anticipated buying by the rumoured distressed institution brings in speculative buying. The exaggerated price response pushes the price up to a sharp peak at around noon, by which time we may conjecture that some of the net short positions of the traders had been unwound, and some may have taken on long positions. When the New York Fed issues its denial at 12.14 pm, the response of the market is sharply downwards, reversing much of the price increase seen in the morning. The market recovers some of its composure by 2 pm, from which time the market trades in relatively tranquil mode until the close.

We believe that this line of investigation may yield theoretical models that do a better job of capturing strategic notions such as overhangs of leveraged positions, short covering and the like.

IV. Conclusions

We have found that the interactions between trades and quote changes in the US Treasury securities market tend to change in important ways when trading conditions are rapid and volatile. We examine trading in the two-year, five-year, and 10-year on-the-run Treasury notes over the period January 1999 to December 2000. The impact of trades on prices tends to become stronger, confirming a common theoretical result in the market microstructure literature. The impact of prices on trades tends to change as well on more volatile days, generally in a positive direction. As a consequence of these two effects, price changes tend to be more positively (or less negatively) autocorrelated on days when conditions are more volatile. This pattern comes through when one compares unusually turbulent days with normal days or unusually quiet days. It also emerges from a close analysis of quotes and trades from 3 February 2000, which was a particularly volatile trading day during this period.

The models commonly used in the analysis of market microstructure emphasise adverse selection effects resulting from the presence of informed and uninformed traders in the market. This helps to explain the impact of trades on prices, but a richer theoretical approach is necessary to capture the impact of prices on trades. Such effects might come out of a model where traders face uncertainty, not just about the fundamental value of an asset, but also about the precision of the signals observed by them and by other traders. In such an environment, a price movement in a given direction could lead a trader to revalue the asset in the same direction, at least for a short period of time. This would lead to positive feedback in trading behaviour and, as a result, in returns over short horizons.

Tables

Table 1a

Statistics on returns, trades and trading volumes (1999-2000)

	2-year	**5-year**	**10-year**
Number of observations	358,361	494,437	506,880
of which:			
% trades only	39.7	22.5	18.9
% quote changes only	49.5	64.7	70.9
% trades and quote changes	10.8	12.8	10.2
Trades			
Number of trades	180,967	174,406	147,546
% buys	52.9	51.1	50.6
Volume per trade ($ millions)			
Mean	12.96	7.28	5.45
Standard deviation	22.65	9.03	7.41
Trading days	501	501	501
Transactions per day	361.21	348.12	294.50
Volume per day ($ millions)	4,622	2,534	1,604
Tick-by-tick returns[1]			
Mean	5.28×10^{-9}	5.64×10^{-10}	-7.02×10^{-9}
Mean absolute value	2.76×10^{-5}	5.38×10^{-5}	0.000101
Standard deviation	4.46×10^{-5}	8.31×10^{-5}	0.000156
Daily returns			
Mean	3.68×10^{-6}	7.07×10^{-7}	-7.20×10^{-6}
Mean absolute value	0.000667	0.001750	0.003065
Standard deviation	0.000882	0.002325	0.004017
Time between ticks (minutes)			
Full sample	0.98	0.76	0.74
High-duration days (top 50)	1.96	1.93	1.81
Low-duration days (bottom 50)	0.67	0.48	0.51
Low trading range days (bottom 50)	1.53	1.00	0.93
High trading range days (top 50)	0.73	0.59	0.61
Low-volatility days (bottom 50)	1.18	1.15	1.06
High-volatility days (top 50)	0.78	0.62	0.62

[1] Log change in midpoint between bid and ask quotes.

Table 1b

Correlations among daily price range, price change, volatility and average duration: two-year note

	Price range	Volatility	Price change[1]
Duration[2]	–0.502	–0.359	–0.031
Price range[3]		0.552	0.093
Volatility[4]			0.129

[1] Difference between daily close and open prices. [2] Daily average time between observations, in minutes, detrended and adjusted for time-of-day and time-of-year effects. [3] Difference between daily high and low prices. [4] Daily standard deviation of tick-by-tick returns.

Table 2

Vector autoregression results: signed trades

This table gives the estimated coefficients from the following vector autoregression:

$$r_t = \sum_{i=1}^{10} \alpha_i r_{t-i} + \sum_{i=0}^{10} \beta_i x_{t-i} + \varepsilon_{1,t}$$

$$x_t = \sum_{i=1}^{10} \gamma_i r_{t-i} + \sum_{i=1}^{10} \delta_i x_{t-i} + \varepsilon_{2,t}$$

r_t is defined as the change from t–1 to t in the log of the midpoint between the prevailing bid and ask quotes. The variable x_t takes the value 1 for a buyer-initiated trade, –1 for a seller-initiated trade, and 0 for a quote revision without a trade. The VAR is estimated over the period from 4 January 1999 to 29 December 2000, and includes only the transactions and quote changes taking place between 7 am and 5 pm. On each day, the estimation starts with the 11th observation after 7 am.

2-year, full sample

	Dept variable: r_t		Dept variable: x_t	
	Coef	**t-stat**	**Coef**	**t-stat**
Lags of r_t				
1	–0.256	–151.86	–130.075	–4.80
2	–0.146	–83.96	267.373	9.57
3	–0.063	–35.66	219.595	7.78
4	–0.022	–12.74	122.318	4.33
5	–0.005	–2.99	74.322	2.63
6	0.002	0.87	34.122	1.21
7	0.006	3.56	13.347	0.47
8	0.010	5.79	37.079	1.32
9	0.003	1.89	12.744	0.46
10	0.001	0.90	50.216	1.88
Lags of x_t [1]				
0	0.665	63.59		
1	0.989	90.95	0.260	153.80
2	0.531	47.98	0.114	64.41
3	0.155	13.96	0.024	13.47
4	0.061	5.49	0.005	2.59
5	–0.014	–1.29	–0.003	–1.50
6	–0.049	–4.45	0.001	0.48
7	–0.041	–3.71	0.003	1.41
8	–0.044	–3.98	0.005	2.60
9	–0.002	–0.19	0.003	1.74
10	–0.010	–0.90	0.003	1.46
$\bar{R}^2$	0.11		0.10	

[1] Coefficient estimates for the r_t equation are multiplied by 100,000.

Table 2 (cont)

5-year, full sample

	Dept variable: r_t		**Dept variable: x_t**	
	Coef	**t-stat**	**Coef**	**t-stat**
Lags of r_t				
1	–0.257	–179.89	155.093	14.67
2	–0.091	–61.85	41.131	3.77
3	–0.035	–23.82	71.799	6.55
4	0.002	1.04	49.575	4.52
5	0.005	3.16	22.074	2.02
6	0.015	10.25	22.190	2.03
7	0.008	5.60	–5.441	–0.50
8	0.014	9.17	–11.409	–1.04
9	0.013	8.79	–5.459	–0.50
10	0.008	5.43	6.509	0.62
Lags of x_t[1]				
0	2.289	118.35		
1	1.728	86.97	0.164	112.78
2	0.998	49.59	0.105	71.22
3	0.328	16.23	0.048	31.90
4	0.065	3.22	0.021	14.12
5	–0.015	–0.76	0.009	6.18
6	–0.065	–3.20	0.002	1.29
7	–0.048	–2.35	0.003	2.15
8	–0.063	–3.14	0.004	2.41
9	0.011	0.57	0.003	1.90
10	–0.018	–0.92	0.003	2.37
$\bar{R}^2$	0.10		0.06	

[1] Coefficient estimates for the r_t equation are multiplied by 100,000.

Table 2 (cont)

10-year, full sample

	Dept variable: r_t		**Dept variable: x_t**	
	Coef	**t-stat**	**Coef**	**t-stat**
Lags of r_t				
1	–0.268	–190.03	38.188	7.47
2	–0.117	–80.15	–38.226	–7.22
3	–0.063	–43.17	–17.908	–3.36
4	–0.019	–12.81	–17.048	–3.19
5	–0.004	–2.95	–19.238	–3.60
6	0.006	4.12	–12.031	–2.26
7	0.003	2.16	–13.565	–2.54
8	0.006	4.10	–10.258	–1.93
9	0.004	3.02	–5.363	–1.02
10	0.007	4.69	–2.859	–0.57
Lags of x_t[1]				
0	3.964	101.70		
1	3.490	87.91	0.129	90.40
2	2.135	53.23	0.079	54.30
3	1.037	25.75	0.035	23.75
4	0.426	10.57	0.014	9.72
5	0.078	1.94	0.006	4.11
6	0.009	0.21	0.004	2.85
7	–0.062	–1.54	0.004	2.68
8	–0.023	–0.56	0.005	3.46
9	–0.087	–2.16	0.005	3.21
10	–0.038	–0.96	0.004	2.88
$\overline{R}^2$	0.10		0.03	

[1] Coefficient estimates for the r_t equation are multiplied by 100,000.

Table 3

Vector autoregression results: signed order flow

This table gives the estimated coefficients from the following vector autoregression:

$$r_t = \sum_{i=1}^{10} \alpha_i r_{t-i} + \sum_{i=0}^{10} \beta_i v_{t-i} + \varepsilon_{1,t}$$

$$v_t = \sum_{i=1}^{10} \gamma_i r_{t-i} + \sum_{i=1}^{10} \delta_i v_{t-i} + \varepsilon_{2,t}$$

r_t is defined as the change from t–1 to t in the log of the midpoint between the prevailing bid and ask quotes. The variable v_t is the size of the trade in millions of dollars, multiplied by the directional indicator x_t defined above. The VAR is estimated over the period from 4 January 1999 to 29 December 2000, and includes only the transactions and quote changes taking place between 7 am and 5 pm. On each day, the estimation starts with the 11th observation after 7 am.

2-year, full sample

	Dept variable: r_t		**Dept variable: v_t**	
	Coef	**t-stat**	**Coef**	**t-stat**
Lags of r_t				
1	–0.212	–126.04	3,129.254	4.36
2	–0.109	–63.42	5,927.097	8.09
3	–0.034	–19.61	3,312.052	4.49
4	–0.003	–1.58	465.159	0.63
5	0.006	3.42	2,078.347	2.82
6	0.007	4.08	967.235	1.31
7	0.009	4.95	794.467	1.08
8	0.012	6.68	722.322	0.98
9	0.004	2.36	1,098.867	1.51
10	0.002	1.23	1,001.097	1.41
Lags of v_t[1]				
0	0.019	48.08		
1	0.018	44.73	0.052	31.05
2	0.012	30.13	0.074	43.52
3	0.005	11.42	0.042	24.89
4	0.001	1.33	0.074	43.67
5	–0.002	–4.11	0.002	1.02
6	–0.003	–7.64	0.016	9.47
7	–0.002	–3.78	0.009	5.17
8	–0.002	–5.54	0.015	8.87
9	–0.001	–3.62	0.007	3.84
10	–0.001	–1.75	–0.006	–3.77
$\bar{R}^2$	0.06		0.02	

[1] Coefficient estimates for the r_t equation are multiplied by 100,000.

Table 3 (cont)

5-year, full sample

	Dept variable: r_t		**Dept variable: v_t**	
	Coef	**t-stat**	**Coef**	**t-stat**
Lags of r_t				
1	–0.223	–156.24	2,642.990	21.52
2	–0.063	–42.95	2,382.348	18.93
3	–0.017	–11.72	2,043.797	16.20
4	0.012	8.11	1,391.112	11.03
5	0.009	6.47	844.571	6.70
6	0.017	11.66	544.473	4.32
7	0.008	5.72	261.360	2.08
8	0.013	8.82	193.415	1.54
9	0.012	8.17	205.151	1.64
10	0.007	4.83	83.945	0.69
Lags of v_t[1]				
0	0.125	75.12		
1	0.091	54.11	0.080	55.47
2	0.056	33.34	0.053	36.71
3	0.023	13.57	0.032	22.37
4	0.006	3.34	0.017	12.02
5	0.002	1.31	0.008	5.46
6	–0.003	–1.54	0.004	2.73
7	–0.002	–1.26	0.007	5.01
8	–0.005	–2.94	0.003	1.76
9	0.000	0.12	0.005	3.44
10	0.001	0.50	0.001	0.63
$\overline{R}^2$	0.06		0.02	

[1] Coefficient estimates for the r_t equation are multiplied by 100,000.

Table 3 (cont)

10-year, full sample

	Dept variable: r_t		**Dept variable: v_t**	
	Coef	**t-stat**	**Coef**	**t-stat**
Lags of r_t				
1	–0.237	–167.74	515.908	11.04
2	–0.091	–62.69	283.281	5.90
3	–0.047	–32.04	316.900	6.58
4	–0.009	–6.48	219.174	4.54
5	–0.001	–0.37	163.829	3.40
6	0.007	4.58	107.184	2.22
7	0.002	1.48	56.687	1.18
8	0.004	2.92	66.828	1.39
9	0.003	1.79	57.360	1.20
10	0.005	3.55	105.165	2.26
Lags of v_t[1]				
0	0.296	69.32		
1	0.183	42.73	0.053	37.34
2	0.130	30.21	0.044	30.84
3	0.065	15.02	0.029	20.08
4	0.021	4.97	0.015	10.50
5	–0.004	–1.00	0.009	6.36
6	0.001	0.16	0.005	3.19
7	–0.013	–2.92	0.006	4.30
8	0.000	0.07	0.007	5.01
9	–0.007	–1.68	0.004	2.91
10	0.008	1.76	0.007	5.01
$\overline{R}^2$	0.07		0.01	

[1] Coefficient estimates for the r_t equation are multiplied by 100,000.

Table 4

VAR coefficients for different subsamples

The table shows the sums of different combinations of coefficients from the following VAR:

$$r_t = \sum_{i=1}^{10}(\alpha_i + \alpha_i^L d_{t-i}^L + \alpha_i^H d_{t-i}^H) r_{t-i} + \sum_{i=0}^{10}(\beta_i + \beta_i^L d_{t-i}^L + \beta_i^H d_{t-i}^H) x_{t-i} + \varepsilon_{1,t}$$

$$x_t = \sum_{i=1}^{10}(\gamma_i + \gamma_i^L d_{t-i}^L + \gamma_i^H d_{t-i}^H) r_{t-i} + \sum_{i=1}^{10}(\delta_i + \delta_i^L d_{t-i}^L + \delta_i^H d_{t-i}^H) x_{t-i} + \varepsilon_{2,t}$$

where d_{t-i}^L is a dummy variable taking the value 1 during the 50 days when average adjusted duration is lowest during the sample, and d_{t-i}^H equals 1 during the 50 days when average adjusted duration is highest. The 401 days on which both dummies equal zero are referred to as "normal" days. The values in the column "Sum of coefs" are the total of the effects estimated for that subsample. Thus, the first figure in the first column is $\sum_{i=1}^{10}\alpha_i$, the second figure is $\sum_{i=1}^{10}(\alpha_i + \alpha_i^L)$, and so on. The values under the column "Vs normal" are the additional effects for that subsample, relative to the effects estimated for the 401 days that are not in either the high-duration or the low-duration subsample. Thus, the first figure in the second column is $\sum_{i=1}^{10}\alpha_i^L$, the second is $\sum_{i=1}^{10}\alpha_i^H$, and so on. The asterisks indicate the significance level for the F-statistic of a Wald test of the hypothesis that the corresponding sum of coefficients is different from zero. Two asterisks indicate rejection at the 5% level or better, while one asterisk indicates rejection at a level between 5 and 10%.

2-year note

	Return equation		Signed trade equation	
	Sum of coefs	**Vs normal**	**Sum of coefs**	**Vs normal**
Coefficients on returns				
"Normal" days	–0.563 **		767.5 **	
Low duration	–0.210 **	0.353 **	912.8 **	145.3
High duration	–0.599 **	–0.036	–134.5	–902.1 *
Coefficients on signed trades[1]				
"Normal" days	2.277 **		0.421 **	
Low duration	3.026 **	0.749 **	0.348 **	–0.073 **
High duration	2.173 **	–0.104	0.404 **	–0.018

[1] Coefficient estimates for return equation multiplied by 100,000.

Table 4 (cont)

5-year note

	Return equation		Signed trade equation	
	Sum of coefs	**Vs normal**	**Sum of coefs**	**Vs normal**
Coefficients on returns				
"Normal" days	–0.288 **		395.4 **	
Low duration	–0.707 **	–0.419 **	229.1 *	–166.3
High duration	–0.325 **	–0.036	–83.5	–478.8 **
Coefficients on signed trades[1]				
"Normal" days	5.066 **		0.364 **	
Low duration	6.876 **	1.809 **	0.293 **	–0.071 **
High duration	5.297 **	0.231	0.381 **	0.018

10-year note

	Return equation		Signed trade equation	
	Sum of coefs	**Vs normal**	**Sum of coefs**	**Vs normal**
Coefficients on returns				
"Normal" days	–0.424 **		–99.7 **	
Low duration	–0.855 **	–0.430 **	–133.0 *	–33.3
High duration	–0.355 **	0.069	–67.7	32.1
Coefficients on signed trades[1]				
"Normal" days	10.759 **		0.286 **	
Low duration	13.443 **	2.684 **	0.241 **	–0.045 **
High duration	10.477 **	–0.282	0.300 **	0.014

[1] Coefficient estimates for return equation multiplied by 100,000.

Table 5

Trading epochs for the two-year note on 3 February 2000

	Return[1]	% buys	Mean duration	Mean bid-ask spread[2]
7 – 11 am	0.00063	52.6	0.61	0.0097
11 am – 12.15 pm	0.00340	65.9	0.53	0.0102
12.15 – 2 pm	–0.00317	40.9	0.48	0.0181
2 – 5 pm	0.00090	66.7	0.96	0.0120
Memo item: *Full sample (1/99–12/00)*	*0.00067*[3]	*52.9*	*0.98*	*0.0065*

[1] Log change in quote midpoint. [2] Difference between prevailing ask and bid quotes. [3] Mean absolute value of daily log quote-midpoint changes.

Table 6

VAR coefficients for 3 February 2000

This table gives the sums of the estimated coefficients from the following vector autoregression for three time periods on 3 February 2000:

$$r_t = \sum_{i=1}^{5} \alpha_i r_{t-i} + \sum_{i=0}^{5} \beta_i x_{t-i} + \varepsilon_{1,t}$$

$$v_t = \sum_{i=1}^{5} \gamma_i r_{t-i} + \sum_{i=1}^{5} \delta_i v_{t-i} + \varepsilon_{2,t}$$

In each quadrant, the table shows the sum of the coefficients on the corresponding variable (eg $\sum_{i=1}^{5} \alpha_i$). The asterisks indicate the significance level for the F-statistic of a Wald test of the hypothesis that the corresponding sum of coefficients is different from zero. Two asterisks indicate rejection at the 5% level or better, while one asterisk indicates rejection at a level between 5 and 10%.

2-year note

	Return equation	**Signed trade equation**
Coefficients on return		
7 – 11 am	–0.588 **	1393.2
11 am – 2 pm	–0.288 *	1224.4 *
2 – 5 pm	–0.477 *	–836.9
Coefficients on signed trade		
7 – 11 am	5.506[1] **	0.164 *
11 am – 2 pm	4.475[1] **	0.444 **
2 – 5 pm	4.291[1] **	0.376 **

5-year note

	Return equation	**Signed trade equation**
Coefficients on return		
7 – 11 am	–0.331 **	501.5
11 am – 2 pm	0.020	50.2
2 – 5 pm	–0.100	–166.2
Coefficients on signed trade		
7 – 11 am	7.221[1] **	0.321 **
11 am – 2 pm	10.893[1] **	0.383 **
2 – 5 pm	12.850[1] **	0.101

[1] Coefficient estimates multiplied by 100,000.

Table 6 (cont)

10-year note

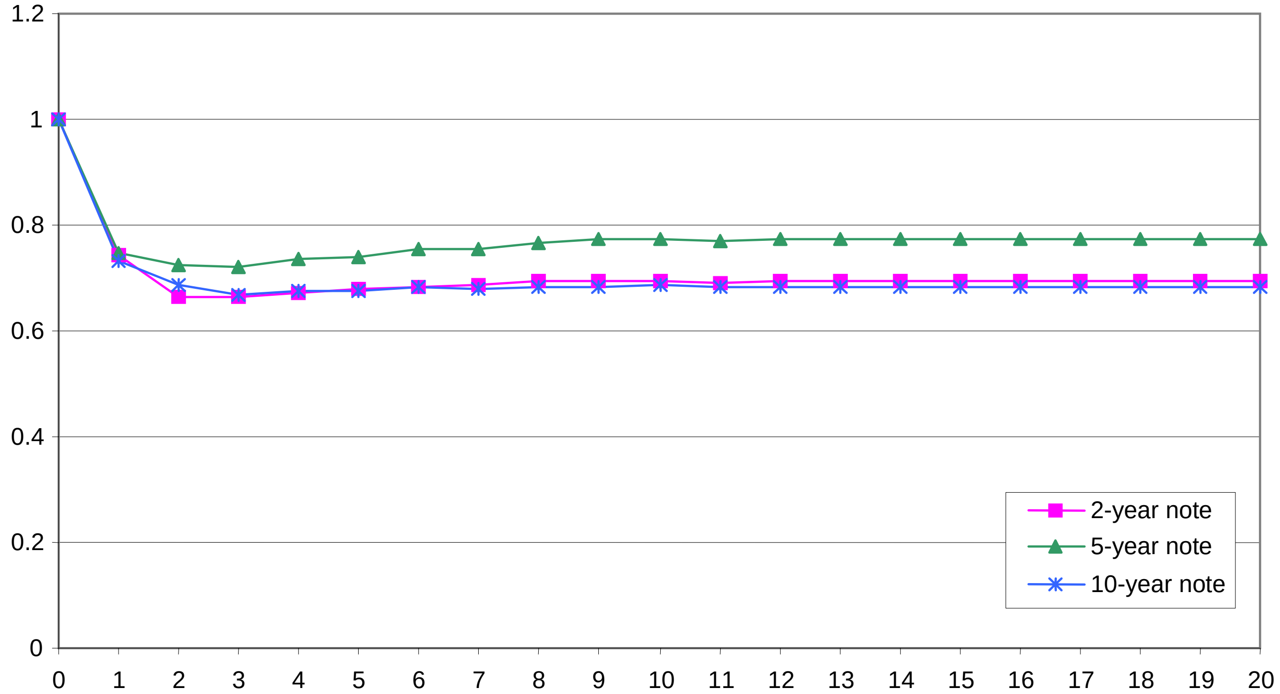
	Return equation	Signed trade equation
Coefficients on return		
7 – 11 am	–0.071	–282.5 **
11 am – 2 pm	0.381 **	50.6
2 – 5 pm	–0.004	–767.9 **
Coefficients on signed trade		
7 – 11 am	26.435[1] **	0.205 **
11 am – 2 pm	10.803[1] **	0.344 **
2 – 5 pm	7.865[1]	0.228 **

[1] Coefficient estimates multiplied by 100,000.

Graphs

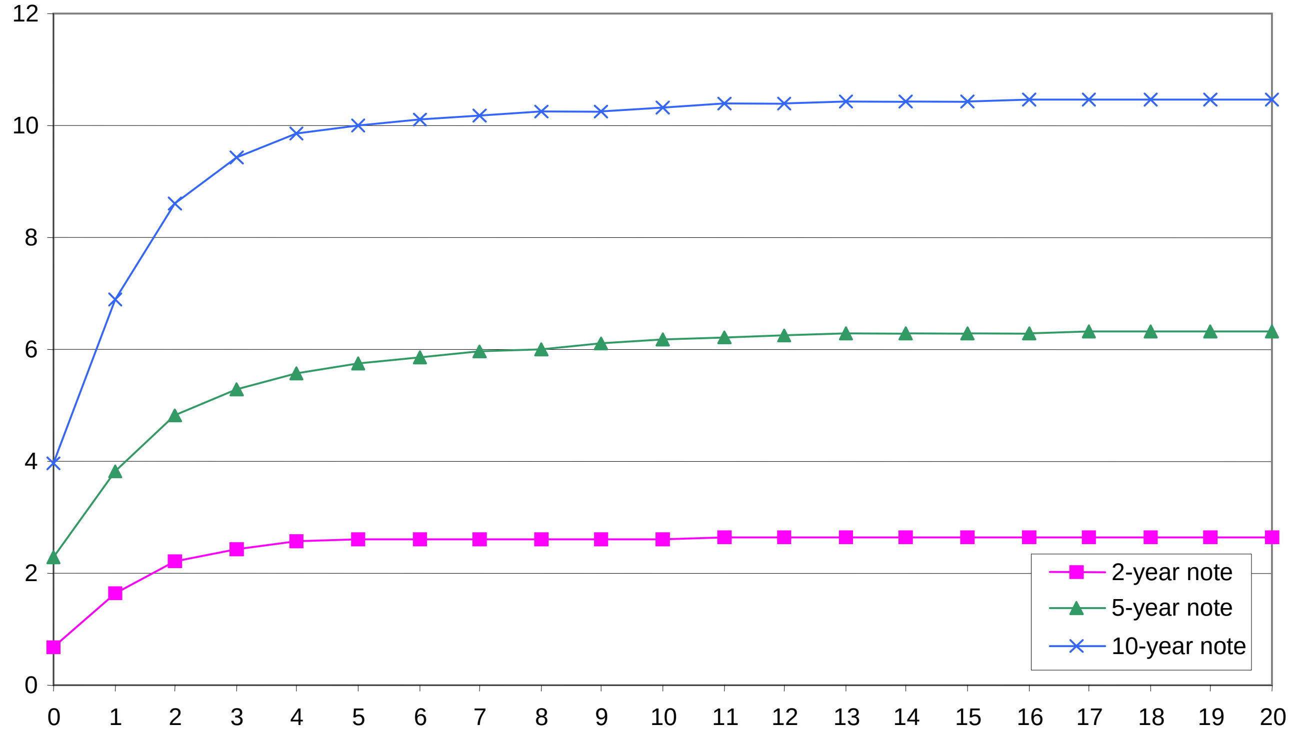
Graph 1

Cumulative effect on return of an additional one unit return



Graph 2

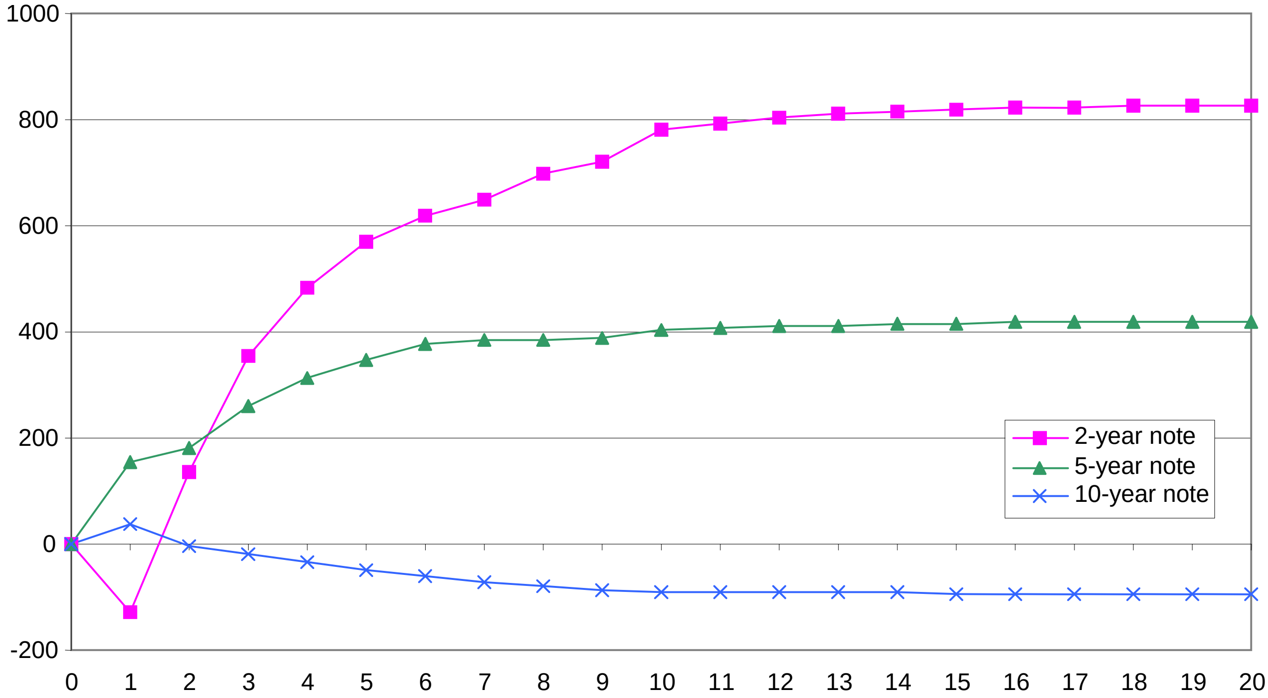
Cumulative effect on return of an additional net buy



Note: Return figures multiplied by 100,000.

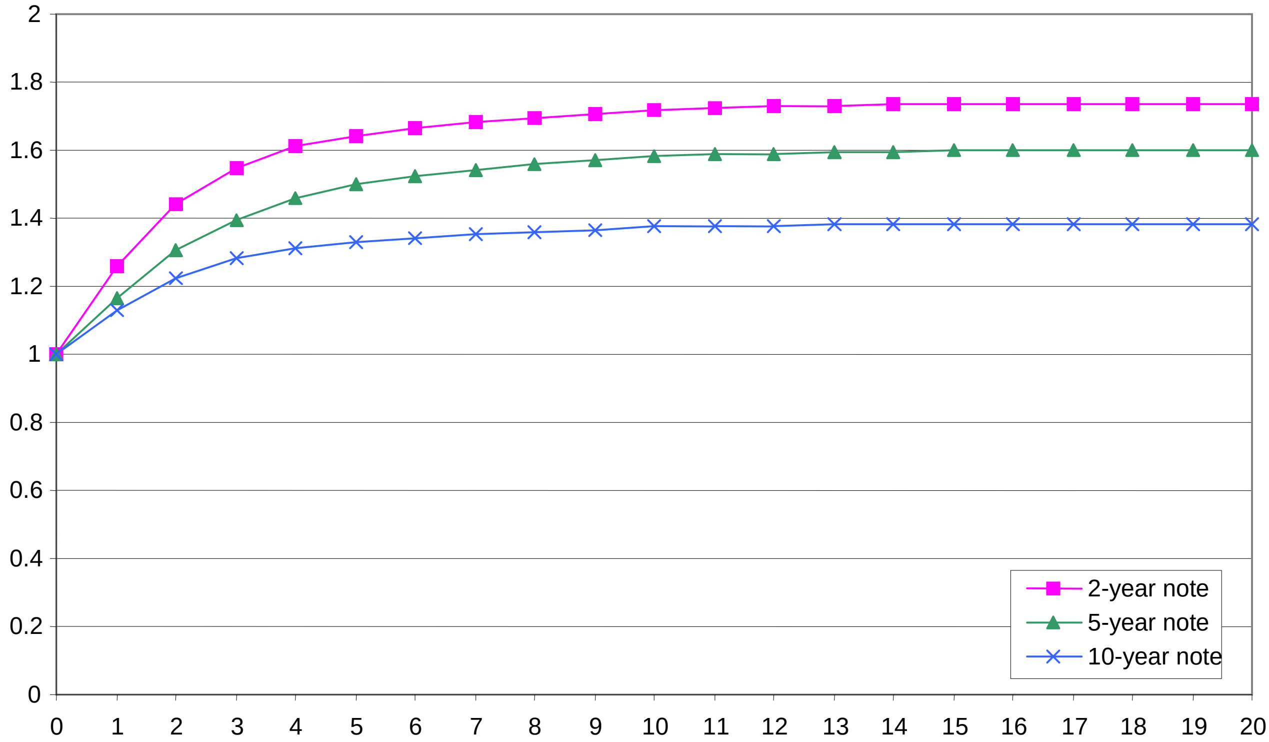
Graph 3

Cumulative effect on net buys of an additional one unit return



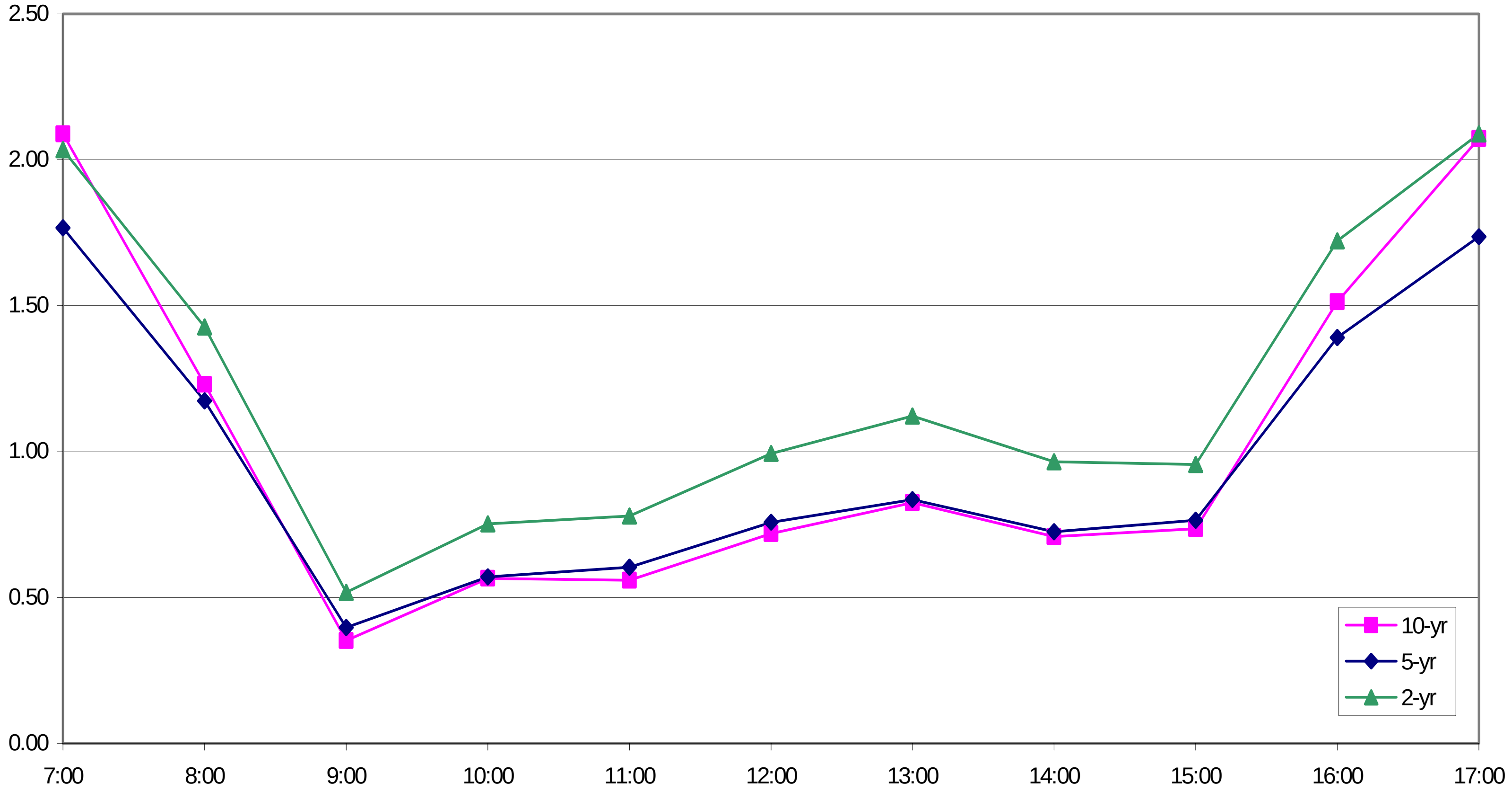
Graph 4

Cumulative effect on net buys of an additional net buy



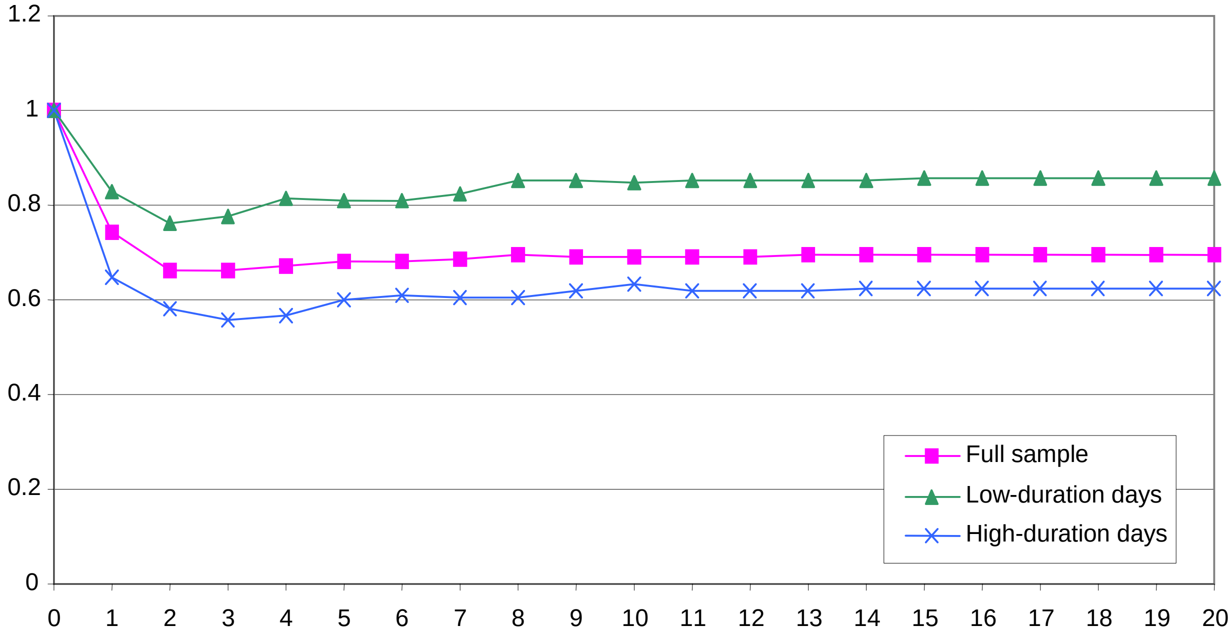
Graph 5

Fitted duration at different times of the day



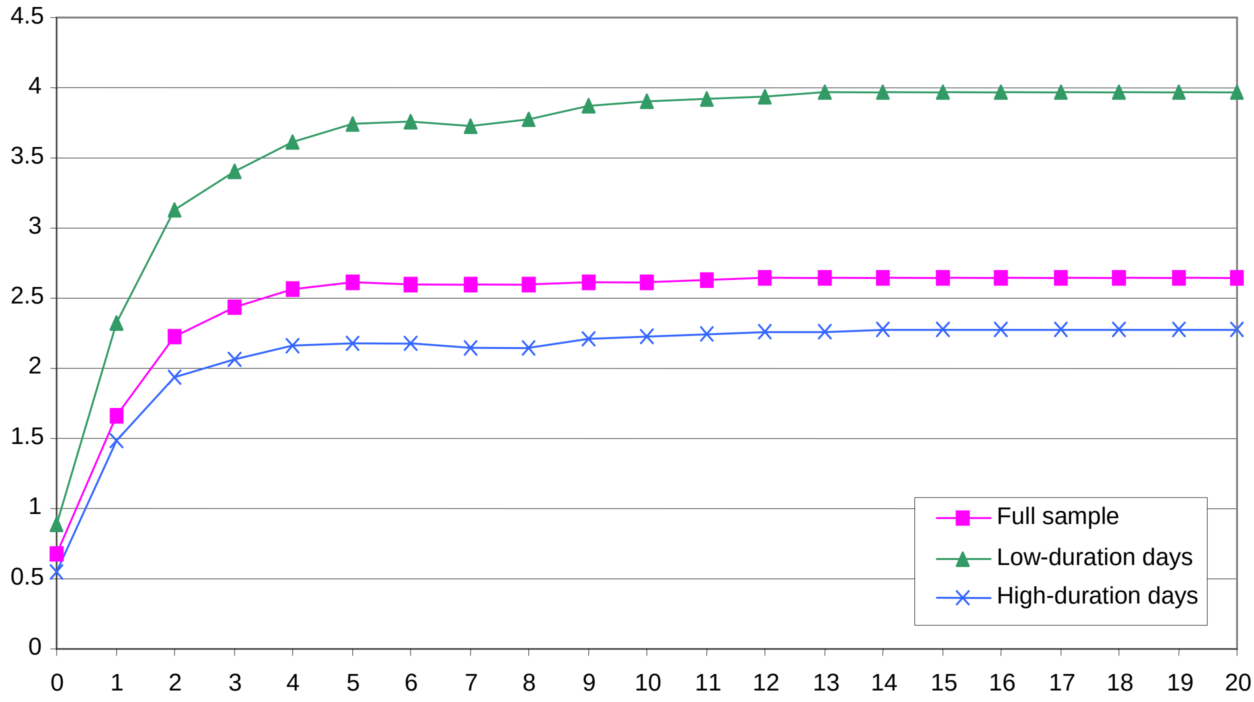
Graph 6a

Cumulative effect on net returns of an additional one unit return: two-year note



Graph 6b

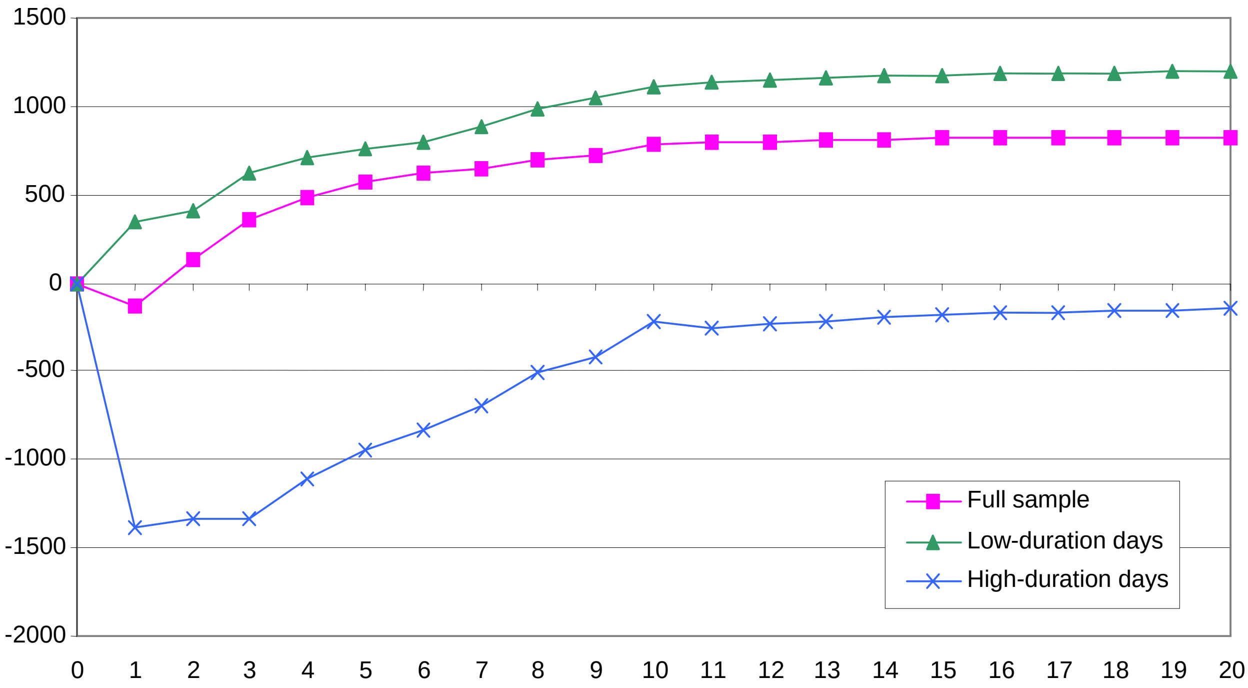
Cumulative effect on return of an additional net buy: two-year note



Note: Return figures multiplied by 100,000.

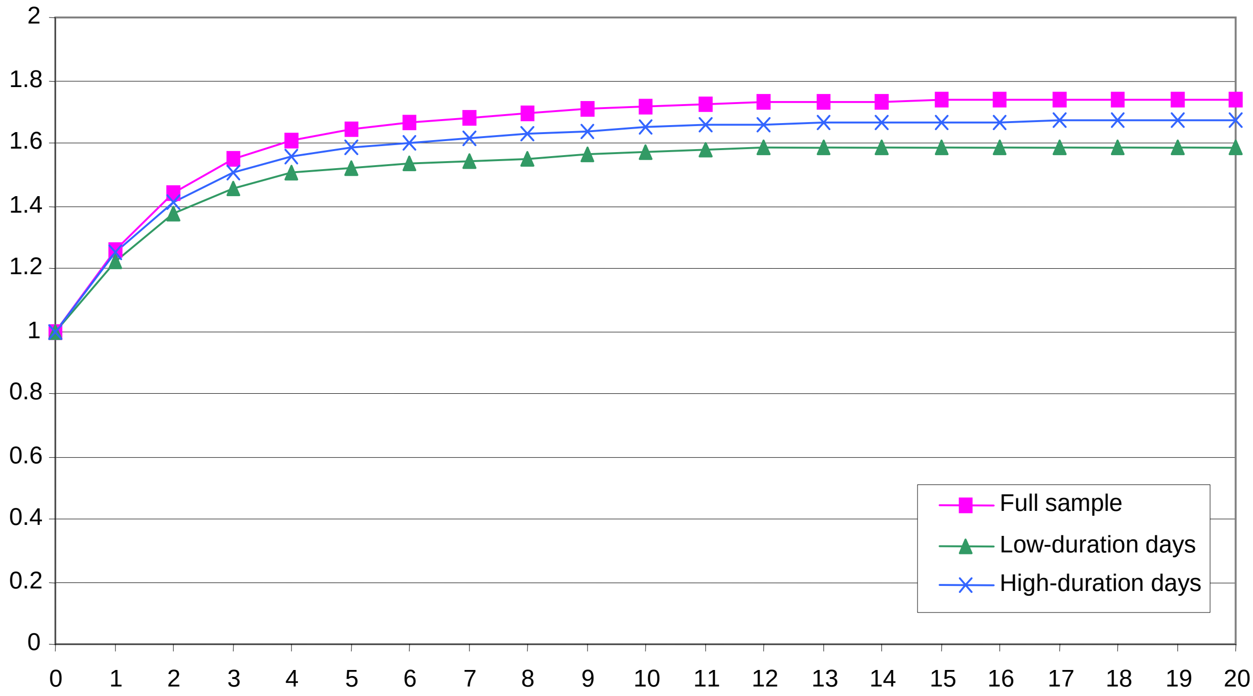
Graph 6c

Cumulative effect on net buys of an additional one unit return: two-year note



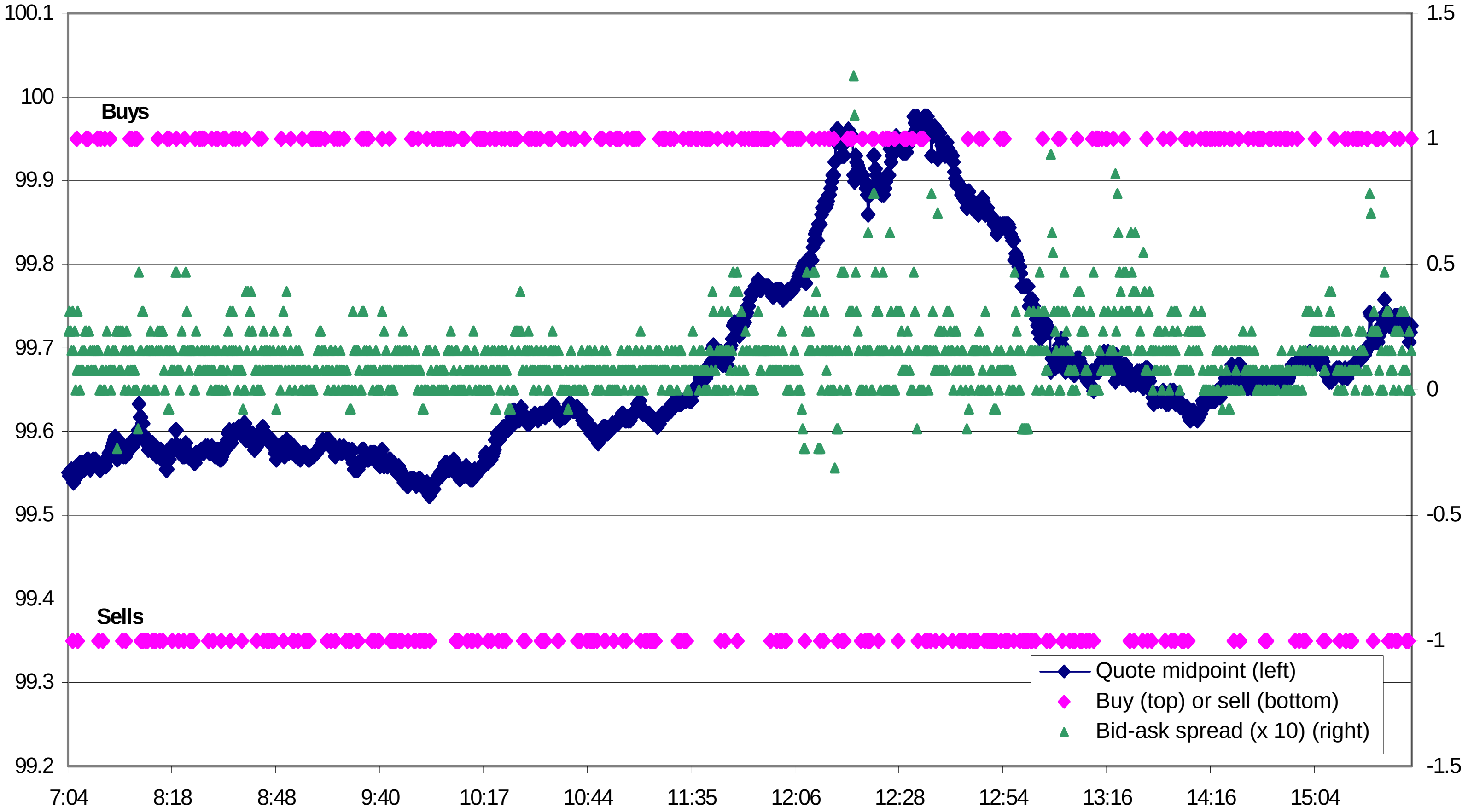
Graph 6d

Cumulative effect on net buys of an additional net buy: two-year note



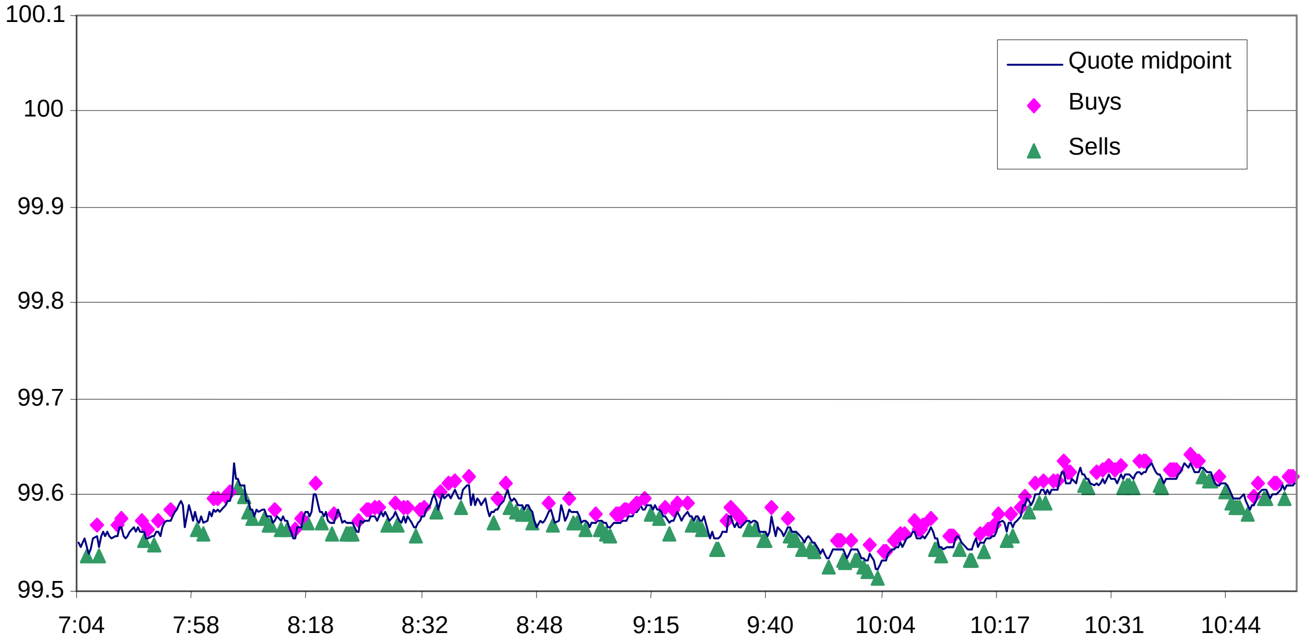
Graph 7

Quotes, trades and bid-ask spreads for the two-year Treasury note: 3 February 2000



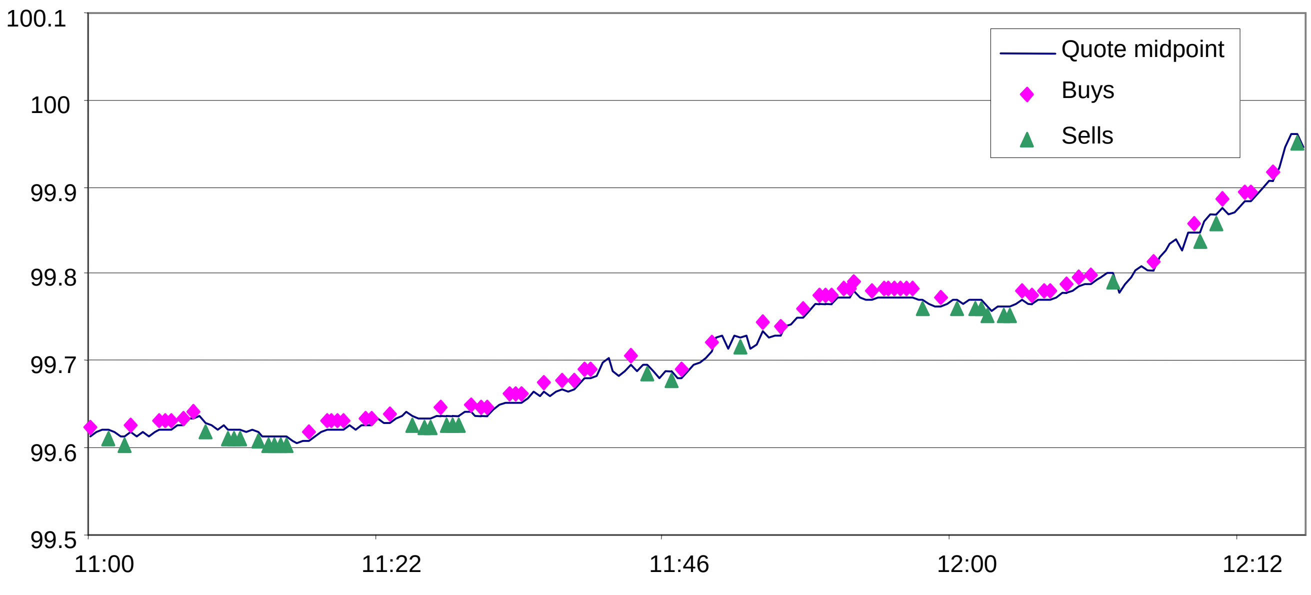
Graph 8a

Quotes and transactions in the two-year note:
3 February 2000, 7 am - 11 am



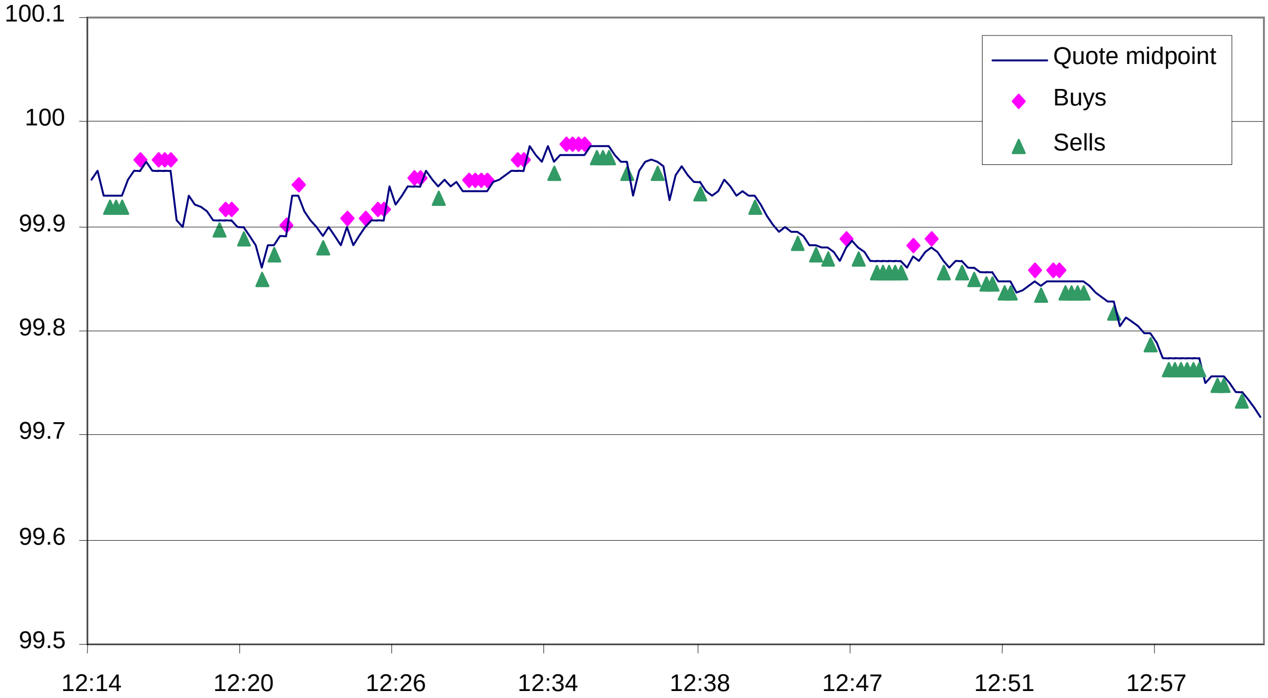
Graph 8b

Quotes and transactions in the two-year note:
3 February 2000, 11 am - 12.15 pm



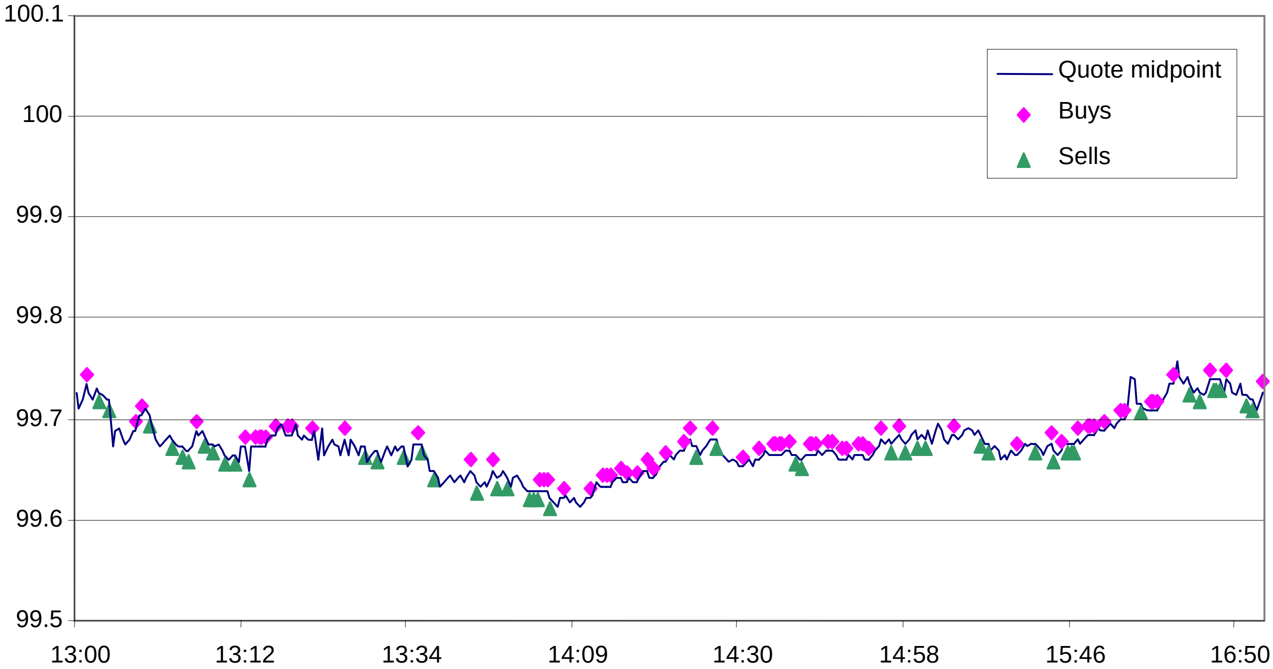
Graph 8c

Quotes and transactions in the two-year note:
3 February 2000, 12.15 pm - 1 pm



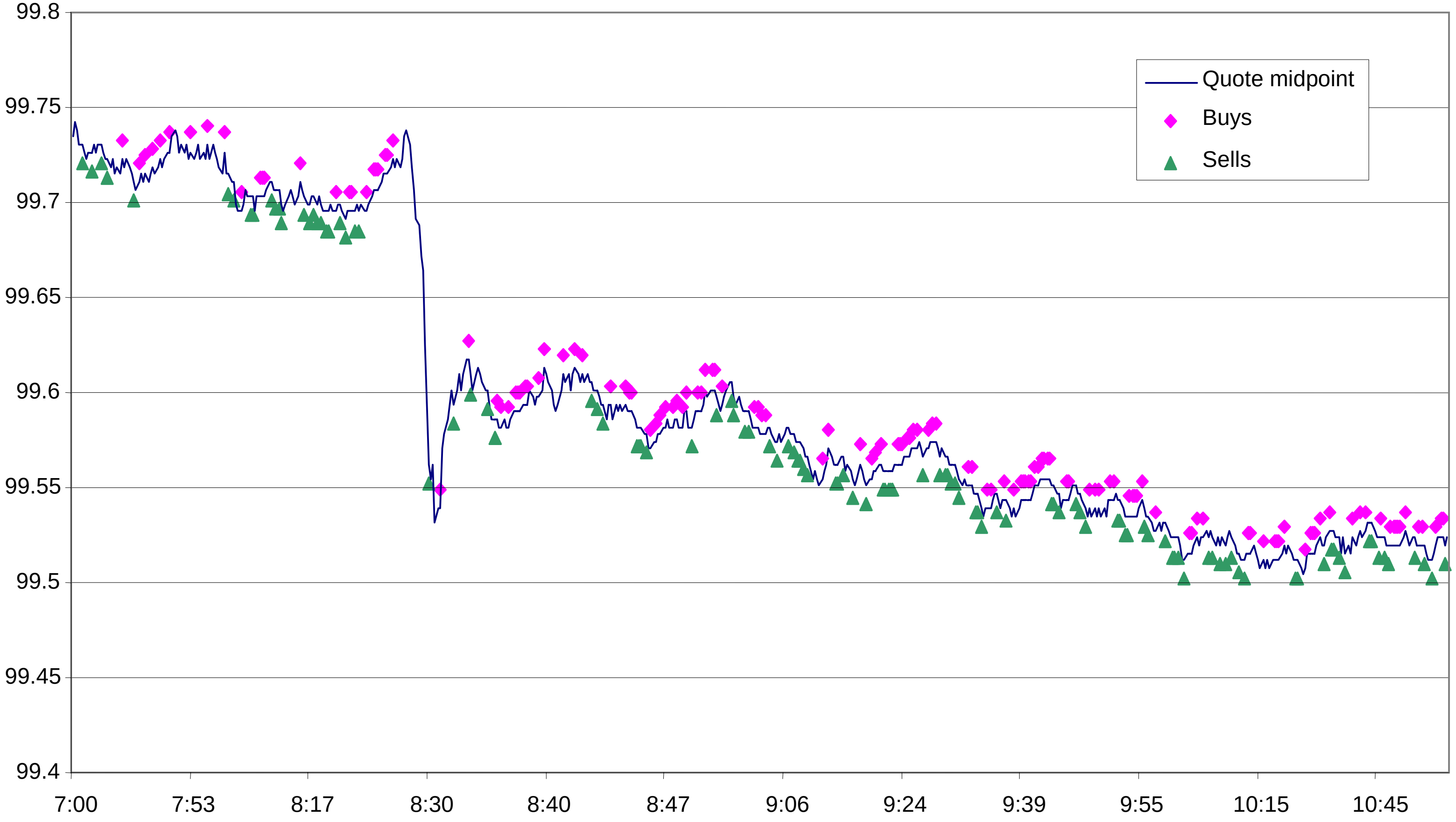
Graph 8d

Quotes and transactions in the 2-year note:
3 February 2000, 1 pm - 5 pm



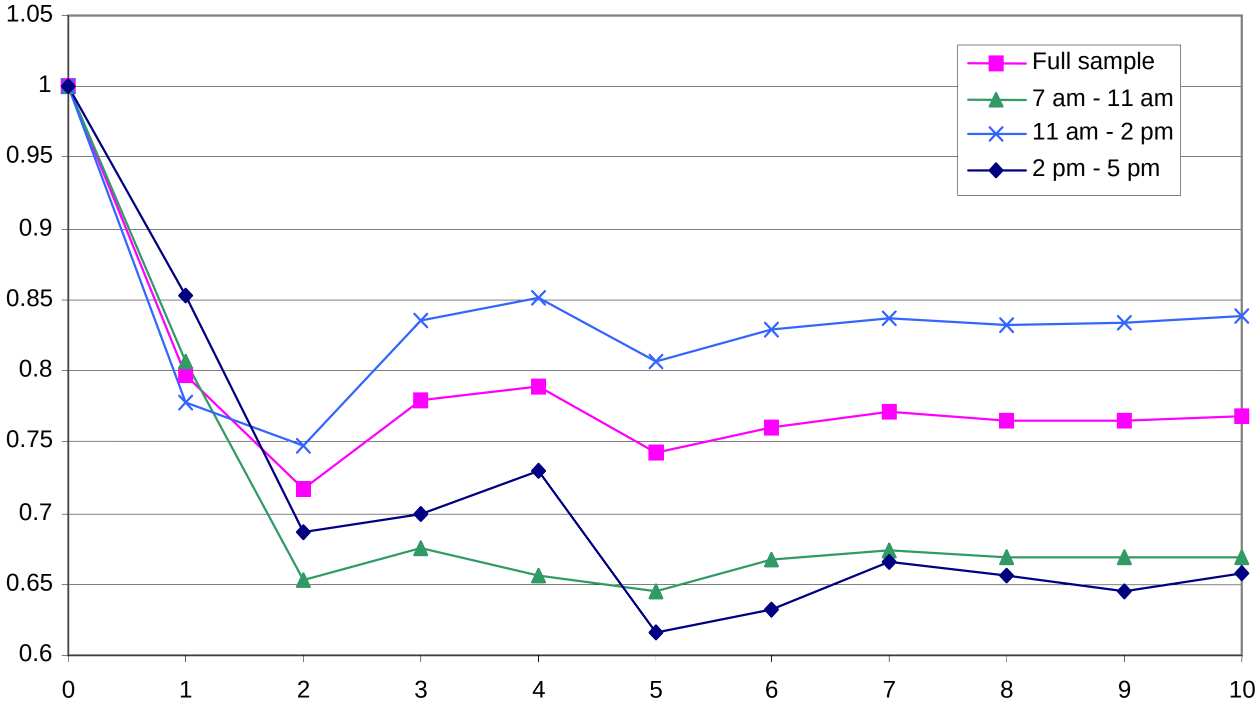
Graph 9

Quotes and transactions in the two-year note:
28 January 2000, 7 am - 11 am



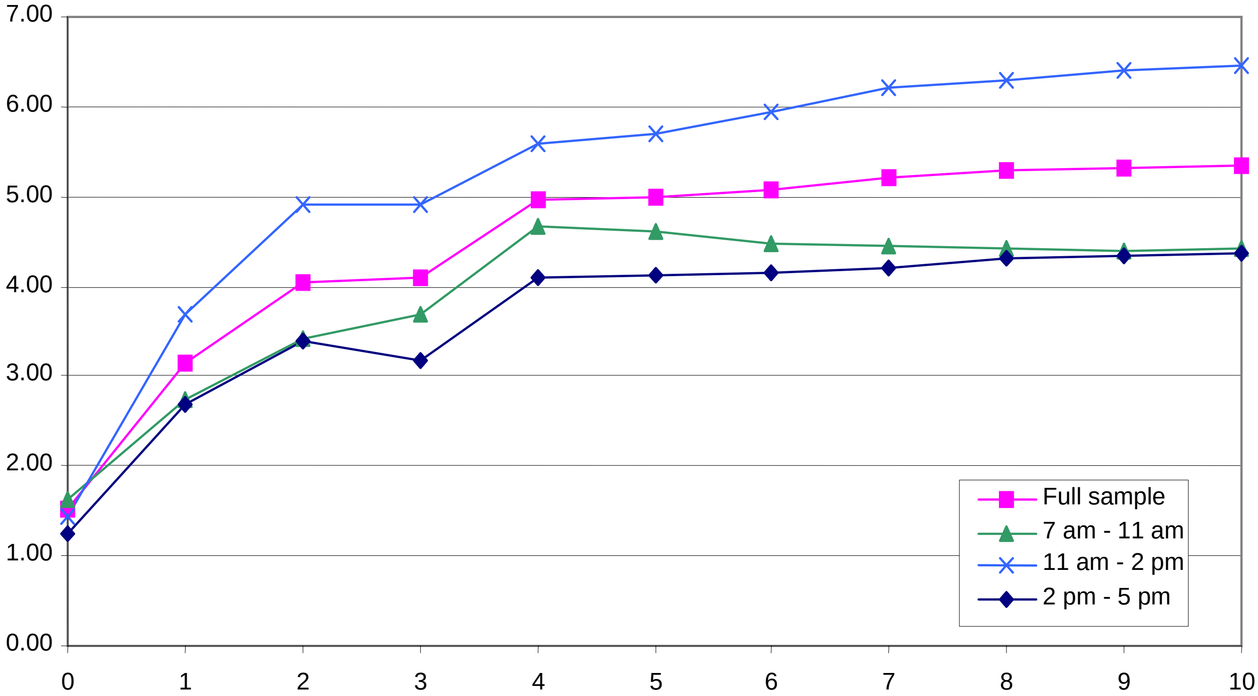
Graph 10a

Cumulative effect on net returns of an additional one unit return: two-year note, 3 February 2000



Graph 10b

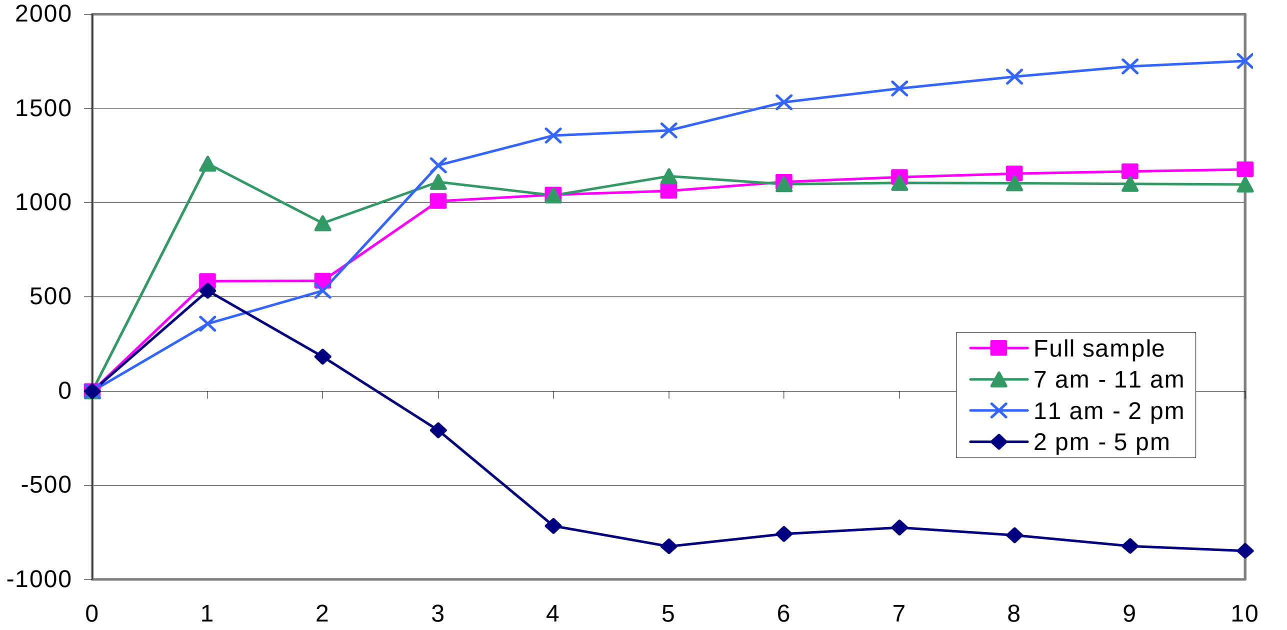
Cumulative effect on return of an additional net buy: two-year note, 3 February 2000



Note: Return figures multiplied by 100,000.

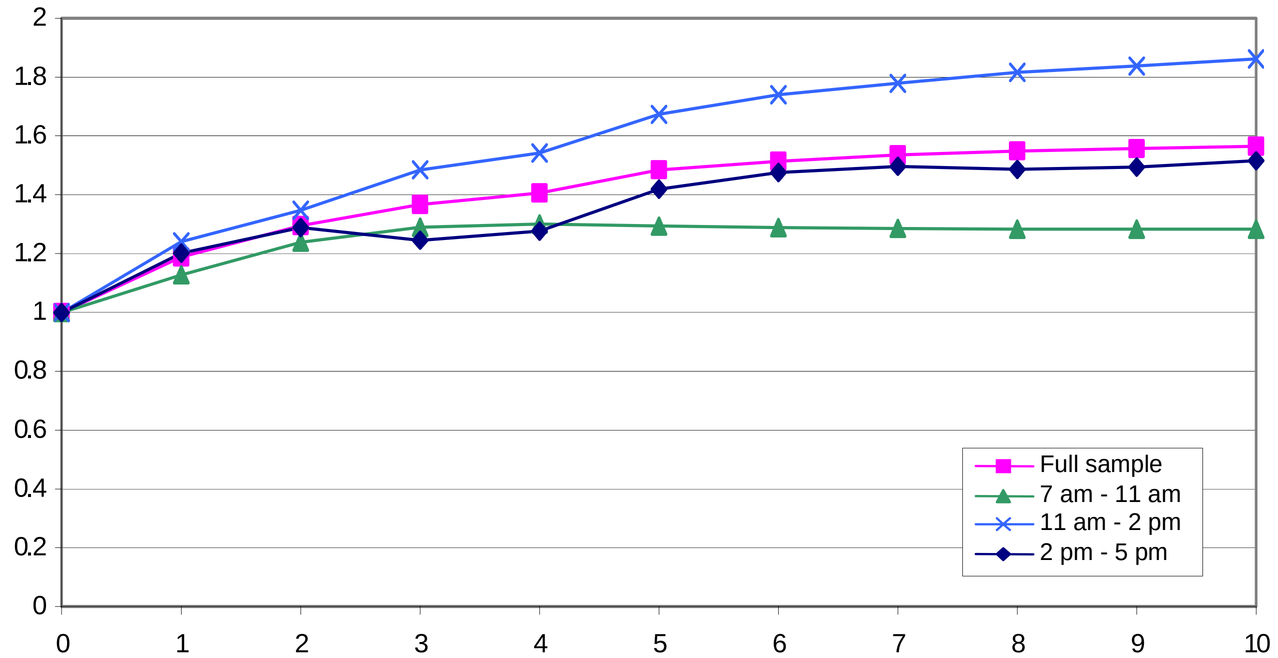
Graph 10c

Cumulative effect on net buys of an additional one unit return: two-year note, 3 February 2000



Graph 10d

Cumulative effect on net buys of an additional net buy: two-year note, 3 February 2000



References

AP Online (2000): "Treasury bond prices mostly up", 3 February.

Boni, Leslie and J Chris Leach (2001): "Depth discovery in a market with expandable limit orders: an investigation of the US Treasury market", working paper, University of New Mexico.

Danielsson, Jon and Richard Payne (2001): "Measuring and explaining liquidity on an electronic limit order book: evidence from Reuters D2000-2", working paper, London School of Economics.

DeLong, J Bradford, Andrei Shleifer, Lawrence Summers and Robert Waldmann (1990): "Positive feedback investment strategies and destabilising rational speculation", *Journal of Finance* 45, 374-97.

Easley, David and Maureen O'Hara (1992): "Time and the process of security price adjustment", *Journal of Finance* 47, 577-605.

Engle, Robert F (2000): "The econometrics of ultra-high-frequency data". *Econometrica* 68, 1-22.

Engle, Robert F and Andrew Patton (2000): "Impacts of trades in an error-correction model of quote prices", working paper, UC San Diego.

Engle, Robert F and Jeffrey R Russell (1998): "Autoregressive conditional duration: A new model for irregularly spaced transaction data", *Econometrica* 66, 1127-62.

Evans, Martin D D and Richard K Lyons (2002): "Order flow and exchange rate dynamics", *Journal of Political Economy* 110, 170-80.

Fleming, Michael (2001): "Measuring treasury market liquidity", Federal Reserve Bank of New York Staff Reports no 133, July.

Fleming, Michael and Eli Remolona (1999a): "Price formation and liquidity in the US Treasury market: The response to public information", *Journal of Finance* 54, 1901-15.

——— (1999b): "The term structure of announcement effects", Bank for International Settlements Working Paper no 71, June.

Gennotte, Gerald and Hayne Leland (1990): "Market liquidity, hedging and crashes", *American Economic Review* 80, 999-1021.

Glosten, Lawrence R and Paul Milgrom (1985): "Bid, ask and transaction prices in a specialist market with heterogeneously informed agents", *Journal of Financial Economics* 14, 71-100.

Grinblatt, Mark, Sheridan Titman and Russ Wermers (1995): "Momentum investment strategies, portfolio performance and herding: A study of mutual fund behaviour", *American Economic Review* 85, 1088-105.

Jegadeesh, N and Sheridan Titman (1993): "Returns to buying winners and selling losers: Implications for stock market efficiency", *Journal of Finance* 48, 65-92.

Hasbrouck, Joel (1991): "Measuring the information content of stock trades", *Journal of Finance* 46, 179-207.

——— (2000): "Trading fast and slow: Security market events in real time", unpublished manuscript, New York University.

Huang, Roger D, Jun Cai and Xiaozu Wang (2001): "Information-based trading in the treasury note interdealer broker market", unpublished manuscript, University of Notre Dame.

Kyle, Albert (1985): "Continuous auctions and insider trading", *Econometrica* 53, 1315-35.

Lyons, Richard K (1996): "Foreign exchange volume: Sound and fury signifying nothing?", Ch 5 in J Frankel, G Galli and A Giovannini (eds), *The microstructure of foreign exchange markets*, University of Chicago Press, 183-201.

Morris, Stephen and Hyun Song Shin (2000): "Market risk with interdependent choice", unpublished working paper, London School of Economics.

Piazzesi, Monika (2001): "An econometric model of the yield curve with macroeconomic jump effects", National Bureau of Economic Research Working Paper no 8246, April.

Roll, Richard, (1984): "A simple implicit measure of the effective bid-ask spread in an efficient market", *Journal of Finance* 39, 1127-40.

Large investors and liquidity: a review of the literature

Matthew Pritsker[1]

Abstract

A growing share of financial assets are held by large institutional investors whose desired trades are large enough to move prices in markets. Because large investors' trades have "price impact", asset markets are not perfectly liquid from their perspective. This illiquidity is likely to influence their decisions of which assets to hold and which assets to trade, and may influence how assets are priced. These insights on illiquidity and large investors motivated Pritsker's (2002) modelling of liquidity in a market with large investors. This article is a companion piece to Pritsker (2002) which reviews the literature on asset liquidity and on large investors and suggests ways in which these research areas can be combined.

1. Introduction

The standard competitive asset pricing paradigm assumes that individual investors' desired trades are sufficiently small that each investor can take prices as given and hence choose their asset holdings while ignoring the price impact of their trades. The price-taking assumption is reasonable when applied to the trades of most individual investors, but it is less tenable when applied to the trades of institutional investors. The observed behaviour of many institutional investors - breaking apart a large trade into several smaller trades, or building up or selling a position over days - suggests that their desired trades have price impact, and that large institutions account for price impact when selecting their trading strategy (Chan and Lakonishok (1995)).

One notion of a perfectly liquid asset is an asset for which individuals can buy and sell all that they want at current prices. This notion of liquidity suggests that many markets are essentially perfectly liquid from the perspective of small investors since prices do not change much, if at all, in response to their desired trades. However, many markets are not perfectly liquid from the perspective of large investors. Because large investors are faced with imperfect market liquidity, the lack of liquidity may influence their investment decisions. For example, large investors who anticipate a potential future need to sell off assets quickly at some unexpected future date to meet cash flow obligations may desire holdings of relatively liquid assets in order to minimise the transaction costs associated with future forced sales. This desire for relatively liquid asset holdings should be reflected in equilibrium asset prices and returns.

The above observations suggest that large investors and asset market liquidity are related topics, and that whether liquidity risk is priced by the market may depend on the trading behaviour of large investors. The purpose of this paper is to review the literature on asset market liquidity and on large investors, and then suggest directions of research which synthesise the two topics. Motivated by the notion that large investors and liquidity are related, Pritsker (2002) studies asset market liquidity in a setting where there are many large and small investors who trade multiple risky assets over a large but finite number of time periods. The analysis in Pritsker builds on other models of large investors. The most closely related research is DeMarzo and Urosevic (2000), Vayanos (2001) and Urosevic (2001). The basic underlying framework in DeMarzo and Urosevic and in Vayanos is nearly identical. Both consider the behaviour of a single large investor and many small investors when the investors

---

[1] Board of Governors of the Federal Reserve System. The views expressed in this paper are those of the author but not necessarily those of the Board of Governors of the Federal Reserve System, or other members of its staff. Address correspondence to Matt Pritsker, The Federal Reserve Board, Mail Stop 91, Washington DC 20551. Matt may be reached by telephone on (202) 452-3534, fax (202) 452-3819, or by e-mail at mpritsker@frb.gov.

trade a single risky and risk-free asset over many time periods, and where all investors have CARA utility of consumption. The models differ in how they depart from this framework. DeMarzo and Urosevic consider a moral hazard setting in which the single large investor also expends costly effort in overseeing the activity of the firm. They model the investors' optimal oversight and portfolio choices together and then examine the implications of these decisions for asset prices. Vayanos modifies the basic setting to instead examine how asymmetric information about the large investor's holdings of risky assets influences the equilibrium behaviour of market prices. The basic framework in Urosevic (2001) extends the basic framework in DeMarzo and Urosevic to allow for multiple large investors and multiple risky assets. Urosevic then proceeds to examine the moral hazard setting when many large investors choose their portfolio holdings and the amount of effort to expend in monitoring the activity of the firms.

The basic modelling framework in Urosevic (2001) is, minus the moral hazard, essentially the same as that in Pritsker (2002).[2] Pritsker uses the basic framework to examine how market liquidity differs across large investors. He also examines how shocks to investors' endowments or to their cash flow needs affect equilibrium asset holdings and prices. Pritsker also examines how rumours or news about potential future financial distress by one large investor affects asset prices, trades and asset market liquidity. This latter exercise is of special interest in the light of the increases in asset market illiquidity that occurred following rumours of financial distress at Long-Term Capital Management, one large investor in financial markets. Pritsker's results to date are as follows:

1. *Asset pricing*: When investors hold Pareto-optimal asset allocations, then asset prices are the same as those in a competitive setting and the CAPM is satisfied. If, instead, investors' asset holdings are not Pareto-optimal, then assets' excess returns over the riskless rate satisfy a multifactor model where one factor is the market portfolio and the other factors correspond to each large investor's endowment.

2. *Market liquidity*: The presence of large investors affects market liquidity. When all investors are small, markets are perfectly liquid: no single investor's order flow has price impact. When large investors are present, then their order flow has price impact and thus they face illiquid markets. Interestingly, the amount of liquidity that is available to different large investors differs with their risk aversion: the lower a large investor's risk aversion, the greater the price impact of his/her trades.[3] Because all information in the model is public, these differences in liquidity across large investors are not related to information asymmetry; instead they are a purely strategic reflection of the imperfect competition features of the model.

3. *Market manipulation*: Risk-sharing and shock absorption are affected by the presence of large investors because large investors behave strategically. In equilibrium, some large investors respond to shocks by following trading strategies which appear to be like front-running. By contrast, when markets are competitive, large investors do not engage in such behaviour. More specifically, when markets are competitive and one investor is hit with a shock which substantially increases his/her risky asset holdings, his/her excess risky asset holdings are rapidly purchased by other large investors, and the market returns to optimal risk-sharing within a single period of trade. By contrast, when there are large investors in the market, if one large investor is hit with a shock which increases his/her supply of risky asset holdings, in equilibrium the other large investors respond by initially selling (not purchasing) risky assets, and then later purchasing them back. The large investors' trades are optimal because they anticipate that future sales by the large investor who was hit with the shock will

[2] Urosevic (2001) does not examine a setting with many large investors and many risky assets. He instead examines a setting with one large investor and many risky assets, or one risky asset and many large investors. In conversations with DeMarzo and Urosevic I learned that Urosevic has solved the multi-asset, multi-large investor case in his PhD thesis. I solved the general multi-asset, multi-large investor case independently by extending the three-period, single large investor, single asset model of Kihlstrom (2001) to allow for multiple time periods, large investors and assets. My earliest work (based on Kihlstrom) only considered investors who live for a large but finite number of time periods. After reading DeMarzo and Urosevic, I modified my model to allow for infinitely lived investors who trade risky assets for a finite number of time periods. My results after these modifications are much more elegant than those that were derived in my earlier analysis.

[3] If liquidity is measured as the slope of the price function with respect to one large investor's trades while holding the trades of other large investors fixed, then investors who are less risk-averse receive less liquidity by this measure. Alternatively, if liquidity is measured as the magnitude of the price impact associated with an investor's selling assets to raise cash, then by this alternative measure, less risk-averse investors continue to receive less liquidity from other investors.

eventually cause prices to decline. They exploit the expected future price decline by front-running the sales, selling just after the shock (selling high) and then purchasing back shares as prices decline (buying low).

4. *Shock propagation*: Shocks to participants' positions in one market affect asset prices in other markets. The price deviations in other markets due to a shock in one market depend on the covariances and variances of the assets' dividend payments. Assets whose dividends do not covary are not susceptible to endowment shocks.[4] Cash flow shocks, ie shocks which cause an individual market participant to sell assets to meet a particular cash need, cause shocks to propagate by another route. In particular, cash flow shocks can cause co-movement between the prices of assets whose dividends are not correlated.[5]

The next two sections review the literature on liquidity and on large investors. The conclusion provides suggestions for further research.

2. Market liquidity

The purpose of this section is to review the literature on market liquidity. The review is structured to cover the definition of liquidity, the sources of illiquidity, the measurement of liquidity and whether liquidity is priced in asset markets.

2.1 Market liquidity defined

Imperfect market liquidity is synonymous with the notion that there are costs associated with transacting. These costs can be explicit, such as the spread between bid and ask prices in securities markets, or they can be implicit, such as the search costs associated with matching buyers with sellers. Liquidity costs are important to the market participants that expect to bear them. These include for example broker-dealers in options markets since those dealers need to dynamically trade through time to hedge their options book.[6] Liquidity also matters to investors who may not expect to trade frequently, but might need to sell assets to meet cash needs in unforeseen circumstances.[7]

Although imperfect market liquidity is synonymous with transaction costs, it is generally impossible to define or to precisely measure the amount of market liquidity associated with a particular asset because liquidity encompasses many different attributes of the implicit and explicit structure of transaction costs. For example, Kyle (1985) describes market liquidity in terms of three attributes of transaction costs: the tightness, depth and resilience of the market, where tightness measures the cost of quickly buying and then selling a position, depth refers to the size of a transaction that is required to change prices, and resilience measures the speed at which prices recover to fundamentals after a non-informational trade. Using these attributes, it should be clear that comparing individual assets' liquidities is problematic because one asset could be more liquid along one dimension of transaction costs while the other is more liquid in a different dimension.

---

4 The deviation of asset prices in market *j* due to a shock which reshuffles asset holdings in market *k* is equal to $\beta_{j,k}$ times the deviation of asset prices in market *k* where $\beta_{j,k}$ measures the covariance between dividends in markets *j* and *k* divided by the variance of dividends in market *k*.

5 There is a large literature on shock transmission within the contagion literature. Models of how contagion occurs through financial markets include Kodres and Pritsker (2002) and Kyle and Xiong (2001).

6 When markets are not statically complete, some market participants will find it optimal to follow a dynamic trading strategy. The standard example is a broker-dealer in a securities market who needs to dynamically hedge an option position.

7 For small investors such circumstances might include medical emergencies or loss of a job. For large institutional investors such circumstances might include paying large insurance claims or mutual fund redemptions.

2.2 The sources of market illiquidity

Despite the lack of a precise definition of illiquidity, it is possible to model why markets may not be liquid. Three sources of illiquidity are commonly used in the academic literature. The first is exogenous transaction costs (Constantinides (1986), Heaton and Lucas (1996), Vayanos (1998), Vayanos and Vila (1999) and Huang (2002)) such as those that might arise from order processing costs, or the costs of commissions. The costs associated with search are another source of exogenous transaction costs (Duffie et al (2001)). The second major source is asymmetric information about asset payoffs (Glosten and Milgrom (1985), Kyle (1985), Kyle (1989) and Eisfeldt (2001)) or about market participants' endowments (Cao and Lyons (1999), Vayanos (1999) and Vayanos (2001)). When there is asymmetric information about asset payoffs, prices change in response to trades because of the information that the trades might convey about asset fundamentals. The resulting price response to trades is an additional cost of transacting. Similarly, if some market participants (such as broker-dealers) have private knowledge of other investors' endowments, they might be able to predict future price movements; and they might trade on this knowledge. As a result, prices will respond to the potential information content of these trades. The third major source of illiquidity is imperfect competition in asset markets due to the presence of large traders (Lindenberg (1979), Kyle (1985), Kyle (1989), Basak (1997), Cao and Lyons (1999), Vayanos (1999), DeMarzo and Urosevic (2000), Vayanos (2001), Kihlstrom (2001) and Pritsker (2002)). As noted above, a trader is large relative to the size of the market if the scale of their desired trading activity would have the effect of causing prices to change.

In addition to the three most common sources, market illiquidity has also been modelled as resulting from Knightian uncertainty. This is the type of uncertainty that might occur if traders in financial markets confront circumstances that are completely unanticipated, and for which it is not clear how to proceed. Securities dealers when confronted with such circumstances may follow very risk-averse strategies which minimise their losses should anyone wish to trade with them. The resulting spreads can be so wide, and hence the market so illiquid, that trading does not take place at the equilibrium spreads (Cherubini and Della Lunga (2001) and Routledge and Zin (2001)). Imperfect market liquidity has also been modelled as resulting from optimal securities design since a firm which sells liquid and illiquid securities can use the differences between the securities' characteristics to price discriminate between investors who care about liquidity and those who do not (Boudoukh and Whitelaw (1993) and DeMarzo and Duffie (1999)).

2.3 How is market liquidity measured?

Liquidity is important to investors because it affects the costs at which they can trade assets. The goal of liquidity measurement is to identify the cost structure which confronts investors, and hence influences their decisions on which assets to hold and when they should be traded.

Because there are many dimensions of the relevant cost structure, there is no single method for measuring market liquidity. Measures which are typically used in the empirical literature on liquidity and asset pricing include bid-ask spreads, various measures of the price impact of order flow, and various measures of order flow.[8] Of these measures, the price impact of order flow is perhaps the one that is used most widely. The advantage of this measure is that it is based on the observed price changes associated with trades. The bid-ask spread is in some sense a more limited measure since it indicates the prices for standardised relatively small trades; as a result many transactions take place at prices other than the bid or the ask. Measures that are solely based on trading volume are also limited, but for a different reason. Trading volume-based measures do not measure the transaction costs associated with trading activity; high volume is typically associated with liquidity, yet it is clear that trading volume could be high and markets could be very illiquid.[9] Despite the advantages of using the price impact of order flow as a measure of liquidity, tricky econometric issues are involved when

---

[8] Measures of the price impact of order flow include price changes regressed on signed volume, or absolute price changes regressed on absolute volume, or daily price changes regressed on daily volume. Measures of volume include numbers of trades and daily volume measured in dollars.

[9] The day of the October 1987 stock market crash involved high volume because many participants wanted to sell stock, but liquidity was reportedly very poor.

using the approach to uncover the cost structure faced by investors when they decide to make a trade. The tricky issues are measurement error, selection bias and simultaneity bias.

Measurement error arises from two sources. The first is that in some specifications of the relationship between asset prices and trades, the appropriate measure of trades indicates whether each trade was initiated by a participant that wanted to buy or sell an asset. Because most data sources do not indicate which side initiated a trade, the designation of the side that initiated a trade is one source of measurement error. A second source of measurement error is due to the price and quote data; often the data which is available to the econometrician is not the same as that which is available to market participants. Differences between the true and observed prices and quotes arise because for some infrequently traded assets, although traders' perceptions of prices may be updated frequently, the publicly observed prices and quotes may only be updated after a trade takes place. If trades are spaced far enough apart in time, then because the notional prices and quotes before a trade are not publicly available, it is difficult to measure the price impact or quote revision associated with the trade. Hence, measures of the price impact of trades will be only imperfectly measured.

Selection biases arise in liquidity measurement because the trades that are observed within a sample are dictated by the amount of market liquidity. To take an extreme example, suppose that markets are highly liquid at some times, and not liquid at others. If trades only occur at the liquid times then measures of liquidity which are based on the price impact of the observed trades will tend to overstate liquidity because they are based only on the select sample of times in which the markets were liquid. Another way in which sample selection biases manifest themselves in this area is that some assets may be so illiquid at all times that investors who tend to do trades above a particular size simply will not take positions in that asset. As a result the illiquidity of those assets for large trades is not identified in the data.

The final source of bias is simultaneity bias. This bias arises when trades and prices are both determined by some other difficult to control for factor such as economic news. When both variables are driven by additional factors it can appear that trades move prices, suggesting a level of market illiquidity, even when there is no relationship between trades and prices.

Most attempts to measure market liquidity using trade and quote data are carried out in the context of the market microstructure literature.[10] The potential sources of noise and bias in liquidity measurement are no doubt well known in this literature. However, there is not a generally accepted methodology to control for these biases. The significance of the potential biases that cannot be controlled for is unknown, and remains an important topic for future research especially since these estimated measures of asset illiquidity are often used to determine whether illiquidity is priced by asset markets. It is to that subject that I now turn.

2.4 Is asset illiquidity priced in asset returns?

If investors care about liquidity risk, and it influences their trading behaviour, then perhaps it should be priced into asset returns. This section reviews some of theoretical and empirical literature on whether liquidity risk is priced.

Theory

One of the earliest theoretical contributions which relates market liquidity and equilibrium expected rates of return is the model of Amihud and Mendelson (1986). Amihud and Mendelson consider a setting with risk neutral investors who differ in the time horizons over which they wish to hold risky assets. The assets in this model vary in their liquidity, where liquidity is modelled as a fixed bid-ask spread. Their principal theoretical result is that there are clientele effects in asset holdings in which investors with short horizons prefer to hold assets with small bid-ask spreads and investors with long horizons prefer to hold assets with larger spreads. As a result of the clientele effects, assets with larger transaction costs are shown to earn larger gross returns, suggesting that asset illiquidity is priced. It is important to stress that the transaction costs in the Amihud and Mendelson model are

---

[10] For example, Huang and Stoll (1997) provide a market microstructure model of the determinants of bid-ask spreads. Stoll (2001) provides a recent review of the empirical and theoretical market microstructure literature.

deterministic, not stochastic. To examine whether there are systematic components to liquidity, and whether these components are priced, a model with stochastic liquidity is required.

Acharya and Pedersen (2002) present a model in which liquidity is stochastic. The model contains small investors who have CARA utility and face stochastic dividends and an exogenous stochastic transaction cost associated with selling assets. Their main insight is that returns net of transaction costs should satisfy the CAPM in this framework. They use this insight to solve for asset prices in an overlapping generations model framework in which each generation of investors lives for two periods. They show that within this framework, asset returns (not net of transaction costs) have a conditional four-factor structure with non-zero alpha:

$$E_{t-1}(r_t^i - r_t^f) = E_{t-1}c_t^i + \lambda_{t-1}\,Cov_{t-1}(r_t^i, r_t^m) + \lambda_{t-1}\,Cov_{t-1}(c_t^i, c_t^m) - \lambda_{t-1}\,Cov_{t-1}(r_t^i, c_t^m) - \lambda_{t-1}\,Cov_{t-1}(c_t^i, r_t^m) \quad (1)$$

where c measures transaction costs, and i and m denote asset i and the market portfolio respectively. This equation makes several contributions to the literature on liquidity. First, it shows how stochastic transaction costs fit into the general asset pricing framework. Second, it shows that when estimating asset pricing models using returns which do not net out transaction costs, then asset returns have a four-factor structure and a non-zero alpha which is related to expected transaction costs.

It is important to stress that the Acharya and Pedersen framework is not truly a four-factor model; the only true factor is the market portfolio. However, the model appears to have four factors because it is written in terms of gross returns (which are irrelevant to investors) instead of returns net of transaction costs (which investors care about).[11] This raises a second issue; since all investors in Acharya and Pedersen's framework have CARA utility and since the only sources of risk are traded asset risk, asset markets in their framework are effectively complete both statically and dynamically.[12] As a result, agents in this framework do not hold liquid assets for their insurance value in meeting unforeseen future cash needs. Further, market participants in Acharya and Pedersen's framework do not have incentives to hedge against changes in future market liquidity. This suggests that in a more realistic setting, asset liquidity may affect asset prices in ways which are not accounted for in this framework.[13]

An issue related to how illiquidity affects asset returns is how it affects asset prices. Duffie et al (2001) address this issue in the context of the prices of durable goods such as houses (or stocks). The source of illiquidity in the model is repeated adverse selection which arises because the seller of the asset knows more about its quality than the buyer.[14] A consequence of adverse selection is that a seller of a house may sometimes choose to forgo some favourable moving opportunities (such as career change) because the adverse selection problem prevents him/her getting a high enough price for the house. Duffie et al show that the discounted expected value of these future missed opportunities is built into the price of the house. A similar mechanism appears to be operating in financial markets with imperfect competition. In Pritsker (2002), imperfect competition in the asset markets causes traders to adjust their asset positions slowly towards Pareto-optimal asset allocations. The discounted future deviations from Pareto-optimal asset allocations are one determinant of the current price of the asset.

---

[11] Another way to see that there is only one factor is to note that the market price of risk of all the factors in the four-factor model are identical, indicating there is really only one factor.

[12] That is, asset prices are the same as they would be if a full set of Arrow-Debreu contingent securities was allowed to be traded in the economy.

[13] Holmstrom and Tirole (2001) differ from Acharya and Pedersen in that they examine how binding, borrowing constraints on corporate borrowers generate a desire for corporations to hold liquid assets to hedge against the market incompleteness generated by the borrowing constraint. Holmstrom and Tirole's analysis is related to corporations' need for liquidity, but it is not related to asset market liquidity where liquidity is measured as a transaction cost.

[14] Their model assumes that the purchaser of an asset may be imperfectly informed about asset quality at the time of purchase, but better informed at the time of asset sale. This is reasonable for houses, but less reasonable for financial assets such as stocks.

Empirical evidence

The early literature on liquidity and asset pricing was motivated by the framework in Amihud and Mendelson (1986), and thus studied whether stocks earn higher returns if they are less liquid, where liquidity is measured by the stock's bid-ask spread as a proportion of asset price. In their analysis, they regressed stocks' excess returns over the riskless rate on estimated market β's and on the proportional bid-ask spread. Their analysis suggested that assets with higher transaction costs, as measured by the spread, earn higher returns. Later work by Brennan and Subrahmanyam (1996) was unable to find reliable evidence that bid-ask spreads were priced. The sources of the differences from Amihud and Mendelson's earlier results are not resolved in their paper, but obvious candidates for explaining the differences are that Brennan and Subrahmanyam used a different econometric testing approach, and additionally they controlled for the factors that Fama and French (1993) showed appear to have power for pricing assets.[15] While Brennan and Subrahmanyam did not find evidence that bid-ask spreads were priced, they found evidence that the market rewarded stocks for which the price impact of trades was higher, where the price impact of trades was estimated based on a market microstructure methodology.

Several new empirical papers have been written on liquidity and asset pricing. These papers are motivated by recent empirical evidence that the liquidities of many assets tend to move together through time, suggesting that there are common factors which determine assets' market liquidity.[16] Pastor and Stambaugh (2001) create data series which measure time variation in the liquidity of individual stocks. They then use market-wide averages of these data series as a proxy for a systematic liquidity factor. The liquidity measure for individual firms is based on the tendency of a firm's excess returns over a market index to experience negative autocorrelation in returns over a two-day period given high trading volume on the first day. This approach builds on the notion that price changes that are due to illiquidity are mean-reverting, and that liquidity measures should be based on the magnitude of price changes relative to volume, with greater price change for a given volume interpreted as evidence of illiquidity.[17,18] They then test whether their measure of market illiquidity is priced by asset markets. They find strong evidence that it is priced even after controlling for the Fama and French (1993) factors. In my view, the Pastor and Stambaugh results are intriguing, but because it is not entirely clear whether they have found a proxy for liquidity or for something else, more work needs to be done on the properties of their proxies before their results can be viewed as entirely convincing.

Acharya and Pedersen (2002) also examine whether liquidity risk is priced, but they use a different measure from that used by Pastor and Stambaugh. Acharya and Pedersen's measures of liquidity are based on daily absolute price changes normalised by daily trading volume. This measure of liquidity is similar to that of Pastor and Stambaugh in that both account for the relationship of volume and price movement; however, the Acharya and Pedersen measure does not condition on the tendency for prices to reverse themselves. Acharya and Pedersen create their proxies of liquidity for individual stocks and for a proxy for the market portfolio. They find that the estimated coefficients on the liquidity variables tend to have the correct sign, and to be economically significant, but usually they are not estimated precisely enough to be statistically significant.

Chordia, Subrahmanyam and Anshuman (2001) also analyse the role of liquidity in asset pricing, but unlike most other analyses, they focus on variability in liquidity as proxied for by variability in measures of trading volume. They hypothesise that risk-averse investors should dislike variability in liquidity and thus stocks with more variability in liquidity should have lower prices and hence earn higher expected returns. In fact, their results are of a somewhat puzzling nature because they find strong evidence that the opposite of their hypothesis is true. My view is that the evidence in Chordia, Subrahmanyam and Anshuman may not be as puzzling as it seems, but that it instead points towards a need for more

---

[15] The Fama and French factors had not been discovered at the time that Amihud and Mendelson wrote their paper.

[16] Chordia et al (2000) document liquidity commonality for stocks. Chordia, Sarkar and Subrahmanyam (2001) show that there are common components which drive liquidity in the stock and bond markets.

[17] Recall that as noted earlier, if trades and prices are both responding to some other factor, such as economic news, then measures of liquidity which are based on the relationship between trades and prices can be misleading.

[18] Pastor and Stambaugh do not carefully justify why their liquidity measure for individual stocks is based on the autocorrelation of excess returns over a market index, as opposed to autocorrelation of the firms' returns.

theoretical research of how liquidity matters in portfolio choice. I discuss this point further in the conclusions.

The next section examines the literature on large investors, which is one source of market illiquidity.

3. Large investors

There has been enormous growth in the share of asset trades that are done by institutional investors. Since these investors often take large positions relative to the size of the markets in which they trade, even in the absence of other transaction costs, some markets may not be liquid from their perspective. Pritsker (2002) examines the role of institutional investors in determining market liquidity. The related literature on large investors in markets can be broadly broken down into three areas: why there are large investors; how they affect equilibrium asset pricing; and whether they stabilise or destabilise asset markets. Most of my discussion is related to the first two of these subjects. A comprehensive review of the literature on the third subject might require a separate paper.

3.1 Why are there large investors?

Many of the models of large investors and asset pricing take as primitives the set of large and small investors in the economy, their preferences and trading mechanisms, and then given this setup solve for the behaviour of asset prices. The contribution of some of the related literature on financial intermediation is that it derives the structure of the participants in financial markets; in essence it establishes why some investors in financial markets are large while others remain small and it establishes why small and large investors can coexist. Ideally, this literature can also be extended to model the behaviour of financial markets when there is intermediation. The full set of related financial intermediation literature is too large to review here. But I will discuss it briefly and highlight a few recent articles that are of interest.

Theories of financial intermediation provide a natural explanation for why there are large investors in financial markets since most financial intermediaries are large. Financial intermediaries such as insurance companies, banks, pension funds and mutual funds issue liabilities to small investors and then purchase assets to back up those liabilities. The traditional basis for why small investors enter into contracts with financial intermediaries includes pooling of risk (insurance companies), pooling of risk from liquidity needs (banks and mutual funds) and economising on the costs of monitoring borrowers (banks).[19] The growth of large institutions which is due to these sources has led to a deepening of markets, and to ever more complex financial products. This trend toward complexity is a self-reinforcing contributor to the increasing role of large institutional investors in markets and to the shrinking role of small investors, since large institutions are the only investors that can afford to pay the high fixed information costs associated with pricing and trading complex products (Allen and Gale (1999)).[20]

Although the amount of direct participation in markets by small investors is shrinking, there may be room for large institutional investors and small investors to both interact in markets. Two recent articles derive roles for small and large investors in the context of theories of mutual funds. The first, by Nanda and Singh (1998), emphasises the liquidity services provided by mutual funds. Their model has two types of small investors: one which is vulnerable to idiosyncratic future liquidity shocks à la Diamond and Dybvig (1983) and the other which is less vulnerable. The vulnerable investors invest in a mutual fund which holds sufficient liquid assets to meet their joint liquidity needs. They pay the mutual funds a

[19] Diamond and Dybvig (1983) model how banks pool liquidity risk. Nanda and Singh (1998) model how mutual funds economise on liquidity risk; they also analyse the relationship between the structure of mutual fund pricing and liquidity. Diamond (1984) models how financial intermediaries economise on monitoring costs.

[20] The high fixed information costs associated with learning to price and trade these products makes it prohibitively expensive for small investors to do so on their own. Instead small investors use large financial intermediaries to trade these products on their behalf. Since these intermediaries can spread the fixed costs of information over many small investors, it becomes economical for small investors to benefit from these products when intermediation is available.

fee for these liquidity services which is reflected in the funds providing performance which is not as good as the market return. Investors who are less vulnerable to liquidity shocks do not invest in the mutual fund and thus remain small. Mamaysky and Spiegel (2002) emphasise a different aspect of mutual funds, namely their ability to reduce the informational costs of following dynamic trading strategies. When markets are statically incomplete, it may be optimal for some investors to follow dynamic trading strategies. Because such strategies may involve closely monitoring market developments, the monitoring costs of a single investor implementing the strategy may be prohibitive. On the other hand, paying an institutional investor to follow a strategy which is customised to the optimal dynamic strategy of each small investor customer is also prohibitively expensive. In the light of these costs, Mamaysky and Spiegel argue that if families of mutual funds advertise funds which follow different dynamic strategies, then an investor who splits his/her wealth among different funds within the same fund family essentially creates a dynamic strategy which may come reasonably close to the optimal dynamic strategy and which may produce a better outcome than paying the costs of continually monitoring market developments. Of course, while there may be some small investors who use mutual funds for the purposes of dynamic strategies, the logic of their model suggests that some small investors for whom markets are “complete enough” will choose to remain small. The contribution of these models to modelling the behaviour of large investors is that they go some distance towards highlighting the differences in objectives of large and small investors. Ideally the differences in these objectives should be reflected in models of large investors and asset pricing. Unfortunately, for reasons of tractability, almost all large investor models assume that market participants have CARA utility or are risk neutral.[21] It is to these asset pricing models that I now turn.

3.2 Large investors and asset pricing

When investors are large enough that they do not take prices as given, then their non-price-taking behaviour is a deviation from the classical price-taking assumption. One issue which the large investor literature seeks to examine is how deviations from price-taking behaviour affect equilibrium returns in asset markets. There are two classes of model which examine how large investors affect equilibrium asset returns; the first class of models are those in which the presence of large investors is the only market imperfection; the second are those in which there are also additional sources of market imperfections. The additional market imperfections typically take the form of asymmetric information about asset payoffs or investors’ asset holdings. A second source of market imperfection involves agency problems between firm management and shareholders. To begin I will address models in which the only market imperfection is the presence of non-price-taking investors.

When there are investors who do not take prices as given, markets are not competitive by definition, hence one would expect that asset prices will vary from their levels in competitive asset markets. Lindenberg (1979) shows that this intuition is correct in static one-period models of asset prices. In particular he shows that when both large and small investors who have mean-variance utility are present, asset returns have a multifactor structure where one factor is the market portfolio and the other factors correspond to the endowments of large investors. By contrast, if the investors in Lindenberg’s model behave competitively, then it is well known that the CAPM holds.

It turns out that in multiperiod models markets are much more competitive than in the case considered by Lindenberg, and can produce the same asset prices as in competitive models in limiting cases. The basic intuition for why multiperiod models produce intense competition is based on Coase (1972). Coase argued that a monopolist selling durable goods today could not credibly commit to not sell the same goods in the future at a lower price, and therefore since durable goods today are close substitutes for durable goods in the future, his/her future and current sales would compete, forcing down prices. Moreover, if the time periods when the monopolist can sell are spaced arbitrarily closely together, then Coase conjectured that the competition across time periods would be so intense that the monopolist would be forced to charge the competitive price in the limit. Kihlstrom (2001) argues that the Coasian logic applies to sales of financial assets since they are also durable goods, and hence competition through time would force the sale price of stocks to be lower than in the monopolist case. DeMarzo and Urosevic (2000) consider an infinite horizon setting and show, in their model of a

[21] A notable exception is Basak (1997). However, because of intractability, Basak does not use his model to study whether the Coasian dynamics that are discussed in the next section are present in his framework, or how they would affect his results.

single large investor, that as the time between trades goes to zero, prices converge to those that would be found in a perfectly competitive model. They also find that when the time between trades is finite, then asset prices contain a risk premium term as a result of non-optimal risk-sharing among investors. In a setting with many large investors one would expect the competition to be more intense than when a single large investor is present, hence one would expect that Coasian dynamics would also be present in such a setting. Urosevic (2001) shows that Coasian dynamics are present in his model. They are also present in Pritsker (2002). Based on the literature with only a single market imperfection, it is clear that in a multiperiod setting, markets can be nearly as competitive as those in which perfect competition is present, with perfectly competitive markets as a limiting case.

When there are additional sources of market imperfections, they sometimes have the effect of reducing the competitiveness of multiperiod models. There is a large literature on the behaviour of large investors who have private information, dating back to Kyle (1985). Kyle's setting has a single large trader who is informed about the liquidation value (or end-of-day value) of a risky asset, noise traders who trade for reasons unrelated to fundamentals, and competitive market-makers who set prices equal to the expected fundamental liquidation value of the asset conditional on the information in traders' order flow. Kyle shows that the large investor's information is incorporated into prices through time, and that even in the limit as the time between trades gets small, the large investor's information is only slowly incorporated into prices, and importantly is incorporated much more slowly than it would be in a competitive framework. This suggests that competitive Coasian dynamics may not dominate the behaviour of asset prices when other sources of market imperfections, such as information asymmetries, are present. Vayanos (1999) considers a different type of information asymmetry, knowledge of large investors' private endowments. More specifically, he considers a setting in which there are several large investors who are subject to endowment shocks which only they observe. Hence they have private information about their own endowments. They trade in a multiple period setting by submitting linear demand curves. The resulting equilibrium price is that which clears markets. In this asymmetric information setting, Vayanos shows that a higher trading frequency does not cause asset prices to become competitive in the limit as the time between trades goes to zero; ie in this setting traders continue to hide their information. Vayanos also shows that if investors' endowments are public information, then the asset dynamics are the same as in a Coasian model, ie prices are competitive. Vayanos (2001) considers a different setting in which there is a single large trader, competitive market-makers and noise traders. The large trader receives endowment shocks as before, and these endowments are privately observed. In each time period, the large investor receives a private endowment shock, then the market-maker forms his/her optimal demand curve. The large trader takes this demand curve as given when choosing the quantity that he wishes to purchase. The large trader's order flow and the demands of the noise trader are submitted together at each market clearing. In each period the resulting price is set to clear the market. Then time passes and a new period starts. In the setting of Vayanos (2001), the informed investor's information is quickly revealed to the market, and asset prices quickly become competitive. It is not clear why information is revealed so quickly in Vayanos (2001) while it is revealed slowly in Kyle (1985) and Vayanos (1999).

The second type of market imperfection is models of agency problems. A standard agency problem is moral hazard resulting from firm management that cannot be perfectly monitored, and that in the absence of monitoring may choose to shirk on their duties by expending too little effort, or worse yet expropriate the shareholders' assets and instead spend them on salary and perquisites for the firm's management. If investors hold widely diversified portfolios, then an individual investor's incentive to monitor a particular firm is small since the benefits accruing to them are small (they hold few shares), but the costs of monitoring may be high. Moreover, each investor has an incentive not to monitor if they believe other investors will do it for them. This free-rider problem can result in an amount of monitoring which is socially suboptimal. If instead there is a large investor, then because their stake in the firm is relatively large, their incentives to monitor are larger as well; hence the presence of a large investor may help to overcome the free-rider problem. On the other hand, a large investor may be underdiversified, so there is a tension between optimal risk-sharing and monitoring. These issues are addressed and discussed by Admati et al (1994), DeMarzo and Urosevic (2000) and Urosevic (2001). One of the interesting findings in DeMarzo and Urosevic (2000) is that the speed of convergence of asset prices depends on whether the agency problem is "small". When it is small enough, asset prices quickly converge to their competitive values when the time between trades goes to zero, but when the agency problem is large enough, they do not, and instead the Coase conjecture does not hold in their setting.

Before proceeding, it is useful to summarise the theoretical results on large investors and asset pricing. It appears that if the time between trades is large enough, or if there are other market imperfections such as asymmetric information, then the presence of large investors may slow the rate at which participants adjust their positions towards optimal risk-sharing. As a result market prices will reflect deviations from optimal risk-sharing. On the other hand, even with some market imperfections, prices can appear to be very close to those in a competitive framework. In the end, whether large investors significantly affect equilibrium asset returns is an unsettled question.[22]

3.3 Do large investors stabilise markets?

One of the reasons that large investors receive so much attention is because they are so often blamed for speculating against currencies, or manipulating markets, causing exchange rate pegs to collapse. In addition, some empirical literature claims that large investors herd or engage in positive feedback trading, and that this activity can destabilise markets.

Some of the large investor models that were discussed in the previous subsection appear to have the feature that large investors can manipulate markets.[23] Vayanos (2001) finds circumstances in which the large investor appears to follow a market manipulation strategy in which he sells more shares to the market-makers than would be required for competitive risk-sharing, and then buys them back; this sell high, buy low strategy looks like market manipulation. Vayanos attributes the resulting price movements to the information asymmetry in his model. Pritsker (2002) also finds that large investors appear to engage in manipulative behaviour by responding to a positive endowment shock to one large investor by initially short selling stock to other price-taking investors, and then buying the stock back as prices decline. This behaviour looks similar to that in Vayanos, but there is no asymmetric information in Pritsker. This suggests information asymmetry may not be required to generate trades that appear to look like market manipulation. It may suffice to have a model with large investors.

One of the most important questions about the role of large investors in financial markets is whether their presence helps to coordinate speculative attacks on a currency. Corsetti, Dasgupta, Morris and Shin (2001) discuss this general issue. They find that the addition of a large investor into a financial market can cause other investors to attack a currency more aggressively, but the net effects of this activity can be small. However, if the large investor can signal his/her position (or trade) to small investors before they act, then other investors' ability to condition on the trades of the large investor help those investors solve a coordination problem, significantly increasing the prospects of the speculative attack's success. Corsetti, Pesenti and Roubini (2001) review the related empirical evidence on the role of large investors during the recent Asian crisis.

4. Conclusions

This paper has reviewed the literature on market liquidity and on large investors. The analysis points towards two areas where more research could be fruitful. The first is more theoretical research on how asset liquidity should affect asset returns. This theoretical research should motivate the empirical literature. To illustrate why such research might be useful, it is useful to first revisit the findings of Chordia, Subrahmanyam and Anshuman (2001). They find that variability in a proxy for assets' liquidity appears to be priced negatively, ie high variability of liquidity implies lower expected returns. One possible explanation for this finding is that variability in liquidity is valuable to investors because it has option value - that is, investors can hold stocks with high variability in liquidity in order to only trade the most liquid assets at any particular point in time. The associated reductions in transaction costs may make it very desirable to hold stocks with high variability in liquidity. As a result, these stocks

[22] Another reason why this question remains unsettled is the intractability of large investor models. The typical model assumes large investors are risk neutral or have CARA utility. These assumptions provide tractability, but they are not without loss of generality. I suspect functional forms for utility in which risk aversion depends on wealth would lead to different results - if anyone could solve a large investor model with such utility functions.

[23] For a model of market manipulation, see Jarrow (1992).

might be expected to have lower expected returns.[24] If this explanation is correct, it suggests that to properly model why liquidity matters, theoretical models need to consider the dynamics of asset trading and assets' liquidity.

An additional area where more research might be fruitful is in empirically relating liquidity premia to institutional investors' asset holdings. Since institutional investors may focus more on liquidity than small investors, careful studies of institutional investors' trading strategies, with a particular focus on the choices of the assets that they choose to hold and choose not to hold, may help to contribute to our understanding of how asset liquidity affects equilibrium stock returns.

Bibliography

Acharya, V V and L H Pedersen (2002): "Asset pricing with liquidity risk", *Working Paper*, Stern School of Business, New York University.

Admati, A, P Pfleiderer and J Zechner (1994): "Large shareholder activism, risk sharing, and financial market equilibrium", *The Journal of Political Economy*, 102, 1097-130.

Allen, F and D Gale (1999): "Innovations in financial services, relationships, and risk sharing", *Management Science*, 45, 1239-53.

Amihud, Y and H Mendelson (1986): "Asset pricing and the bid-ask spread", *Journal of Financial Economics*, 17, 223-49.

Basak, S (1997): "Consumption choice and asset pricing with a non-price taking agent", *Economic Theory*, 10, 437-62.

Boudoukh, J and R F Whitelaw (1993): "Liquidity as a choice variable: a lesson from the Japanese government bond market", *Review of Financial Studies*, 6, 265-92.

Brennan, M J and A Subrahmanyam (1996): "Market microstructure and asset pricing: on the compensation for illiquidity in stock returns", *Journal of Financial Economics*, 41, 441-64.

Cao, H H and R K Lyons (1999): *Inventory information*, mimeo, Haas School of Business, University of California at Berkeley.

Chan, L K C and J Lakonishok (1995): "The behavior of stock prices around institutional trades", *Journal of Finance*, 50, 1147-74.

Cherubini, U and G Della Lunga (2001): "Liquidity and credit risk", *Applied Mathematical Finance*, 8, no 2, 79-95.

Chordia, T, R Roll and A Subrahmanyam (2000): "Commonality in liquidity", *Journal of Financial Economics*, 56, 3-28.

Chordia, T, A Sarkar and A Subrahmanyam (2001): "Common determinants of bond and stock market liquidity: the impact of financial crises, monetary policy, and mutual fund flows", *Working Paper*, The Federal Reserve Bank of New York.

Chordia, T, A Subrahmanyam and V R Anshuman (2001): "Trading activity and expected stock returns", *Journal of Financial Economics*, 59, 3-32.

Coase, R (1972): "Durability and monopoly", *Journal of Law and Economics*, 15, 143-9.

Constantinides, G M (1986): "Capital market equilibrium with transaction costs", *Journal of Political Economy*, 94, 842-62.

Corsetti, G, A Dasgupta, S Morris and H S Shin (2001): *Does one Soros make a difference? A theory of currency crises with large and small traders*, mimeo, Yale University.

[24] Chordia, Subrahmanyam and Anshuman mentioned this optionality aspect as one possible explanation for their findings in early drafts of their paper but later removed references to it because they did not have a way of empirically testing the explanation.

Corsetti, G, P Pesenti and N Roubini (2001): *The role of large players in currency crises*, mimeo.

DeMarzo, P M and D Duffie (1999): "A liquidity-based model of security design", *Econometrica*, 67, 65-99.

DeMarzo, P M and B Urosevic (2000): "Optimal trading by a large shareholder", *Working Paper*, Haas School of Business, University of California at Berkeley.

Diamond, D W (1984): "Financial intermediation and delegated monitoring", *Review of Economic Studies*, 51, 393-414.

Diamond, D W and P Dybvig (1983): "Bank runs, deposit insurance, and liquidity", *Journal of Political Economy*, 105, 928-56.

Duffie, D, N Gârleanu and L H Pedersen (2001): "Valuation in dynamic bargaining markets", *Working Paper*, Graduate School of Business, Stanford University.

Eisfeldt, A (2001): *Endogenous liquidity in asset markets*, mimeo, Kellog Graduate School of Management, Northwestern University, April.

Fama, E F and K R French (1993): "Common risk factors in the returns on stocks and bonds", *Journal of Financial Economics*, 33, 3-56.

Glosten, L R and P R Milgrom (1985): "Bid, ask, and transaction prices in a specialist market with heterogeneously informed traders", *Journal of Financial Economics*, 14, 71-100.

Heaton, J and D J Lucas (1996): "Evaluating the effects of incomplete markets on risk sharing and asset pricing", *Journal of Political Economy*, 104, 443-87.

Holmstrom, B and J Tirole (2001): "LAPM: a liquidity-based asset pricing model", *The Journal of Finance*, 56, 1837-67.

Huang, M (2002): "Liquidity shocks and equilibrium liquidity premia", *Journal of Economic Theory*, forthcoming.

Huang, R D and H R Stoll (1997): "The components of the bid-ask spread: a general approach", *Review of Financial Studies*, 10, 995-1034.

Jarrow, R A (1992): "Market manipulation, bubbles, corners, and short squeezes", *Journal of Financial and Quantitative Analysis*, 27, no 3, 311-36.

Kihlstrom, R (2001): "Monopoly power in dynamic securities markets", *Working Paper*, The Wharton School, University of Pennsylvania.

Kodres, L E and M Pritsker (2002): "A rational expectations model of financial contagion", *Journal of Finance*, 57, 769-99.

Kyle, A S (1985): "Continuous auctions and insider trading", *Econometrica*, 53, 1315-35.

——— (1989): "Informed speculation with imperfect competition", *Review of Economic Studies*, 56, 317-56.

Kyle, A S and W Xiong (2001): "Contagion as a wealth effect", *Journal of Finance*, 56, 1401-40.

Lindenberg, E B (1979): "Capital market equilibrium with price affecting institutional investors", in *Portfolio theory 25 years after: essays in honor of Harry Markowitz*, Elton, E J and M J Gruber, eds, North Holland, pp 109-24.

Mamaysky, H and M Spiegel (2002): "A theory of mutual funds: optimal fund objectives and industry organization", *Working Paper*, Yale School of Management.

Nanda, V and R Singh (1998): "Mutual fund structures and the pricing of liquidity", *Working Paper*, University of Michigan Business School.

Pastor, L and R F Stambaugh (2001): "Liquidity risk and expected stock returns", *Working Paper*, University of Chicago Graduate School of Business.

Pritsker, M (2002): *Large investors, implications for equilibrium asset returns, shock absorption, and liquidity*, mimeo, Federal Reserve Board.

Routledge, B R and S E Zin (2001): *Model uncertainty and liquidity*, mimeo, July.

Stoll, H R (2001): “Market microstructure”, *Working Paper* 01-16, Financial Markets Research Center, Vanderbilt University.

Urosevic, Branko (2001): “Dynamics of insider ownership: theory and evidence”, *Working Paper*, Haas School of Business.

Vayanos, D (1998): “Transaction costs and asset prices: a dynamic equilibrium model”, *Review of Financial Studies*, 11, 1-58.

——— (1999): “Strategic trading and welfare in a dynamic market”, *Review of Economic Studies*, 66, 219-54.

——— (2001): “Strategic trading in a dynamic noisy market”, *Journal of Finance*, 56, 131-71.

Vayanos, D and J L Vila (1999): “Equilibrium interest rate and liquidity premium with transaction costs”, *Economic Theory*, 13, 509-39.

Hedging demand and foreign exchange risk premia

David Tien[1]
Haas School of Business, UC Berkeley

Abstract

This paper develops and tests a model of unobservable risk premia in the foreign exchange market. Risk premia in our model arise from non-marketable income shocks that risk-averse agents hedge by trading foreign currency. We construct a proxy for currency hedging demand and find that it explains approximately 45% of the variation in currency returns at the monthly horizon. We find that hedging demand appears to Granger cause speculative flows. We also show that hedgers exhibit negative feedback behaviour. Our results show that correlation between order flow and currency returns is consistent with risk-sharing among market participants.

1. Introduction

Exchange rate economics has long struggled to reconcile the empirical behaviour of currency fluctuations with rational theories of exchange rate dynamics. Numerous studies (see Meese and Rogoff (1983) and Flood and Rose (1995)) have demonstrated the failure of models based on macroeconomic fundamentals to explain a significant proportion of the variation in exchange rates at horizons of one year or less.[2] Recent research applying tools from the market microstructure literature has been more successful in explaining currency dynamics in terms of order flows between various types of agents; see Lyons (2001) for a recent survey of the literature. The current interpretation of the results from the FX microstructure literature is somewhat counterintuitive. Many researchers take the observed correlation between order flow and currency returns as evidence that some traders have private information. The existence of asymmetric information in the currency market runs counter to the general perception that currency markets are among the most informationally efficient markets in existence.[3]

The key to reconciling the existence of asymmetric information with the perceived informational efficiency of the foreign exchange market lies in identifying the nature of the informational asymmetry. If certain traders have private information about the distribution of endowment shocks or changing risk appetites across the economy, then the market serves as a mechanism for distributing information to aid in the optimal allocation of risk across all agents as described by Hayek (1945). If, however, some traders have information regarding future statistical releases or central bank policy changes, the market is still serving as a mechanism to disseminate information, but the liquidity of the market may be quite low as other traders would be hesitant to trade against a better informed counterparty. A better understanding of the information structure in currency markets would give us a clearer picture of the role of speculators in the market: are they a stabilising influence as posited by Friedman or the scourge of financial markets as described by some leaders of emerging market countries?

1 I thank Ash Alankar, John Briginshaw, Mintao Fan, Hayne Leland, Richard Lyons, Terry Marsh, Mark Rubinstein, Jacob Sagi, Mark Seasholes and Harry Stordel as well as seminar participants at UC Berkeley, Santa Clara University and the Third Joint Central Bank Research Conference on Risk Measurement and Systemic Risk for helpful comments and suggestions. I am indebted to Bob Lawrie and Mike Rosenberg for their insights and access to their data.

2 There is evidence that standard macroeconomic models have significant explanatory power over longer horizons; see Flood and Taylor (1996).

3 A recent BIS study estimates daily turnover in the spot foreign exchange market at USD 1.5 trillion. Also, past studies have shown that the over-the-counter currency markets trade billions of dollars with a bid-ask spread in the neighbourhood of a few hundredths of a cent.

This paper demonstrates that the observed relationships between trading variables and currency returns are completely consistent with a market where the only motive for trade is risk-sharing. We develop and test a simple model of the foreign exchange risk premium where non-marketable cash flows generate hedging demands from risk-averse agents. We derive equilibrium hedging demands and risk premia in an economy with two types of risk-averse agents: hedgers who face non-marketable risks, and liquidity providers who stand ready to share risk with the hedgers for the right price. Our empirical tests of the model make use of the fact that hedging demands are proportional to the foreign exchange risk premium. Using data from the currency futures markets, we construct a proxy for hedging demand in currency futures and test for the existence of unobservable systematic risk factors in five major currencies. The market for currency futures is a natural setting in which to test the implications of our model. First, the zero entry cost for futures, standardised contract specifications and relatively low transaction cost make these markets very liquid, attracting a wide variety of traders. Second, the open outcry nature of the futures pits adds a measure of transparency that serves to discourage informed speculators from entering these markets, allowing us to more accurately measure hedging flows. Lastly, the small size of the currency futures market relative to the entire over-the-counter market diminishes the likelihood of price pressure in the futures market affecting the aggregate market for spot foreign exchange.[4]

Our results affirm the interpretation of hedging demand as a proxy for risk premia. We find that, on average, hedging demand explains 45% of the variation in currency returns. To examine what might be driving this result, we consider the forecasting power of hedging demand over future realisations of bilateral trade balances and find that hedging demand has significant forecasting power over these flows for the Canadian dollar and Japanese yen. We then compare the performance of hedgers versus non-hedgers in the futures market. We find, intuitively, that hedgers tend to lose money to non-hedgers. These losses can be interpreted as compensation to speculators for insuring the hedgers. Along these lines, we also test causal relationships between hedging and speculative flows and find, consistent with the theory, that changes in hedging demand Granger cause changes in speculative demands in four of the five currencies at the weekly level. We also test to see if the observed effects could be due to some type of positive feedback trading on the part of hedgers, and find that hedgers actually tend to be negative feedback traders. Lastly, to see if these results are directly related to the findings of Evans and Lyons (2002), we compare hedging demands to data on customer order flow from a major international bank. We find that futures market hedging demand is not related to the aggregate order imbalance in customer order flow.

There is a vast literature which tries to explain the short-run variability of exchange rates. Previous studies of the foreign exchange risk premium[5] have examined the conditional variance of exchange rates as a proxy for risk premia (Domowitz and Hakkio (1985)), considered consumption-based CAPM models (Mark and Wu (1998)) and examined the possibility of “peso problem” effects (Evans (1996)).[6] The difficulty in identifying a risk premium in currency returns is analogous to the “equity premium” puzzle in the asset pricing literature. Observed fundamentals do not appear volatile enough to justify the volatility of floating exchange rates. Other attempts to identify the risk premium have used survey data on exchange rate forecasts of market participants to control for expectational errors (Frankel and Froot (1989)) and statistical models to identify time-varying risk premia (Baillie and Bollerslev (1994)). Some non-risk-related explanations of the forward discount bias include irrationality (Froot and Thaler (1990)), regime shifts driven by policy changes (Engel and Hamilton (1990)) and learning (Roberts (1995)).

This study is similar in spirit to recent research on the microstructure of the foreign exchange market and the literature on futures risk premia. Evans and Lyons (2002) show that signed order flow in the inter-dealer market possesses significant explanatory power for exchange rate returns, but their model is agnostic as to whether the results are driven by private information about future returns or

---

4 The aggregate notional amount of outstanding positions in the currency futures market is USD 103 billion compared with USD 1.5 trillion in average daily volume for the spot foreign exchange market. Though spot-futures arbitrage may confound some results, even these flows should be minuscule relative to the entire market.

5 The literature on exchange rate risk premia is vast; see Engel (1996) for a recent survey.

6 The term “peso problem” refers to the possibility of agents attaching a small probability to some extreme event that has not yet been observed in the data. The term comes from the experience of the Mexican peso in the 1970s when agents appeared to expect a huge devaluation despite the fact that such an event had never been observed.

risk-sharing motives. This study extends this literature in two ways. First, the data set used here spans 15 years at a monthly frequency, providing a much more expansive study than any previous research using trading activity in the currency market. Second, by focusing on hedging demand, we are able to more clearly identify the link between currency returns, some macro fundamentals and time-varying risk premia. Research on futures risk premia is also closely related to this study. Bessembinder (1992) and de Roon et al (2000), testing a theory of futures pricing developed in Hirshleifer (1990), find that hedging pressure risk is priced in the futures market. We extend their results to show that the effects they observed appear to affect the broad market for foreign exchange.

This research has policy implications in the debate on the transaction taxes in the currency market; see Eichengreen et al (1995). The case for transaction taxes rests on the assumption that irrational traders can destabilise currencies by engaging in positive feedback strategies or herding together to drive exchange rates away from their fundamental values. Our results indicate that, at least for the major currency markets, the imposition of a transaction tax could have significant welfare implications for firms trying to hedge their exposures to currency risk.

This paper is organised as follows. Section 2 develops the model relating hedging demand and risk premia. Section 3 describes the data used in this paper, while Section 4 outlines our estimation procedures and discusses the results. Section 5 concludes.

2. The model

In this section, we develop a model that relates expected returns on foreign exchange to observable variables, namely interest rate differentials and hedging demand. Our model is unique in that it bridges the gap between traditional asset pricing and microstructure models by using trading variables as proxies for risk premia derived in a standard risk-sharing environment. This feature is important in that previous work has interpreted the strong contemporaneous correlation between order flow and currency returns as evidence of the existence of private information in the foreign exchange market. Our model shows that the observed correlations are consistent with risk-sharing in a symmetric information environment. The model developed in Wang (1994) is similar in some respects to our model in that he also links the microstructure variables to expected returns by exploring the behaviour of trading volume in dynamic rational expectations economies.

The theoretical setting we consider is a simple economy populated by two types of risk-averse agents with utility functions that exhibit constant absolute risk aversion (CARA): hedgers and speculators. Agents can invest in either domestic or foreign risk-free bonds that yield r_{t-1}^{D} and r_{t-1}^{F} from period t - 1 to t, respectively. In order to purchase foreign bonds, agents must purchase foreign currency; P_t is the amount of domestic currency that can be exchanged for one unit of foreign currency. Hedgers are unique in that they also receive a non-tradable flow of stochastic income that yields r_t^N and is correlated with exchange rate returns. To close the model, we assume that all agents have symmetric information and that all assets are in zero net supply.[7]

Let W_t^h be the wealth of the hedger at time t

$$W_t^h = \left(1 + r_{t-1}^F\right)\left(1 + r_t^p\right)\lambda_t^h W_{t-1}^h + \left(1 + r_{t-1}^D\right)\left(1 - \lambda_t^h - \overline{\beta}\right)W_{t-1}^h + \left(1 + r_t^N\right)\overline{\beta}\, W_{t-1}^h \tag{1}$$

where λ_t^h is the hedger's portfolio holding of foreign assets chosen at time t - 1 and held through time t, $\overline{\beta}$ is the fixed proportion of wealth that the non-tradable income stream comprises, and r_t^p is the foreign currency return. In other words, if r_t^p is positive, then the foreign currency has appreciated relative to the domestic currency. Note that the domestic return on foreign bonds, $\left(1 + r_{t-1}^F\right)\left(1 + r_t^p\right)$, is

[7] The assumption that all assets are in zero net supply is purely for mathematical convenience; all of the results are essentially unchanged if there is a positive net supply of foreign and domestic bonds.

approximately equal to $1+r_{t-1}^{F}+r_{t}^{p}$.[8] Using this approximation, we can write down the wealth of the hedger as

$$W_t^h = W_{t-1}^h + \left(r_{t-1}^F + r_t^p\right)\lambda_t^h W_{t-1}^h + r_{t-1}^D\left(1-\lambda_t^h-\overline{\beta}\right)W_{t-1}^h + \left(r_t^N\right)\overline{\beta}\, W_{t-1}^h \tag{2}$$

Similarly, the wealth of the speculator at time *t* is W_t^s where

$$W_t^s = W_{t-1}^s + \left(r_{t-1}^F + r_t^p\right)\lambda_t^s W_{t-1}^s + r_{t-1}^D\left(1-\lambda_t^h\right)W_{t-1}^s \tag{3}$$

There are only two sources of uncertainty in this economy: exchange rate risk and the stochastic non-tradable income. We assume that both random variables are conditionally normally distributed:

$$r_t^p = E_{t-1}\left[r_t^p\right] + \varepsilon_t\,, \qquad \varepsilon_t \sim N\left(0,\sigma_\varepsilon^2\right)$$

$$r_t^N = \theta + \eta_t\,, \qquad \eta_t \sim N\left(0,\sigma_\eta^2\right)$$

$$E_{t-1}\left[\varepsilon_t\eta_t\right] \equiv \sigma_{t-1}^{\varepsilon\eta} \neq 0$$

where θ is a fixed scalar. For simplicity, we assume that both the hedger and speculator have CARA preferences and maximise utility over wealth next period. In this setting, agents effectively maximise the one-period return on wealth, r_t^{Wh} and r_t^{Ws}, for the hedger and speculator, respectively. Their preferences imply that their choices will only depend on the mean and variance of return on wealth.

$$E_{t-1}\left[r_t^{Wh}\right] = \left[\lambda_t^h\left(r_{t-1}^F + E_{t-1}\left[r_t^p\right]\right) + \left(1-\lambda_t^h\right)r_{t-1}^D + \theta\overline{\beta}\right] \tag{4}$$

$$Var_{t-1}\left[r_t^{Wh}\right] = \left(\lambda_t^h\right)^2\sigma_\varepsilon^2 + \overline{\beta}^2\sigma_\eta^2 + 2\lambda_t^h\overline{\beta}\sigma_{t-1}^{\varepsilon\eta} \tag{5}$$

The expressions for the mean and variance of the speculator are very similar and omitted for the sake of clarity. Thus, the hedger's investment problem is equivalent to

$$\max_{\lambda_t^h} E_{t-1}\left[r_t^{Wh}\right] - \frac{1}{2}\rho^h Var_{t-1}\left[r_t^{Wh}\right] \tag{6}$$

where ρ^h is the hedger's coefficient of constant absolute risk aversion. Taking the first-order conditions for (6) and solving for the hedging demand yields

$$\lambda_t^h = \frac{\left(r_{t-1}^F - r_{t-1}^D\right) + E_{t-1}\left[r_t^p\right] - \rho^h\overline{\beta}\sigma_{t-1}^{\varepsilon\eta}}{\rho^h\sigma_\varepsilon^2} \tag{7}$$

Similarly, the speculative demand is

$$\lambda_t^s = \frac{\left(r_{t-1}^F - r_{t-1}^D\right) + E_{t-1}\left[r_t^p\right]}{\rho^s\sigma_\varepsilon^2} \tag{8}$$

Since the bonds are in zero net supply, combining (6) and (7) and imposing market clearing, ie $\lambda_t^h + \lambda_t^s = 0$, yields

$$E_{t-1}\left[r_t^p\right] = \left(r_{t-1}^D - r_{t-1}^F\right) + \frac{\rho^h\rho^s}{\rho^h+\rho^s}\overline{\beta}\sigma_{t-1}^{\varepsilon\eta} \tag{9}$$

Equation (9) shows that the foreign exchange risk premium is driven by the covariance of the non-tradable income shocks and exchange rate returns. Unfortunately, these income shocks are unobservable to the econometrician. To find an observable proxy for the risk premium, we can substitute (9) into (7), which yields

---

[8] $\left(1+r_{t-1}^F\right)\left(1+r_t^p\right) = 1 + r_{t-1}^F + r_t^p + r_{t-1}^F r_t^p \approx 1 + r_{t-1}^F + r_t^p$. Note that, at the monthly level, bond and currency returns are likely to be less than 1% per month, implying that the term $r_{t-1}^F r_t^p$ will generally be less than 0.01%.

$$\lambda_t^h = -\frac{\rho^h}{\left(\rho^h + \rho^s\right)\sigma_\varepsilon^2}\,\overline{\beta}\,\sigma_{t-1}^{\varepsilon\eta} \tag{10}$$

Rearranging (10) and substituting back into (9) yields

$$E_{t-1}\left[r_t^p\right] = \left(r_{t-1}^D - r_{t-1}^F\right) - \left(\sigma_\varepsilon^2 \rho^s\right)\lambda_t^h \tag{11}$$

Equation (11) is now completely in terms of observables and can be estimated with the data.

The model developed above is very much in the spirit of the consumption-based CAPM developed by Rubinstein (1976) and Breeden (1979). Demand for foreign exchange is driven by the desire of hedgers to purchase assets that hedge their stochastic income stream; as their income becomes more correlated with currency returns, they demand less. Though we have assumed a single source of income uncertainty, multiple sources of uncertainty would increase or decrease hedging demands depending on their covariances. The model is also related to portfolio balance models of exchange rate determination described in Branson and Henderson (1985). In that class of models, currency risk premia arise from the imperfect substitutability of foreign and domestic bonds. In our model, foreign and domestic bonds are not perfect substitutes because the non-tradable income stream received by hedgers is correlated with currency fluctuations, making foreign bonds effective hedging instruments.

Though we do not explicitly model the random income shock, one can think of it as a domestic firm's income from a foreign subsidiary that repatriates profits quarterly or as receipts to a firm that exports goods overseas. Our non-marketable income stream is consistent with Obstfeld and Rogoff (2000) in that nominal price rigidities, pricing to market, and trading costs could induce non-tradable income shocks that cause firms to hedge in the futures market. From an asset pricing point of view, the non-marketability of the uncertain income stream violates the necessary conditions for a representative agent representation for this economy and forces us to identify an observable proxy for the risk premium.

3. The data

We use monthly observations on the aggregate positions of commercial traders in the currency futures markets to construct our hedging demand proxy; these data are collected and distributed by the Commodity Futures Trading Commission (CFTC). Our data set includes five currency futures contracts, the Canadian dollar (CAD), Swiss franc (CHF), Deutsche mark (DEM), pound sterling (GBP) and Japanese yen (JPY) over the period from January 1986 to December 2000.[9]

In each market, the CFTC classifies large traders as either commercial or non-commercial, where a trader is typically classified as a commercial trader if she is "engaged in business activities hedged by the use of the futures or option markets".[10] We follow Bessembinder (1992) and de Roon et al (2000) and treat commercial traders as hedgers and non-commercial traders as liquidity providers. These positions are reported to the public on a weekly basis in the Commitment of Traders Report; the reported positions typically account for 70-80% of the open interest in any given contract; summary statistics for each contract are reported in Table 1; note that these statistics are for the period January to December 2000.[11]

We form our measure of hedging demand in each currency as

$$h_t = \frac{\textit{number of long hedge contracts} - \textit{number of short hedge contracts}}{\textit{total number of hedge contracts}} \tag{12}$$

---

[9] We only study the Deutsche mark up to the introduction of the euro in January 1999.

[10] From the Commitment of Traders Report Backgrounder, CFTC, October 2000, available at http://www.cftc.gov/opa/backgrounder/opacot596.htm.

[11] The percentages reported for bank participation are for December 2000 only, but are fairly representative of average participation.

This definition was used in de Roon et al (2000) and is simply the *relative* net position of hedgers in the market. This is a natural measure of hedging activity because it captures the net portfolio weight the average hedger has in each currency. Summary statistics on the statistical properties of h_t are also reported in Table 1.

Figures 1 and 2 plot the historical path of the spot Japanese yen exchange rate and the hedging demand, h_t, for the yen, respectively. An alternative measure of hedging demand used in Bessembinder (1992) is the *absolute* net position of hedgers, or

$$h_t^a = \textit{number of long hedge contracts} - \textit{number of short hedge contracts}$$

We also construct speculative demand proxies analogously. We construct a relative net speculative demand series, x_t, and an absolute net speculative demand series, x_t^a

$$x_t = \frac{\textit{number of long non-commercial contracts} - \textit{number of short non-commercial contracts}}{\textit{total number of non-commercial contracts}}$$

$$x_t^a = \textit{number of long non-commercial contract} - \textit{number of short non-commercial contracts}$$

We use the absolute net demand measures in Section 4.3 when we examine the causal relationship between hedging and speculative activity.

Identifying the major players in the currency futures markets is quite difficult. Using a similar data set, Kodres and Pritsker (1995) find that commercial banks, broker-dealers and hedge funds typically account for approximately 35% of the open interest in currency futures markets. Their study, however, did not include non-financial corporations. Anecdotal evidence suggests that the currency futures markets probably mirror activity in the interbank market. Major corporations typically do not transact in the futures market because they face very low transaction costs in spot and forward markets. Major currency dealers occasionally use futures markets to lay off inventory risk with hedge funds, commodity trading advisors (CTAs) or other "local" traders. Since commercial banks are classified as commercial traders by CFTC guidelines, it is likely that the dynamics in hedging activity are driven by changes in positioning by interbank dealers.

To compare the behaviour of trading currency futures to the spot market in foreign exchange, we also utilise a database of customer-dealer trades done by a major international bank.[12] This database contains over 800,000 transactions in all spot currency markets over the period from January 1998 to March 2000. While we have some data at the transaction level (ie customer locale, transaction size and rate), transactions are not time-stamped. We aggregate these trades to make them comparable to our futures data set. Anecdotal evidence suggests that the bank's customer base was fairly diverse and included a significant proportion of hedge fund customers along with more traditional corporate customers.

We use spot exchange rate data released by the Federal Reserve Bank of New York. These rates are collected daily at 12 pm Eastern Standard Time. Our forward rate data consist of 30-day forward rates obtained from Datastream. In calculating our currency returns and expected depreciation, we restate all spot and forward exchange rates in terms of US dollars per unit of foreign currency to remain consistent with our modelling framework. Our data on bilateral trade flows come from the US Census Bureau.[13]

4. Estimation and empirical results

In this section, we test the model developed in Section 2 and perform some robustness checks against plausible alternative explanations for the results. The first set of results directly test (11) in Section 2. The next subsection explores the relation between hedging demand and future goods trade. The

[12] Estimates of this bank's market share in the spot FX market range around 10%.

[13] Data on bilateral trade flows are available from the Census Bureau's website at http://www.census.gov/foreign-trade/www/.

following subsections test the robustness of the results to two plausible alternatives: the private information hypothesis and the positive feedback trading hypothesis. This section concludes by comparing hedging demand and customer order flow to see if our results are generic to any type of order flow.

4.1 Hedging demand and exchange rate dynamics

Table 2 documents the results of standard uncovered interest parity (UIP) regressions on the five currencies we study, where p_t is the natural logarithm of the spot exchange rate quoted in terms of US dollars per unit of foreign currency at time t and f_t is the 30-day forward rate as of time t. The estimates in Tables 2 and 3 were calculated taking all five currencies as a system of equations using a generalised least squares (GLS) framework. GLS provides uniformly better estimates than OLS equation by equation in cases where the residuals are correlated across equations, as is likely to be the case here because all of the exchange rates we study are US dollar-based. The results in Table 2 mirror the findings of previous studies. The forward discount has extremely poor explanatory power over future changes in spot rates. The well documented forward discount bias is evident in the coefficients for the Swiss franc and the Japanese yen, ie that the β for these currencies is negative. These results are troubling because all but one of the coefficients are significantly less than one.

Table 3 reports the results of the regression which implements (11).[14] First, note how the coefficients on the forward discount term for the yen and franc have become more positive while the coefficient in the pound equation has basically remained unchanged. The β coefficients for the Canadian dollar and Deutsche mark have both become more negative, but these results may be confounded by current account flows in the case of the Canadian dollar and euro convergence trading for the Deutsche mark. Second, the coefficients on the hedging demand term are negative and significantly different from zero for all currencies; the sign of the coefficients is consistent with the theory. The sign of the coefficient indicates that when hedgers buy yen forward, for instance, the yen tends to depreciate relative to the US dollar, ie hedgers tend to lose money. Third, Table 3 also reports the implied price impact of trading 10,000 contracts for each market. Interestingly, the price impact of 10,000 contracts (roughly USD 1 billion for all contracts) is similar in magnitude to the price impact estimated in Evans and Lyons (2001). Finally, the adjusted R^2 for all of the equations has increased dramatically.

The impact of adding a hedging demand variable to the UIP regression is very similar to the effect observed in Evans and Lyons (2002), where they use signed inter-dealer order flow instead of hedging demand. An important difference between our results and the previous microstructure literature lies in the time period and horizon studied. By working at the monthly horizon over a 15-year sample, our results conclusively show that the effects we observe are economically meaningful and persistent. Another key difference between their work and this research is that we explicitly attribute the relationship between order flow and returns to a hedging motive. In the portfolio shifts model developed by Evans and Lyons (2002), the initial customer order flow which drives trading for the rest of the day is exogenous; it can be driven by either private information or hedging. Thus, their model cannot distinguish between informed speculation and risk-sharing as the driver of the relationship between order flow and exchange rate dynamics.

4.2 Hedging demand and the balance of payments

Table 2 documents the strong contemporaneous correlation between hedging demand and currency returns. These results beg the question, “What are these traders hedging?”. In this subsection, we study goods trade as a possible motivation for hedging activity. More specifically, we examine the forecasting power of hedging demand in the currency futures market over future realisations of bilateral trade balances.

Trade in goods and services is an intuitive place to begin the search for the non-tradable income streams discussed in Section 2. International trade induces currency exposures for firms because of

---

[14] Though (11) relates currency returns to interest rate differentials and risk premia, the regression equation estimated is still equivalent to (11) by the covered interest parity condition, $f_t - p_{t-1} = r^F_{t-1} - r^D_{t-1}$.

the long lags between the time when a transaction is completed and the time when payment is physically made.[15] Firms uncomfortable with the uncertainty involved in receiving a fixed payment in foreign currency can easily hedge the transaction using either futures or forward contracts.

If firms actively use currency futures to hedge international transactions in goods and services, then one would expect currency hedging demand to have forecasting power over bilateral trade balances. The intuition here is that once a transaction is initiated, firms extending standard credit terms can expect payment within one to three months. If firms begin to hedge once they become aware of the currency exposure, then hedging demands should lead actual trade balance flows by one to three months. To explore this hypothesis, we test the in-sample forecasting power of currency hedging demand on bilateral trade balances. We do this by estimating autoregressive moving average with exogenous regressor (ARMAX) models for each currency pair in our study. Using the Box-Jenkins methodology, we estimate ARMAX(1,1,1) models of the form

$$\frac{tb_t - tb_{t-1}}{tb_{t-1}} = \alpha + \rho tb_{t-1} + \beta h_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1} \tag{13}$$

where ε_t is a white noise process and tb_t is the bilateral trade balance at time t with the United States taken as the home country. We report the results in Table 4.

The results are mixed, with the only significant results coming from the trade balances with Canada and Japan, the United States' first and third largest trading partners, respectively. The coefficient on the hedging demand term is of the correct sign in that purchases of Canadian dollars forward tend to lead increases in the trade balance. The coefficient on hedging demand for the Japanese trade balance does not have the expected sign. The lack of significance in the Swiss and UK regressions is not too surprising because of the relatively small bilateral trade between those countries and the United States.[16] The mixed results for both the Canadian and Japanese trading balances could be due to the use of natural or economic hedges by firms. Given the large volumes of trade between the United States and Canada and Japan, many firms may choose to locate their operations in foreign countries[17] to denominate their cost and revenue streams in a common currency to reduce their net exposure to currency fluctuations.

The weak relationship between hedging demand and trade flows is consistent with the types of agents that typically trade in currency futures. As described in Section 2, hedgers in the currency futures markets comprise large commercial banks and medium-sized corporations. Trading activity from banks is likely to reflect conditions in the interbank market while the corporate players in the futures markets probably account for a small portion of the total volume of bilateral goods trade.

4.3 Speculators: informed "insiders" or insurance providers

The previous subsection showed that hedging demand in currency futures markets does not appear to be driven by income shocks related to goods trade. While this result is not totally surprising given the relative magnitudes of trading volume in currencies versus the amount of bilateral trade between countries, it may imply that motives other than risk-sharing may be driving the results in Table 3.

An alternative hypothesis that is consistent with the results in Tables 3 and 4 is that hedgers are in fact noise traders who trade against much better informed speculators. Under this hypothesis, hedging demand should not be related to trade flows since hedgers trade in a random fashion and hedgers should, on average, lose money to informed speculators, leading to the negative coefficients on γ in Table 3. To explore the validity of this hypothesis, we regress exchange rate returns on speculative

---

15 Currency exposures induced by trade are generally referred to as transaction exposures in the international corporate finance literature.

16 In 2000, the volume of trade between the United States and Switzerland was roughly USD 20 billion as compared to USD 80 billion traded between the United Kingdom and the United States or the USD 400 billion of trade between Canada and the United States.

17 Examples of these natural hedges include the construction of semiconductor fabrication plants in Ireland and Germany by Intel and AMD, both US firms, and the large manufacturing capacity that Japanese car manufacturer Toyota Motor Corporation has developed in North America, producing almost 20% of its output there.

and hedging demands to check that speculative demand is positively related to currency returns; these results are reported in Table 5.

These results indicate that when speculators buy a given currency, that currency appears to appreciate relative to the US dollar. This behaviour would be consistent with a Kyle (1985) setting where speculators have private information about future returns. The nature of trading in the futures pits implies that the speculators' gains come at the expense of the hedgers. The hypothesis that these results are due to an informational advantage held by speculators is somewhat suspect. First, the magnitude and stability of these returns imply that speculators have extremely good information about future returns. Second, the sustained losses by hedgers over the sample period seem too great to justify their continued existence.

The risk-sharing environment developed in Section 2, however, also predicts the observed relationship between speculators and hedgers. The intuition here is that hedgers "pay" speculators a premium for bearing risks that they do not wish to hold. Thus, under this interpretation one can view the losses of the hedgers as an insurance premium. The key difference between the information and risk-sharing scenarios is the causality between hedging and speculative demands. In the Kyle setting, speculators enter the market and induce hedgers to take the other side of their trades, while, in the risk-sharing model, hedgers are the initiators of trade.

To differentiate between these competing models, we run Granger causality tests to identify the causal relationship between innovations in hedging and speculative flows at the weekly level; the results are reported in Table 6.[18] The results are quite striking: in all currencies except the Canadian dollar, innovations in hedging demand Granger cause changes in speculative demand; even for the Canadian dollar, the results point towards hedging demand Granger causing speculative demand, but the results are not significant. Though Granger causality is at best a rough measure of causality, the results are fairly clear in that none of the tests indicates reverse causality. The consistency of the Granger causality tests lends strong support to the risk-sharing interpretation of the results. The findings make intuitive sense in that it is hard to believe that speculators could sustain an informational advantage over such a long period *while at the same time* hedgers continued to accumulate losses.

4.4 Hedging and feedback trading

The previous subsection showed that hedgers appear to be driving the trading dynamics in the futures market, lending support to the theory developed in Section 2. Another alternative model that could be driving the results is that hedgers are simply irrational feedback traders. The literature has typically focused on positive feedback, or momentum, trading as an irrational trading strategy. Many authors have shown the fragility of financial markets when positive feedback traders are present. Here, we study the nature of trading by hedgers that are following some type of positive feedback strategy.

Table 7 documents the relationship between hedging demand and lagged currency returns. These results suggest that hedgers tend to act as negative feedback traders, ie hedgers tend to purchase a currency after it has depreciated. Negative feedback trading is much more difficult to justify using behavioural arguments, as it requires traders to buy after prices go down. This finding, coupled with the results from Table 8, sheds interesting new light on previous studies that documented positive feedback trading in futures markets; see Kodres (1994).

Our results suggest that destabilising speculation of the sort described in de Long et al (1990) is unlikely. In their model, rational speculators may bid up the price of a security, inducing noise traders who use positive feedback strategies to enter the market, subsequently selling out at a higher price. While some subset of traders classified as speculators may indeed fit the description of a positive feedback noise trader, the presence of hedgers who are on average negative feedback traders should drastically reduce the net susceptibility of the market to rational destabilisation.

[18] The results presented use two weekly lags; to measure changes in demand, we simply use the absolute net change in position for each class of trader. The results are essentially unchanged when one includes lags from one to four weeks.

4.5 Futures hedging and customer-dealer order flow

Recent research on the microstructure of the foreign exchange market indicates that aggregate foreign exchange order flow is significantly related to currency returns; see Rime (2000). In this subsection, we test to see how futures hedging demands are related to a data set containing customer-dealer order flow.

Table 8 documents the relationship between customer order flow normalised by USD 100 million, Δx_t^c, and currency returns at the weekly level. Note that we do not test the Deutsche mark here because it effectively stopped trading half way through our sample. The almost complete lack of explanatory power is surprising given prior research that has generally associated order flow with returns quite strongly. The large market share and diverse customer base of the bank we study go some way to explaining these results. Given that foreign exchange dealers are extremely reluctant to hold positions overnight, net daily customer order flow in the aggregate should fluctuate randomly around zero.

This set of results indicates that the effects we observe in previous tables are not due to a generic order flow effect. These results also have important implications for future research. It appears that researchers would be well served to study specific components of customer order flow to identify structural relationships in the market. Intuitively, the potential for informational gains from disaggregating order flows is similar to the benefits from studying cointegrating relationships versus simply differencing a non-stationary time series.

5. Conclusion

This paper developed a model of unobservable risk premia in a stylised foreign exchange market based on the need of some agents to hedge non-marketable income flows. Using data on hedging demand in the currency futures market, we tested the implications of the model and found broad support for it. We tested our results against the specific alternative that the observed results were due to information-based trading rather than risk-sharing. Consistent with our theory, we found that hedgers tended to lose money at the expense of speculators and changes in hedging demands Granger cause changes in speculative demand. We also ruled out the possibility that the influence of hedgers is driven by some type of naive positive feedback strategies. Lastly, we compared the explanatory power of futures hedging demand over currency returns to that of customer-dealer order flow from a major international bank. We found that our customer order flow data had little or no explanatory power over exchange rate returns over weekly horizons.

The consistency of the empirical findings with our theoretical predictions suggests that risk premia are present and identifiable in the foreign exchange market. Equivalently, the results suggest that -sharing can explain a significant proportion of the variation in exchange rates. Our findings intuitively show that the foreign exchange market is an efficient mechanism for allocating risk across the economy. The type of information which is privately held appears to be information related to risk premia and not future payoffs. This finding is consistent with previous evidence of asymmetric information in currency markets as well as the enormous depth and liquidity of the major currency markets. Traders are more willing to transact because they are less likely to be trading against someone with superior information.

While our theoretical model is straightforward, the result that hedging demand is closely related to risk premia is quite general. Unfortunately, this generality precludes a straightforward explanation of what drives the risk premium, but provides a fruitful area for future research. The composition of the large players in the futures markets and the lack of a relationship between hedging demand and trade balances suggest that the effects we observe reflect conditions in the interbank market. In future research, we plan to explore the process whereby risk-sharing among dealers and other speculative traders can drive short-term currency dynamics while macroeconomic forces enforce long-term cycles in exchange rates.

Our results also have practical implications. First, the observation that futures hedging demand is priced in the aggregate foreign exchange market implies that currency trading provides risk reduction benefits to a non-trivial group of agents. This suggests that the imposition of transaction costs to reduce speculation, at least in developed markets, could have significant welfare costs. Second, the lack of explanatory power of our aggregate customer order flow data set suggests that future research should focus on components of order flow which have an economic relation to variables of interest.

Tables

Table 1 shows some summary information on the specification of the currency futures contracts used in this study. The average daily volume and bank participation statistics reported below were calculated using data from January to December 2000 as this was the longest span over which these data were publicly available.

The next table provides summary statistics for our hedging demand proxy, h_t, by currency. ACF(i) corresponds to the *i*th term of the series' autocorrelation function and PACF(1) refers to the value of the first term of the series' partial autocorrelation function.

Table 1

Summary statistics for currency futures contracts

Currency futures contract specifications and summary information

	Canadian dollar	**Swiss franc**	**Deutsche mark**	**Pound sterling**	**Japanese yen**
Contract size	CAD 100,000	CHF 125,000	DEM 125,000	GBP 62,500	JPY 12.5 m
Delivery months	3, 6, 9, 12	3, 6, 9, 12	3, 6, 9, 12	3, 6, 9, 12	3, 6, 9, 12
Avg open interest	42,248	45,412	79,109	39,849	74,736
Avg daily volume	9,672	12,862	649	8,054	15,736
Bank participation	29.7%	40.4%	NR	16.0%	32.7%

Statistical properties of hedging demand proxy, h_t, by currency

	Canadian dollar	**Swiss franc**	**Deutsche mark**	**Pound sterling**	**Japanese yen**
Mean	– 0.14	0.06	0.01	– 0.01	0.09
Std deviation	0.41	0.45	0.31	0.43	0.39
Median	– 0.15	0.10	0.02	0.02	0.12
Minimum	– 1.00	– 0.84	– 0.65	– 0.89	– 0.92
Maximum	0.73	0.88	0.82	0.90	0.72
ACF(1)	0.56	0.44	0.45	0.34	0.56
ACF(2)	0.33	0.16	0.25	0.10	0.33
PACF(2)	0.02	– 0.04	0.07	– 0.01	0.02

Table 2 shows the relationship between currency returns and the expected returns in the currency forward market, commonly referred to as the uncovered interest parity relation. The statistics below are from the regression

$$p_t - p_{t-1} = \alpha + \beta(f_{t-1} - p_{t-1}) + \varepsilon_t$$

where p_t is the natural logarithm of the US dollar price of one unit of foreign currency at time t and f_t is the natural logarithm of the one-month forward price in US dollars of one unit of foreign currency. We use monthly data from January 1986 to December 2000 and estimate the system of equations together using generalised least squares (GLS). Standard errors are reported in parenthesis.

Table 2

Uncovered interest parity without hedging demand

Currency	α	β	Adj R^2	D-W
Canadian dollar	– 0.0003 (0.0010)	0.1167 (0.4274)	0.00	2.10
Swiss franc	0.0003 (0.0025)	– 0.5746 (0.4482)	0.01	1.77
Deutsche mark	– 0.0004 (0.0024)	– 0.0020 (0.3646)	0.01	1.91
Pound sterling	– 0.0002 (0.0023)	0.5602 (0.5287)	0.00	1.81
Japanese yen	0.0031 (0.0032)	– 0.5013 (0.6577)	0.00	1.76

Table 3 shows the impact of adding the hedging demand proxy, h_t, to the regression of currency returns on the expected return in the currency forward market. Formally, we run the regression

$$p_t - p_{t-1} = \alpha + \beta(f_{t-1} - p_{t-1}) + \gamma h_t + \varepsilon_t$$

where p_t is the natural logarithm of the US dollar price of one unit of foreign currency at time t, f_t is the natural logarithm of the one-month forward price in US dollars of one unit of foreign currency at time t, and h_t is the relative net hedging demand proxy for that currency. We use monthly data from January 1986 to December 2000 and estimate the system of equations together using generalised least squares (GLS). Standard errors are reported in parenthesis.

Table 3

Uncovered interest parity with hedging demand

Currency	α	β	γ	Price impact for 10,000 contracts	Adj R^2	D-W
Canadian dollar	– 0.0032 (0.0009)	0.0482 (0.3328)	– 0.0208 (0.0019)	– 34 bp	0.39	2.33
Swiss franc	0.0045 (0.0019)	– 0.1414 (0.3583)	– 0.0406 (0.0028)	– 54 bp	0.47	2.02
Deutsche mark	0.0035 (0.0018)	– 0.2215 (0.2934)	– 0.0537 (0.0046)	– 32 bp	0.47	2.17
Pound sterling	0.0008 (0.5718)	0.5718 (0.4085)	– 0.0389 (0.0029)	– 44 bp	0.48	1.98
Japanese yen	0.0070 (0.0025)	0.3273 (0.5356)	– 0.0533 (0.0045)	– 34 bp	0.38	1.88

Table 4 shows the effectiveness of the hedging demand proxy, h_t, in forecasting the bilateral trade balance between the United States and each country in our study. The results below are from the ARMAX(1,1,1) model

$$\frac{tb_t - tb_{t-1}}{tb_{t-1}} = \alpha + \rho tb_{t-1} + \beta h_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1}$$

where tb_t is the bilateral trade balance between the United States and the foreign country reported in US dollars. The model is fitted on monthly data from January 1986 to December 2000. Standard errors are reported in parenthesis.

Table 4

Bilateral trade balances and hedging demand

Currency	α	ρ	β	θ	Adj R^2	D-W
Canadian dollar	0.1079 (0.0195)	0.6401 (0.1561)	0.1727 (0.0801)	– 0.8216 (0.1183)	0.04	2.09
Swiss franc	– 2.7528 (2.2908)	0.7900 (1.2612)	3.8625 (4.5578)	– 0.7783 (1.2916)	– 0.02	2.02
Deutsche mark	– 0.1113 (0.2104)	– 0.2740 (1.6258)	– 0.5116 (0.6763)	0.3143 (1.6049)	– 0.01	2.01
Pound sterling	– 1.1399 (1.5193)	– 0.5394 (1.2986)	– 3.3372 (3.5313)	0.5316 (1.3310)	– 0.01	1.97
Japanese yen	0.0205 (0.0042)	0.1204 (0.1229)	– 0.0327 (0.0154)	– 0.7236 (0.0852)	0.26	1.97

Table 5 shows the relationship between currency returns and relative net hedging and speculative demands, $h_{i,t}$ and $x_{i,t}$, respectively. The table contains the results from the following regressions:

$$p_t - p_{t-1} = \alpha_h + \gamma_h h_t + \varepsilon_t$$

$$p_t - p_{t-1} = \alpha_s + \gamma_s x_t + \varepsilon_t$$

where $p_{i,t}$ is the natural logarithm of the US dollar price of one unit of foreign currency *i* at time *t*. Note that a positive γ_s coefficient implies that as speculators increase their net positions in the currency futures contract, the currency *appreciates* versus the US dollar. The regressions are run on monthly data from January 1986 to December 2000. Standard errors are reported in parenthesis.

Table 5

Profitability of hedgers versus speculators

Currency	α_h	α_s	γ_h	γ_s
Canadian dollar	– 0.0033 (0.0008)	– 0.0020 (0.0008)	– 0.0210 (0.0019)	0.0134 (0.0014)
Swiss franc	0.0052 (0.0019)	0.0061 (0.0021)	– 0.0578 (0.0042)	0.0371 (0.0034)
Deutsche mark	0.0035 (0.0018)	0.0047 (0.0020)	– 0.0872 (0.0062)	0.0500 (0.0043)
Pound sterling	0.0005 (0.0017)	0.0001 (0.0018)	– 0.0503 (0.0034)	0.0286 (0.0026)
Japanese yen	0.0082 (0.0021)	0.0100 (0.0023)	– 0.0575 (0.0055)	0.0372 (0.0039)

Table 6 outlines the results of Granger causality tests on the relationship between the absolute net hedging and speculative demand measures, h_t^a and x_t^a, respectively. The test is done on weekly data from 13 October 1992 to 26 December 2000 using two weekly lags. The test consists of running the bivariate regressions

$$h_t^a = \alpha_1 + \beta_1 h_{t-1}^a + \beta_2 h_{t-2}^a + \gamma_1 x_{t-1}^a + \gamma_2 x_{t-2}^a + \varepsilon_t$$

$$x_t^a = \alpha_2 + \theta_1 x_{t-1}^a + \theta_2 x_{t-2}^a + \lambda_1 h_{t-1}^a + \lambda_2 h_{t-2}^a + \eta_t$$

The hypothesis that "h_t^a does not Granger cause x_t^a" corresponds to a test of the hypothesis $\lambda_1 = \lambda_2 = 0$. If we cannot reject the hypothesis that x_t^a does not Granger cause h_t^a but do reject the hypothesis that h_t^a does not Granger cause x_t^a in the second regression, then we say that Granger causality runs one way from hedging demand to speculative demand.

Table 6

Granger causality test of weekly hedging versus speculative flows

Currency	**Hypothesis**	**F-statistic**	**p-value**
Canadian dollar	h_t^a does not Granger cause x_t^a	1.879	0.154
	x_t^a does not Granger cause h_t^a	0.641	0.527
Swiss franc	h_t^a does not Granger cause x_t^a	8.070	0.000
	x_t^a does not Granger cause h_t^a	2.526	0.081
Deutsche mark	h_t^a does not Granger cause x_t^a	4.480	0.012
	x_t^a does not Granger cause h_t^a	1.954	0.144
Pound sterling	h_t^a does not Granger cause x_t^a	3.088	0.047
	x_t^a does not Granger cause h_t^a	2.028	0.133
Japanese yen	h_t^a does not Granger cause x_t^a	5.927	0.003
	x_t^a does not Granger cause h_t^a	0.676	0.509

Table 7 shows the dependence of the hedging demand proxy, h_t, on past currency returns. Specifically, the results below are from the regression

$$h_t = \alpha + \beta(p_{t-1} - p_{t-2}) + \varepsilon_t$$

using monthly data from January 1986 to December 2000. p_t is the US dollar price of a unit of foreign currency at time *t*. Standard errors are reported in parenthesis.

Table 7

Hedging demand and past currency returns

Currency	α	β	Adj R^2	D-W
Canadian dollar	– 0.1470 (0.0279)	– 11.3914 (2.0545)	0.15	1.43
Swiss franc	0.0939 (0.0289)	– 5.9398 (0.7173)	0.14	1.82
Deutsche mark	0.0451 (0.0206)	– 3.5058 (0.5644)	0.07	1.96
Pound sterling	– 0.0032 (0.0302)	– 5.5366 (0.8990)	0.11	1.91
Japanese yen	0.1192 (0.0258)	– 4.7413 (0.6685)	0.13	1.32

The table below shows the results of the regression

$$p_t - p_{t-1} = \alpha + \beta x_t^c + \varepsilon_t$$

using weekly data over the period from January 1998 to March 2000. p_t is the natural logarithm of the US dollar price of a unit of foreign currency for currency i at the end of week t and $x_{i,t}^c$ is the net customer order flow in a particular currency over week t. Standard errors are reported in parenthesis.

Table 8

Customer order flow and currency returns

Currency	α	β	Adj R^2	D-W
Canadian dollar	– 0.0002 (0.0008)	– 0.0003 (0.0009)	0.00	2.38
Swiss franc	– 0.0017 (0.0014)	– 0.0012 (0.0010)	0.01	2.12
Pound sterling	0.0002 (0.0009)	0.0004 (0.0004)	0.01	2.20
Japanese yen	0.0011 (0.0020)	0.0005 (0.0004)	0.02	1.81

Figure 1

Plot of spot Japanese yen/US dollar exchange rate
January 1986 to December 2000

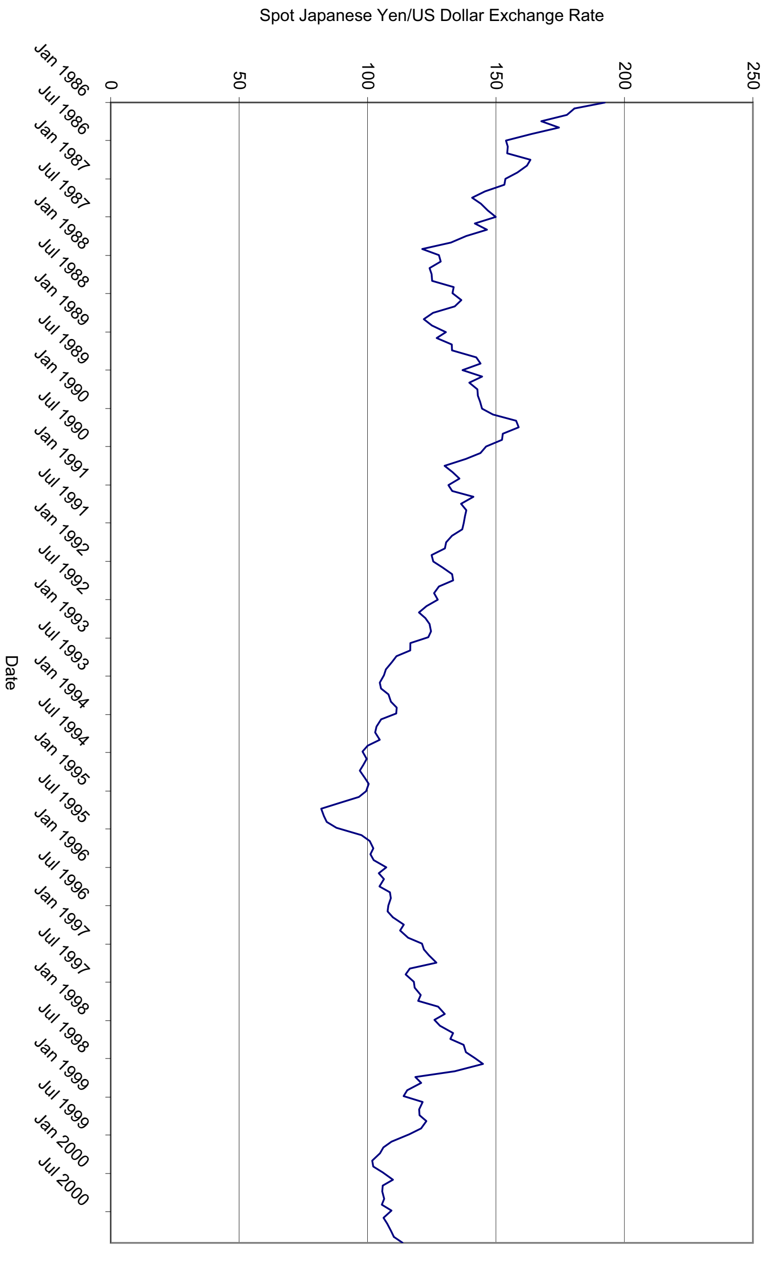
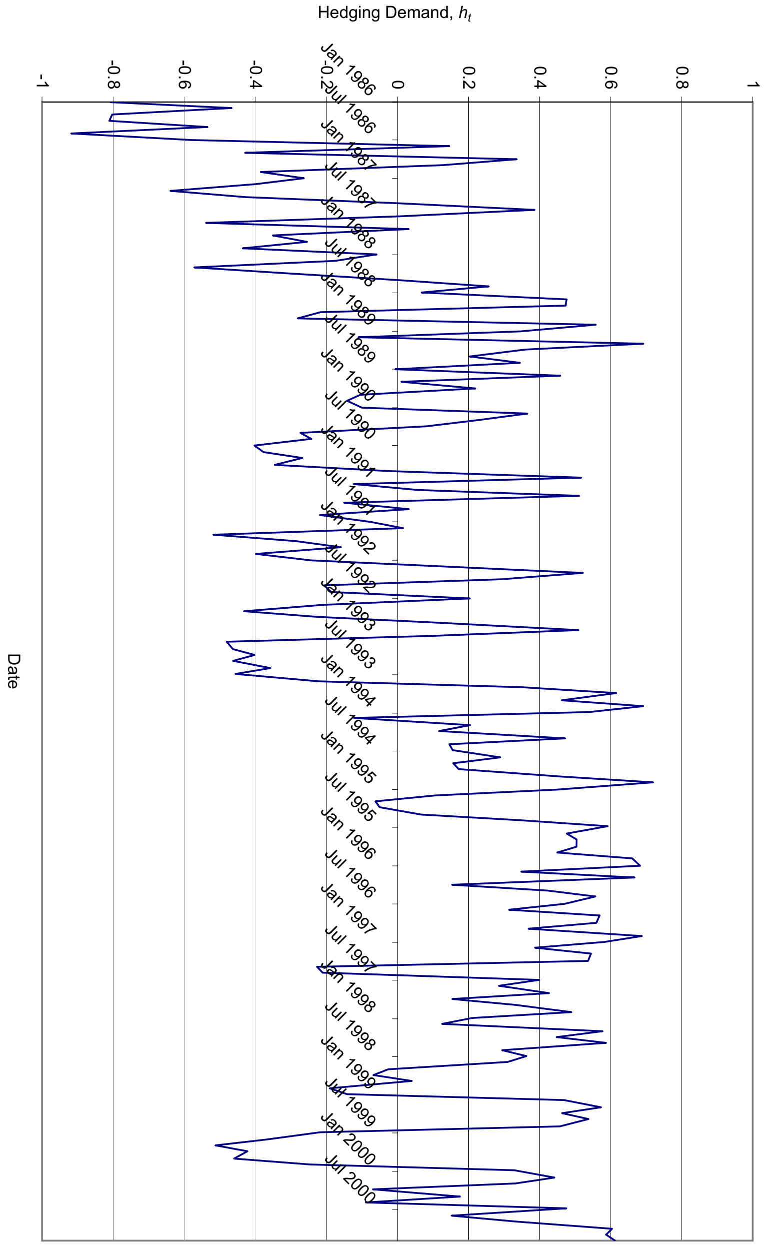


Figure 2

Plot of hedging demand, h_t, in Japanese yen January 1986 to December 2000



References

Baillie, Richard T and Tim Bollerslev (1994): "The long memory of the forward premium", *Journal of International Money and Finance* 13, 565-71.

Bessembinder, Hendrik (1992): "Systematic risk, hedging pressure, and risk premiums in futures markets", *Review of Financial Studies* 5, 637-67.

Branson, William H and Dale W Henderson (1985): "The specification and influence of asset markets", in Ronald W Jones and Peter B Kenen (eds), *Handbook of International Economics* (Elsevier, New York).

Breeden, Douglas T (1979): "An intertemporal asset pricing model with stochastic consumptions and investment opportunities", *Journal of Financial Economics* 7, 265-96.

de Long, J Bradford, Andrei Shleifer, Lawrence H Summers and Robert J Waldmann (1990): "Positive feedback investment strategies and destabilising rational speculation", *Journal of Finance* 45, 379-95.

de Roon, Frans A, Theo E Nijman and Chris Veld (2000): "Hedging pressure effects in futures markets", *Journal of Finance* 55, 1437-56.

Domowitz, Ian and Craig S Hakkio (1985): "Conditional variance and the risk premium in the foreign exchange market", *Journal of International Economics* 19, 47-66.

Eichengreen, Barry, James Tobin and Charles Wyplosz (1995): "Two cases for sand in the wheels of international finance", *Economic Journal* 105, 162-72.

Engel, Charles (1996): "The forward discount anomaly and the risk premium: a survey of the recent evidence", *Journal of Empirical Finance* 3, 123-92.

Engel, Charles M and James D Hamilton (1990): "Long swings in the dollar: are they in the data and do the markets know it?", *American Economic Review* 80, 689-713.

Evans, Martin D D (1996): "Peso problems: their theoretical and empirical implications", in G S Maddala and C R Rao (eds), *Statistical Methods in Finance* (North-Holland, Amsterdam).

Evans, Martin D D and Richard K Lyons (2001): "Portfolio balance, price impact, and secret interventions", NBER Working Paper No 8356.

——— (2002): "Order flow and exchange rate dynamics", *Journal of Political Economy* 110, 170-80.

Flood, Robert P and Andrew K Rose (1995): "Fixing exchange rates: a virtual quest for fundamentals", *Journal of Monetary Economics* 36, 3-37.

Flood, Robert P and Mark P Taylor (1996): "Exchange rate economics: what's wrong with the conventional macro approach?", in Jeffrey A Frankel, Giampaolo Galli and Alberto Giovannini (eds), *The Microstructure of Foreign Exchange Markets* (University of Chicago Press, Chicago).

Frankel, Jeffrey A and Kenneth A Froot (1989): "Forward discount bias: is it an exchange risk premium?", *Quarterly Journal of Economics* 104, 139-61.

Froot, Kenneth A and Richard H Thaler (1990): "Anomalies: foreign exchange", *Journal of Economic Perspectives* 4, 179-92.

Hayek, Friedrich August (1945): "The use of knowledge in society", *American Economic Review* 35, 519-30.

Hirshleifer, David (1990): "Hedging pressure and futures price movements in a general equilibrium model", *Econometrica* 55, 411-28.

Kodres, Laura E (1994): "The existence and impact of destabilising positive feedback traders: evidence from the S&P 500 index futures market", Working Paper 94-9, Board of Governors of the Federal Reserve System Finance and Economics Discussion Series.

Kodres, Laura E and Matthew Pritsker (1995): "Directionally similar position taking and herding by large futures market participants", *Risk Measurement and Systemic Risk: Proceedings of a Joint Central Bank Research Conference,* 221-72.

Kyle, Albert S (1985): "Continuous auctions and insider trading", *Econometrica* 53, 1315-35.

Lyons, Richard K (2001): “New perspective on FX markets: order-flow analysis”, *International Finance* 4.

Mark, Nelson C and Yangru Wu (1998): “Rethinking deviations from uncovered interest parity: the role of covariance risk and noise”, *Economic Journal* 108, 1686-1706.

Meese, Richard A and Kenneth Rogoff (1983): “Empirical exchange rate models of the seventies: do they fit out of sample?”, *Journal of International Economics* 14, 3-24.

Obstfeld, Maurice and Kenneth Rogoff (2000): “The six major puzzles in international macroeconomics: is there a common cause?”, in Ben S Bernanke and Kenneth Rogoff (eds), *NBER Macroeconomics Annual* (MIT Press, Cambridge).

Rime, Dagfinn (2000): “Customer trading and information in foreign exchange markets”, Working Paper, Memorandum 30/2000, Department of Economics, University of Oslo.

Roberts, Mark A (1995): “Imperfect information: some implications for modelling the exchange rate”, *Journal of International Economics* 38, 375-83.

Rubinstein, Mark (1976): “The valuation of uncertain income streams and the pricing of options”, *Bell Journal of Economics* 7, 407-25.

Wang, Jiang (1994): “A model of competitive stock trading volume”, *Journal of Political Economy* 102, 127-68.

Session 4

Liquidity II

Measuring and explaining liquidity on an electronic limit order book: evidence from Reuters D2000-2[1]

Jón Daníelsson and Richard Payne,
London School of Economics

Abstract

The conference presentation focused on recent results on dynamic trading patterns in limit order markets, primarily foreign exchange and money markets. Clear feedbacks are observed between liquidity, volatility and volume. These results suggest that any regulatory regime for market liquidity should appreciate these feedback rules, and treat liquidity risk as endogenously determined, rather than an exogenous process.

1. Introduction

Liquidity risk has emerged as one of the most significant risk factors in the global financial economy, being a significant contributor to several financial crises such as the 1987 stock market crash and the Russia crisis of 1998. In spite of the importance of liquidity for financial stability, academic understanding of liquidity is very limited. On a general level, liquidity facilitates trading, where a liquid market is one in which participants can trade desired amounts quickly, cheaply and without greatly affecting prices.

The objective of this presentation is to discuss how methodologies developed in the field of market microstructure can aid in understanding liquidity in a particular trading venue or market. The task of studying liquidity within this context is complicated by the fact that no single definition of liquidity exists. However, Kyle's (1985) three component classification of liquidity, covering tightness, depth and resilience, is well known, and serves as a useful starting point. Unfortunately, not only do most extant empirical studies of liquidity fail to fully explore Kyle's notions,[2] we feel that his concept of liquidity is limited in the sense that it only reflects a static picture of market conditions, and not the dynamic environment of modern financial markets. This is especially important in the study of financial stability where it is necessary to explicitly consider the evolution of liquidity over time, and the interdependence of liquidity with other market variables, eg prices. Given the importance of liquidity, any threat to liquidity supply has the potential for adverse economic implications.

Daníelsson and Payne (2002a) analyse the dynamics of liquidity using one week of transaction data for the USD/DEM spot rate on the Reuters D2000-2 system. The properties of this data set are extensively documented in Daníelsson and Payne (2002b).[3] Since the data are unusually detailed, containing information on all D2000-2 orders whether or not they were traded, while market participants only see a subset of the data, it is possible to analyse market dynamics which are beyond

[1] Corresponding author Jón Daníelsson, j.danielsson@lse.ac.uk. Our papers can be downloaded from www.riskresearch.org. The authors thank Reuters Group PLC for providing some of the data. All errors are our own responsibility.

[2] Most empirical studies focus solely on tightness, ie spreads. There are many reasons for this. First, the inventory control and asymmetric information literature developed in the 1970s and 1980s gives clear predictions regarding the determination of bid-ask spreads; see eg Ho and Stoll (1983), Glosten and Milgrom (1985) and Easley and O'Hara (1987). Second, estimators of spread components were successfully developed based upon these theories; see eg Roll (1984), Stoll (1989) and Huang and Stoll (1997). Last, most microstructure databases contain little/no liquidity information outside the spread.

[3] Given the short temporal span of the data, the analysis is limited in the types of empirical analysis that can be conducted. For example, macro-level analysis of exchange rate determination is clearly not possible.

the scope of most other market microstructure studies, eg high-frequency order placement decisions. The study by Daníelsson and Payne (2002b) casts light on the strategic trading behaviour of market participants, and documents the resulting trading patterns. On a theoretical level, it argues that most of the observed results are consistent with asymmetric information theories.

Daníelsson and Saltoglu (2002) take advantage of the insights of Daníelsson and Payne (2002a) in their analysis of the recent Turkish financial crises, and find that market microstructure liquidity patterns played a key role in the evolution of the crises.

The key objective of the papers discussed above is the analysis of various aspects of liquidity. First, the determination of conditions when liquidity is supplied or demanded. Second, the impact of trading strategies on liquidity supply/demand. Third, to what extent changes in liquidity supply/demand and trading strategies help predict market crashes. Finally, what is the dynamic relation between liquidity, volatility, volume and financial crises.

2. Data and models

In recent years, electronic brokers have become increasingly important in inter-dealer FX trading. The data set used by Daníelsson and Payne (2002a) (DP) consists of one week of trading in the USD/DEM spot rate on the Reuters D2000-2 electronic broking system. The D2000-2 is one of the two main electronic brokers in the market, the other being EBS.

D2002 operates as a pure limit order market governed by rules of price and time priority. A D2002-2 screen displays to users the best limit buy and sell prices as well as quantities available at those prices and a record of recent transaction activity for up to six currency pairs. It is important to note that, unlike many other limit order markets, information about limit orders away from the best prices is not available to users, ie the order book is *closed*. In addition, orders are not allowed to “walk up the book”. The data set used by DP contains all orders entered into the system, both limit orders and market orders, making it possible to construct the entire order book in real time. This enables DP to analyse the role of information and how traders form expectations and react to unexpected events in this type of limit order markets.

An example of these order books is given in Figures 1 and 2.

Figure 1

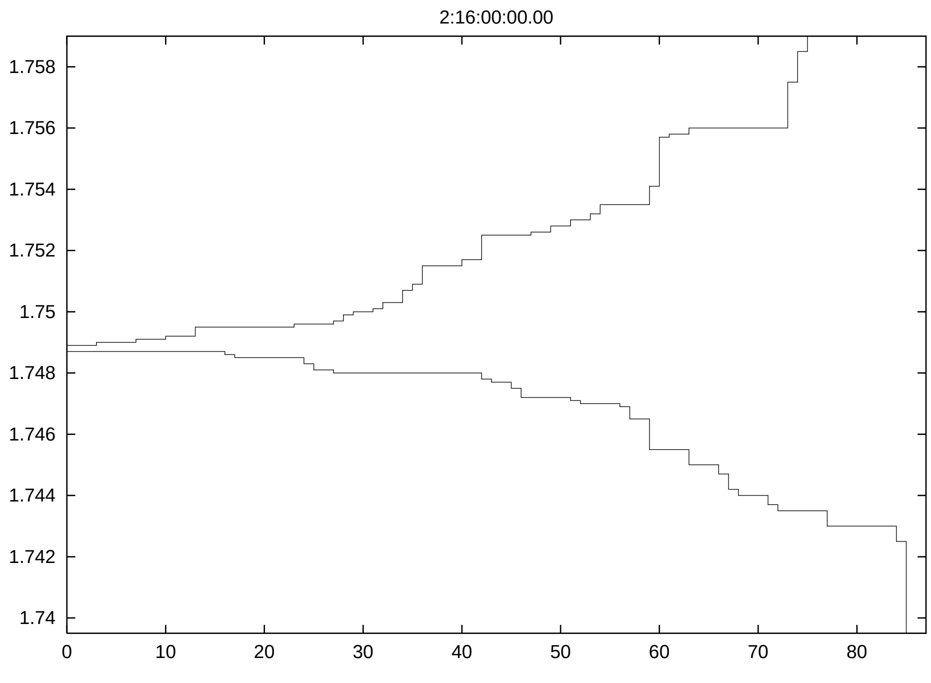


Figure 2

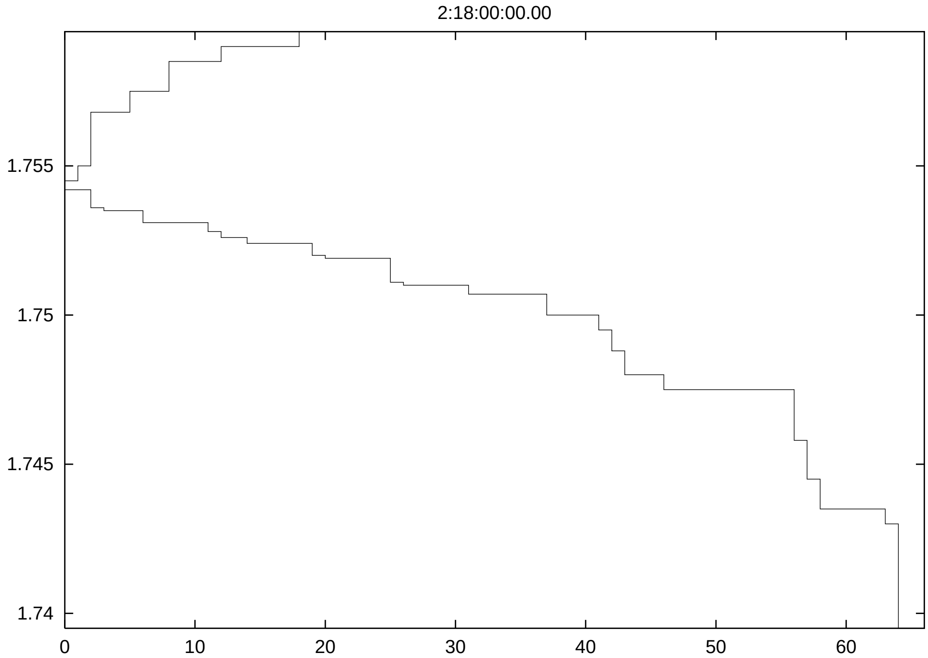


Figure 1 shows the order book at 4 pm on the second day of the sample; the best ask price is 1.749 DEM/USD with a spread of one pip (1/100 pfennig). There is about USD 80 million in the book on both the bid and ask side of the book, where the book is more or less symmetric. An interesting observation is a small amount in the book, given the overall volume in the FX markets. Indeed, on average USD 80 million enters the book each minute during peak trading times, and 80 million exits, via either trading or cancellations. This indicates that much volume sits outside the order book, ready to enter at a moment's notice. This is a key reason why DP suggest that it is important to consider the dynamic aspects of liquidity, both the dynamics of how the order book changes shape and the flow in and out of the book. The change in the order book shape is apparent in Figure 2, which shows the market two hours later. At this time the spreads are wider, and the order book contains less money, 20 million on the ask side, and 65 on the bid side. This is primarily because 6 pm is late in the day, and the trading day is beginning to wind up.

Since D2000-2 is only one of the two electronic brokers operating in the inter-dealer market, and we observe neither direct inter-dealer trading nor customer-dealer activity, we are not able to provide a picture of overall FX market activity. However, since the data set is unusually rich, DP are able to analyse the codetermination of liquidity, volatility and transaction activity in a given trading venue and the richness of the data set opens the possibility of studying high-frequency order placement decisions, something not possible with most other market microstructure data sets. They employ a variety of both event and calendar time techniques. For example, they study dynamic order placement patterns in event time by looking at both multiperiod transition matrices as well as the location of new limit orders in the order book. In calendar time, they consider vector autoregressions (VAR) where order entries, volatility and traded volume are all included, explicitly taking into account trader expectations and reactions to unexpected events.

Daníelsson and Saltoglu (2002) apply the methodology and insights from this study to analyse financial crises. The data set they use consists of all transactions on the Turkish overnight repo money market from January 2000 to March 2001. This sample includes two major financial crises. The Turkish money market is also an electronic limit order market just like the Reuters D2000-2 market. They find that interest rates are significantly correlated with order flow, spreads, realised volatility and trading imbalances. Furthermore, the interrelationship between those key variables changes fundamentally around crisis periods.

3. Empirical results

The results from Daníelsson and Payne (2002a) provide new insights into the interplay between liquidity, volatility and market activity. Taken in isolation, liquidity supply is found to be self-regulating, ie low extant liquidity leads to higher liquidity supply in the future, and conversely, abnormally high liquidity tends to be reduced in the future. Furthermore, liquidity supply temporally clusters on one side of the market and removal of liquidity at the front of one side of the book implies increased probability of seeing fresh liquidity at the front of the book and lower chances of seeing subsidiary liquidity supply on that side of the book.[4] These effects are time persistent.

However, by jointly analysing liquidity supply, volatility and volume, a different picture emerges. Liquidity, volatility and volume are interrelated, with strong feedbacks between those variables.

When focusing on order submission strategies, in times of uncertainty the relative number of limit orders vs market orders increases. While this might seem to imply that liquidity increases when markets are uncertain, this liquidity supply is poorly priced, thus spreads are high and depth low. Hence, we observe a positive relationship between risk and the price of liquidity. These results are reinforced by calendar time analysis using vector autoregressions. By focusing on volatility in particular, we find that when observing episodes of high volatility, liquidity is low, and conversely when volatility is low liquidity is high. Furthermore, these patterns are self-reinforcing. Similar evidence emerges from the study by Daníelsson and Saltoglu (2002) of the Turkish financial crises, which were characterised by extreme movements in interest rates. They run a similar vector autoregression to Daníelsson and Payne (2002a), but with daily data. They find that there are significant positive feedbacks between realised volatility, liquidity and interest rate changes - exactly the same observations as were found on foreign exchange markets. Furthermore, they find that this interdependence becomes more strongly significant prior to and during crisis periods.

4. Interpretation and analysis

A key result from the previous section is the presence of feedbacks between key variables. The theoretical environment that may generate such outcomes is of some interest. There are at least two possible theoretic explanations. The first main area of microstructure research focuses on dealer inventory management issues (Amihud and Mendelson (1985), Stoll (1989) and Huang and Stoll (1997)). Lyons (1995) demonstrates that such inventory control is a very important part of FX dealer behaviour. However, we do not believe that this strand of theory can help us explain the patterns we see in the data. Rather, we appeal to the second main area of microstructure theory - asymmetric information theory.

In response to potentially informed trades, we observe that transaction activity increases subsequent volatility while reducing the liquidity, both spreads and depth. This happens because limit orders are repriced and the order book thins out as liquidity suppliers guard against being picked off by traders with superior information. Furthermore, market buy activity causes a decrease in the limit sell side depth and an increase in the limit buy side depth. This strengthens our belief that trades are providing information on the likely future direction of market prices. In a market with both informed and noise traders, we would expect an increase in the information asymmetry to widen spreads and reduce depth. A very high degree of information symmetry can easily drive extreme spreads, liquidity and volatility.

---

[4] By subsidiary liquidity supply we mean submission of limit orders at prices inferior to the extant best limit price.

5. Conclusion

This presentation focused on the dynamic evolution of limit order markets, in particular foreign exchange markets and emerging market interest rate markets in crisis. It is shown that clear dynamic patterns exist where key variables are jointly determined and, more importantly, jointly affect each other.

The analysis discussed above opens as many questions as it answers. The fact that the dynamic dimension of liquidity and information play such an important role in the market suggests that considerable research remains to be done before we can fully understand limit order markets. In addition, the fact that established market microstructure patterns seem to break down in crisis suggests that relying on analysis made in *normal* market conditions as a guide to how financial markets behave in crisis would seem to be misguided.

From the point of view of economic policy, we feel that these results demonstrate that market variables are determined in a dynamic environment and all are interdependent. This implies that any regulatory environment needs to consider how regulations may affect the dynamic structure of the market. Furthermore, an in-depth understanding of the market microstructure of financial markets can be invaluable to policymakers interested in financial stability and containment of financial crisis.

References

Amihud, Y and H Mendelson (1985): "Dealership markets: market making with inventory", *Journal of Financial Economics*, 8, 31-53.

Daníelsson, J and R Payne (2002a): "Liquidity determination in an order driven market", *London School of Economics*, www.RiskResearch.org.

——— (2002b): "Real trading patterns and prices in spot foreign exchange markets", *Journal of International Money and Finance*.

Daníelsson, J and B Saltoglu (2002): "Anatomy of a crisis: market microstructure analysis of the Turkish financial crisis", *London School of Economics*, www.RiskResearch.org.

Easley, D and M O'Hara (1987): "Price, trade size and information in securities markets", *Journal Financial Economics*, 19, 69-90.

Glosten, L and P Milgrom (1985): "Bid, ask and transaction prices in a specialist market with heterogeneously informed traders", *Journal Financial Economics*, 14, 71-100.

Ho, T and H Stoll (1983): "The dynamics of dealer markets under competition", *Journal of Finance*, 38 (4), 1053-74.

Huang, R and H Stoll (1997): "The components of the bid-ask spread: a general approach", *Review of Financial Studies*, 10, 995-34.

Kyle, A (1985): "Continuous auctions and insider trading", *Econometrica*, 53 (6), 1315-35.

Lyons, R (1995): "Tests of microstructural hypotheses in the foreign exchange market", *Journal Financial Economics*, 39, 321-51.

Roll, R (1984): "A simple implicit measure of the effective bid-ask spread in an efficient market", *Journal of Finance*, 39, 1127-39.

Stoll, H (1989): "Inferring the components of the bid/ask spread: theory and empirical tests", *Journal of Finance*, 44, 115-34.

The impact of market liquidity in times of stress on corporate bond issuance

Paul Harrison[1]

Abstract

This paper investigates the impact of liquidity shocks on the composition of firms that enter the corporate bond market. When liquidity is at a premium, larger bonds by better known firms are much more prominent, which squeezes smaller issuers and the high-yield market, in particular. This paper shows that bond size is a liquidity factor, at least for some corporate debt, and that both pricing and issuance are impacted by market liquidity.

1. Introduction and motivation

In the wake of the Russian default and the Long-Term Capital Management crisis in 1998, the corporate bond market was plagued by a lack of liquidity. Trading dried up, price quotes were reportedly difficult to come by, and positions could not be liquidated either to stem losses or to meet cash demands (see, for instance, BIS (1999) or Wall Street Journal (1998a,b)). This liquidity shock had a significant and persistent impact on the corporate bond market and on the ability of firms to raise funds in that market.

Faced with an illiquid market in the autumn of 1998, bond issuance fell dramatically from a May peak of over 150 bonds per month to less than 40 per month in September and October (Exhibit 1). While issuance bounced back following the Federal Reserve's emergency October rate cut and the subsequent narrowing of spreads, the downward trend in bond issuance that was begun in September did not reverse direction until early 2001 when interest rates plummeted following aggressive easing by the Federal Reserve.

The picture in Exhibit 1 is, of course, only suggestive. Rising interest rates and heightened risk concerns also helped damp issuance following the 1998 liquidity crisis, confounding the identification of any effect from illiquidity. Furthermore, while there would seem to be little room for argument about the presence of a break in the series in autumn 1998, one might examine the issuance rebound in early 1999, or even late 1998, and argue that there was no lingering impact. To this extent the relatively quick rebound in issuance potentially hides lingering effects in the *composition* of issuers, rather than in the number of issuers or amount of issuance.

This paper, in part, documents the impact of liquidity shocks on the composition of firms that enter the corporate bond market. One difference is evident from Exhibit 1, which is that the share of investment grade issuance rose relative to high-yield ("junk") issuance. Throughout 1997 and 1998 the share of junk issuance climbed, and after August 1998 the share fell significantly (and the gap between the moving average of total issues and high-yield issues widened). While credit concerns certainly played a roll in the decline of high-yield issuance, I am going to argue that the bigger compositional effect was via the market's emphasis on issue liquidity - in particular on issue "size" and "familiarity". When liquidity is at a premium, larger bonds are much more prominent.

---

[1] Presented at the BIS "Third Joint Central Bank Conference on Risk Measurement and Systemic Risk" under the title "The impact of market liquidity in times of stress on the corporate bond market: pricing, trading, and the availability of funds during heightened illiquidity". I thank conference participants and Daniel Covitz for helpful comments and suggestions. Sandeep Sarangi provided excellent research assistance. I am responsible for errors and omissions. The views expressed do not necessarily represent those of the Federal Reserve Board, System, Staff, or Governors.

Exhibit 2 suggests a spike in the relative issuance of larger bonds after the LTCM crisis, and that there was some persistence in this change in composition. I will show that this shift was driven, at least in part, by a demand for liquidity by investors and underwriters and distinguish it from various alternatives that could also account for the change. Of course, since large bonds are more likely to be issued by larger companies, it could well be that issuer characteristics rather than issue characteristics prompted the shift to larger bonds. This explanation is not really independent of my liquidity hypothesis, since the liquidity of an issue may be influenced by multiple factors, including issuer characteristics. For instance, the size and "familiarity" of the issuer may matter for liquidity because investors have done more research on these companies and there is potentially less private information.

While it is difficult to measure "familiarity", one proxy in the context of the debt markets is the amount of debt that the firm has issued. Not only is past issuance evidence of past (and ongoing) investor scrutiny, but it may also suggest some substitutability between bonds of the same issuer which could generate liquidity. Exhibit 3 is suggestive of some impact from the LTCM crisis onto the debt outstanding of bond issuers at the end of 1998.

The paper proceeds by reviewing the literature establishing that bond size could be a factor in the amount of trading activity, and therefore liquidity. Then, using multivariate regressions to control for observable issue and issuer characteristics, I establish that issue size, and certain measures of familiarity, are priced liquidity factors. In particular, the price depends crucially on whether the economy is experiencing an illiquidity shock. Moreover, the estimated effect is likely to understate the true effect as the sample of bonds issued tends significantly towards bigger bonds in times of illiquidity.

This new evidence that bond size is a liquidity factor contributes to our understanding of liquidity in the corporate debt markets. First, it helps establish that both issuer *and* issue characteristics matter for an asset's liquidity. The fact that first issues, issues by private firms and issues into the 144a (private) market are all more expensive suggests that information problems are priced at issuance. Likewise, the fact that multiple issues and large issues are discounted suggests that the prospects of wider ownership translate into more trading and more liquidity for the securities. Both of these are consistent with theories of liquidity. Second, it seems clear that the effects of liquidity, or illiquidity, go beyond market pricing and extend to the composition of who is in the market. From the perspective of market watchers, this is a hidden cost of heightened illiquidity.

The paper continues in Section 2 with a discussion of the previous theoretical and empirical literature on the sources of liquidity as well as some extensions to thinking about the corporate bond market. Sections 3 and 4 present the empirical tests of size and other liquidity factors. Section 5 then concludes.

2. Previous literature and the plausibility of issue characteristics as liquidity factors

2.1 Previous theory

The market microstructure theory from equity markets provides a basis for hypothesising that size matters. In general the bid-ask spread, which proxies for liquidity, has been modelled as dependent on three factors: order processing costs, inventory costs and adverse-selection costs (see, for instance, O'Hara (1995)). Empirical work on the contribution of these three factors to the bid-ask spread vary tremendously (see, for instance, Stoll (1989), George et al (1991) and Huang and Stoll (1997)), although both the theoretical and empirical literature has come to emphasise the roll of information problems (adverse-selection costs). But the relevant point here is that the same factors can be thought of as operating in the debt markets. While it is not necessary, it can clearly be argued that issue size could impact relative costs across any of those three dimensions.

The basic idea motivating size as a liquidity factor is that large issues will trade more frequently. Information costs may also be reduced, not only by more trading activity, but because investors will be more knowledgeable about a larger issue because it is more widely held and analysed - it is more transparent (these are the same motivations offered in Crabbe and Turner (1995)). Trying to distinguish between what is issue-specific and issuer-specific liquidity is one of the goals of the paper.

2.2 Intuition for liquidity in the corporate bond market and the LTCM effect

In the Appendix I propose a stylised model of trading in the corporate bond market to help think about the rise of liquidity problems and its effect on the market. In the model illiquidity is the result of an information problem about the correct market prices, which generates a lemons problem in the sense of Akerloff (1970). The lemons problem, in this case, is mitigated by "informed" traders because they compete with each other for trades (rather than with the market-maker, as in the equity microstructure literature (see O'Hara (1995)), which instead generates the lemons problem when there are too many, not too few, informed traders). Thus, the extent of liquidity is determined by the availability of "informed" traders in what amounts to a search framework. *Liquidity is therefore linked to size* because larger bonds will be more widely held and disseminated, leading to more informed traders, and more liquidity, in bigger bonds.

Informed traders may also be determined by their "familiarity" with the bond being traded, or with close substitutes - close substitutes may be other bonds issued by the same issuer. Both paths lead to more informed traders, and more liquidity, for larger bonds. This secondary market phenomenon can translate into reduced issuance during illiquid times because firms (issuers) may not want to pay a large liquidity penalty. Underwriters are also less likely to bring small deals due to the same lemons problem. Underwriters must take the bonds into inventory and then sell them to investors, and during illiquid times they are less likely to do that. They must also be willing to act as dealers and make a market in the bond to help ensure liquidity.

Underwriters (dealers) do not like to hold unhedged inventory (particularly over quarter-end, and especially over year-end) because inventory is risky and firm capital must be set aside to account for that. But if the inventory can be easily hedged, dealers' positions are protected. When dealer willingness to take positions is reduced and/or the cost of hedging climbs, then dealers will not provide liquidity - they will simply be another informed investor. This distinguishes dealers from market-makers, of course, since they are not required to take the other side of trades.

In 1998, dealers suffered a shock across three related dimensions. Bond trading positions suffered losses, and dealer hedges blew up. This gave dealers losses on their positions and on their hedges while also dramatically increasing the cost of hedging. Trading losses led Wall Street firms to cut all positions, including dealer positions that were not necessarily related. At the same time the dealers' own losses gave them an incentive to reduce inventory exposure.

In 1998, the typical way for corporate bond dealers to hedge inventory was with a short position in the 10-year Treasury security. When that hedge proved ineffective - corporate prices fell while a flight to quality drove up Treasury prices - the cost of hedging climbed. Hedges that protected against spread risk were required, and since corporate bond futures and options are non-existent the swap market was the only alternative.[2] Swap spreads skyrocketed and thus so did the cost of hedging. Dealer inventories were slashed and new bond issuance was curtailed.

The importance of inventory for liquidity is an old idea. Demsetz (1968) views inventory costs, and thus the bid-ask spread, as dependent upon "waiting costs" which depend on the frequency of transactions. Thus bonds that trade more often have lower costs and spreads - they are more liquid. Demsetz (1968) shows that the specialist ends up taking more positions in slow-trading stocks - consistent with the specialist taking on more inventory and hence setting higher spreads. Dealers' sensitivity to inventory is also pursued by Ho and Stall (1981), who show that if dealers accumulate too much inventory they will lower their offer price and increase the bid-ask spread to accumulate trades on the other side. The assumption that dealers will want to reduce exposure to inventory is similar to theirs derived from a maximisation problem. That is, one could imagine dealers (and other informed investors) incrementally widening spreads as too many sell orders arrive. Spulber's (1996) search model for bid-ask spreads is similar.[3] He has no "explicit costs of search", rather the search time is the transactions cost, but it yields each "dealer" some local monopoly power. Grossman and Miller's (1988) analysis also focuses on liquidity as the "price of immediacy". Routledge and Zin (2001) instead emphasise the role of the hedge available to the market-maker.

---

[2] Hedging strategies related to short positions in the asset would require selling the asset and thus put the dealer in the same position as everyone else.

[3] Hall and Rust (2001) extend Spulber (1996) to show how dealers and market-makers can coexist.

2.3 Previous empirical evidence

Surprisingly limited previous empirical examination exists on liquidity in debt markets, although the LTCM collapse and declining supply of Treasury debt has sparked recent interest (see, for instance, Fleming (2001)). Studies of the corporate debt market have been even rarer, presumably because of the lack of trading-level data.

Much more analysis has occurred on equity markets, where the availability of "tick" data and market quotes exists. The equity literature speaks a bit to the question of the relation between liquidity and issue size. In the equity market literature it is well established that small stocks are more subject to non-trading effects (Lo and MacKinlay (1990)) and to larger relative bid-ask spreads (see, for instance, Campbell et al (1997), Section 3.2). Less liquid stocks have also been shown to be more sensitive to trade size (Hausman et al (1992)).

The same has been assumed to be true for bond markets. For instance, Fenn (2000, p 397), in discussing a regression with spreads as the dependent variable, asserts that the "expected sign on issue size is negative, as larger issues are thought to be somewhat more liquid". Fenn (2000) indeed, in an analysis of 144a issues, finds significant results consistent with this expectation. Blackwell and Kidwell (1988), however, in a comparison of public and private bonds, find no significant link between issue size and yield. Crabbe and Turner (1995), in a narrower investigation of the MTN market, also find no significant link between issue size and yield.

Research on Treasury market liquidity has been more extensive than for the corporate market, but still limited relative to equities. Analysis of the Treasury market has focused on *measures* of liquidity, such as trading volume, trading frequency, trade and quote size, bid-ask spreads and the on-the-run/off-the-run spread, and the effect of liquidity on prices (see, for instance, Fleming (2001)). Little work has focused on the factors causing liquidity in the bond market, except for going off-the-run. In one exception, Sarig and Warga (1989) show that the age of the bond is a liquidity factor. The link between age and liquidity is assumed to be that bonds eventually end up in buy-and-hold portfolios and so cease to trade. If true, this also supports the contention that size is a liquidity factor, since the amount outstanding to be traded should be proportional to size.

3. Existence of a large bond liquidity premium

If large bonds are indeed more liquid then this liquidity should be priced by the market. One standard, and relatively clean, way to test this is to put bond spreads at issuance as the dependent variable of a regression and determine if the liquidity factor affects bond spreads in the predicted direction (as in Fenn (2000) and Blackwell and Kidwell (1988)). That too is the approach taken in this paper. Spreads at issuance are, in fact, preferable since they are typically quite accurately observed.

To test the hypothesis that issue size is a liquidity factor I use data on all US non-financial straight bond issuance from 1994 to 2001. The spread is calculated as the issue's yield to maturity over that of the nearest on-the-run Treasury. The data source is SDC's New Issues database. Restricting the sample to straight debt simplifies the comparisons, since yields on convertibles are misleading without accounting for the equity piece. Pass-throughs, floaters, medium-term note programmes, asset-backed, lease- or mortgage-related, equipment trusts, and bonds with guarantees are all eliminated. That leaves 2,639 bond issues in the full sample.

The key to specifying this test is to control for the macroeconomic, issue and issuer characteristics that will also move the spread. Within this framework we can also control for alternative hypotheses regarding what drives liquidity or for why size might matter for non-liquidity reasons. For instance, a prominent alternative explanation for why size might matter for spreads is that it is a default risk factor. Therefore the independent variables include: (1) variables for testing the size-liquidity hypothesis, (2) variables measuring issue characteristics, (3) variables measuring market conditions, and (4) variables measuring issuer characteristics. The main variable used to test the size-liquidity hypothesis is the issue size. I also use the time since previous issue or a dummy variable for previous issuance within the year. Other liquidity measures include a dummy variable for multiple issues on the same day and a dummy variable for first bond issue in sample. The first issue dummy uses issuance back to 1993, but earlier issuance is excluded, so if a firm issued a bond in 1992 and 1994 the 1994 issue would be counted as a "first issue" in my analysis. I also use the total debt outstanding from Compustat as a potential measure of liquidity via "familiarity".

The macroeconomic controls include the 10-year constant maturity Treasury yield, the yield curve premium defined as 30-year minus five-year Treasury, the on-the-run premium between the on-the-run Treasury and the fitted synthetic off-the-run yield curve, the spread between BBB-rated and AAA-rated bonds, and the spread between AAA-rated bonds and Treasuries. The last two are important because I give them additional interpretation. The BBB-AAA spread I consider to be the credit spread, since it reflects the reward for the risk differential between those two classes. The AAA-T spread I consider to be the liquidity spread, since short-maturity AAA bonds have essentially zero credit risk. The liquidity spread will be dependent on flight-to-quality and other moves that push investors into Treasuries. While these two spreads are positively correlated, that correlation is only .34, suggesting that they are indeed independent sources of information.

Issue characteristics include the rating notch, coded on a continuum from AAA=1 to CCC=20, so that a higher grade means greater risk (Fenn (2000) shows that a single rating variable fits the data as well as individual dummy variables), the issue maturity, whether the issue had a put or call option, whether the issue was subordinated, and whether it was issued in the 144A market. Issuer characteristics include industry dummy variables and whether the issuer was a private firm. The data are then merged with Compustat to add other issuer characteristics such as firm leverage and coverage, in a more constrained sample.

3.2 Empirical results

Exhibit 4 reports results for the basic spread regression outlined above. Column 1 presents the baseline model. The coefficients on the macroeconomic variables are all significant in the expected direction. Increases in the on-the-run premium increase the spread, presumably due to a decline in market liquidity. A 1 basis point increase in the premium is estimated to raise issuance spreads by 1.3 basis points. Increases in the 10-year Treasury yield also increase the spread, perhaps due to their directional link with overall economy via monetary policy. A 100 basis point increase in the 10-year Treasury is estimated to raise issuance spreads by nearly 14 basis points. The slope of the yield curve, which is well known to flatten before recession and steepen before recovery, affects spreads inversely - a 10 basis point increase in the term structure reduces spreads by 4 basis points. Both the credit spread and liquidity spread push up issuance spreads. A 10 basis point move in the credit spread boosts spreads by 11 basis points, a nearly one-for-one effect, while a 10 basis point move in the liquidity spread boosts issuance spreads by nearly 5 basis points.

Skipping over (for now) the variables for the size-liquidity hypothesis, the other issue and issuer variables are all significant in the expected direction. The coefficient on rating indicates that, conditional on everything else, a one-notch downgrade adds 22 basis points to the spread. The estimated coefficient on maturity indicates that every additional year of length costs .8 of a basis point. Including an embedded put option, which is protection for the bondholder, reduces the spread by 44 basis points, while having an embedded call option, a cost to the bondholder, only increases the spread by 7 basis points and, as seen in later regressions, is one of the few non-robust estimates. The value of the call option appears to be captured by the interest rate and other issuer characteristic variables.[4] A bond issued by a private firm is estimated to pay nearly 62 basis points extra, a subordinated issue to pay an extra 86 basis points, and a 144A issue to pay an extra 65 basis points. The private-firm and 144A market effects may both reflect a penalty paid by firms which may not have to provide as much disclosure, or relatedly, a liquidity penalty by less well known firms. The industry dummies are not broken out for presentation, but they are jointly significant.

3.3 Tests of liquidity and size

The overall fit of the basic regression seems good, suggesting that it is a reasonable model for testing what premium investors attach to issue size, as well as to other liquidity indicators. All of the included liquidity variables are highly significant in column 1. First issues pay a 14 basis point penalty, while multiple issues get a 14 basis point reward. The size of the bond issue has a significant coefficient of

---

[4] Call options appear in almost 30% of the bonds. It may be that different types of calls receive different valuation, but, in general, they receive little apparent value.

-0.034, so that the estimated effect of increasing a bond offer by $100 million is to reduce spreads by 3.4 basis points. One standard deviation for issue size in the cross section is about $277 million, yielding an estimated spread change of nearly 10 basis points.

Adding the time, in years, since the issuer's previous issue, shown in column 2, barely changes the results. The coefficient is marginally significant and each additional year since issuance is estimated to add 2.5 basis points to the spread. Including that variable adds a small boost to the size coefficient, and lowers the coefficient and significance of both the call option dummy and the on-the-run premium variable. Adding, instead, a dummy variable for whether the issuer issued a bond previously within the last year, shown in column 3, changes the estimates even less (from column 1). The coefficient on the recent issuance dummy is also marginally significant, implying that a recent bond issue reduces spreads by 7.5 basis points.

Finally, in columns 4 and 5, the Compustat data are added. Both leverage (debt-to-assets) and coverage (interest expense-to-operating income) ratios are significant in the expected direction. Firms with weaker balance sheets and weaker cash flow must pay higher spreads. Total debt outstanding, however, is not significant. This casts doubt on the robustness of the "familiarity" argument, at least as proxied for by that variable. For instance, it is insignificant even when the time-since-last-issue variable is excluded (column 4).

Moreover, the estimated size effect is also weakened. In the reduced sample with the presence of the leverage and coverage variables the estimated effects for a number of the other coefficients are altered and standard errors are larger due to the smaller sample size. For instance, the on-the-run premium and Treasury yield effects are eliminated, the liquidity spread effect is weakened, and the 144a effect is weakened.

Hence, the general conclusion from Exhibit 4 must be that liquidity factors are important for bond pricing, and that issue size appears to be rewarded with lower spreads, but the result is not completely robust. However, in ongoing research (Harrison (2002)) I show that these results change when one includes an interaction between issue size and the liquidity spread to test the hypothesis that the pricing of liquidity during illiquid times is the most sensitive. The effect of size on spreads is completely altered by adding this interaction term. It now appears that the effect of size by itself actually has a positive impact on spreads - that is, pays a liquidity penalty. This is plausible since larger issues must find more buyers for them. One way to attract more investors and to keep the deal from languishing in the underwriter's inventory is to raise the spread.

However, the liquidity premium on size is dependent upon the amount of liquidity in the market, as measured by the liquidity spread. Harrison (2002) shows that the more illiquid the period, the greater the premium on large bonds. The estimated coefficient is robustly significant, suggesting that bond size matters more during illiquid time periods.

4. Bond issue size and liquidity

As discussed in the introduction, since issue size is not exogenous it is very likely that the selection of bonds issued during illiquid periods is biased toward large bonds. This question is pursued in Exhibit 5, which puts the size of bond issuance as the dependent variable and then determines how the macroeconomic liquidity influences (or "determines") the bond size. The results are striking. In particular, the divergence between the investment grade and high-yield results is suggestive. The size of high-yield issues appears to be extremely sensitive to the state of illiquidity. A change in the liquidity spread from 0.74 to 1.34, such as after the LTCM blow-up, is estimated to increase the average bond size by $200-300 million, a more than doubling of the average size. For investment grade firms, the estimated effect is either insignificant or even in the opposite direction.

Notice, however, that the investment grade results on bond size are very sensitive to the credit spread measure, while the high-yield bond size is not at all. This is true even if the high-yield spread is used as the measure of credit risk. This suggests a link between bond size and credit quality for investment grade firms and between bond size and liquidity for high-yield firms. The credit risk channel for investment grade firms may reflect a disclosure-related mechanism that is actually due to the size of the issuer, rather than the issue. The liquidity risk channel for high-yield firms appears to be something specific about the *bond* size. In the Compustat sample the amount of long-term debt that the firm has outstanding is the only significant indicator for bond issue size, which may be a liquidity factor or simply something else related to firm size.

Other liquidity measures besides size are also potentially influenced by the state of illiquidity. Importantly, rating grade is not, suggesting that the changing quality of the sample is not driving the findings related to issuer size. Rating grade matters in every regression, but it does not appear to be systematically moving with illiquidity. This is consistent with recent anecdotal history. For example, in the aftermath of the LTCM liquidity crisis, the first issuers back in the high-yield market were the speculative telecoms firms. The market's appetite for high-risk and low-rated telecoms debt would not sate until the sector's overcapacity became apparent in 2000.

5 Discussion

Recent experience shows that a severe liquidity shock (1998) is in some ways as bad for the corporate bond market as a severe credit quality shock (2000-01). In both cases credit spreads widen, even though in the case of the credit quality shock spreads widen more. But issuance was more strongly curtailed in the case of the liquidity shock (1998). This shuts some firms out of the public debt market, and thus makes it more difficult for them to obtain financing. However, the reality is that most firms do not need to come to the bond market very often, and thus a temporary closing of that financing venue (even for a period of three months) does not pose serious consequences to the underlying economy.[5]

Rather, this finding simply emphasises that the effect of liquidity on the corporate bond market goes well beyond the secondary market by also affecting the primary market. The impact of illiquidity on investors, and on trading activity, may well be more troublesome than the impact on issuance. Nonetheless, problems in the primary market reflect the problems in the secondary market. Central bankers interested in monitoring liquidity can therefore also look to the primary market. Of course, liquidity problems in US fixed income markets were mitigated by emergency Federal Reserve rate cuts in both October 1998 and January 2001.

Examining the primary market provides additional insights into what issue and issuer characteristics may be fundamental liquidity factors. This study, in particular, focuses on the roll of issue size and its sensitivity to illiquidity. By looking for liquidity factors in market prices, I am assuming that the market recognises and prices liquidity. Identifying fundamentals therefore only helps in our understanding of how liquidity works and what attributes are valued by the market. This could be helpful in building "liquidity" portfolios and identifying liquidity returns. Merrill Lynch, for instance, tracks a corporate bond index of the 175 most active high-yield bonds, as well as both "large cap" and "small cap" high-yield indices. Such evidence is also useful for theoretical considerations of the sources of market liquidity.

[5] For instance, I find that only between 5 and 10% of high-yield firms issue bonds in a given quarter, and only around 10% will issue additional bonds within a year.

Appendix
Stylised model of liquidity in the corporate bond market

Model set-up

The true market value of a bond is uncertain. It is distributed uniformly on an interval +/- Φ around P*, with E(P) = P*. Investors go to the market to buy or sell and must search for a partner to trade with. The search is random but costly. The partner can be either informed or uninformed. Informed traders exist in the population in the proportion $\forall$, to be described later. Uninformed traders are (1-$\forall$) likely. Assume that the seller is informed. Informed traders know the correct price, P^I, a draw from the interval around P*.

Consider a seller who solicits an offer from an uninformed trader. Ignore the search costs for now. What offer does the uninformed trader make? The expected price is P*, but to offer P* is not optimal since it invites trades from an informed seller only when $P^I < P^*$. To avoid this adverse selection the uninformed traders must offer $P^{LO} = P^* - \Phi$. This is the lemons problem in the corporate bond market. If there are only uninformed traders (except the seller) then no trading occurs, unless the seller must sell for other reasons - in which case P^{LO} prevails.

Now consider a seller soliciting an offer from an informed trader, again ignoring the search costs. The informed trader offers P^I, since to offer anything lower than that is to lose the difference to the next informed trader that the seller can find. The informed partner only has monopoly power up to the cost of searching for the next informed trader, and thus this is the extent of the price concession that they can extract from an informed seller. (For the sake of bargaining, imagine that it is costless to refresh a previous offer.)

Assume that the cost of searching is $\exists$, for now take it as a fixed cost, but it can also be a variable cost, which may be important for sellers needing to sell off a particularly large position. Since it costs $\exists$ to replace a partner, each offered uninformed price will actually be P^{LO}-$\exists$; due to the search costs even the uninformed trader can extract rents. For the offered informed price it still costs $\exists$ to find a new price, but the informed partner is more difficult to replace since they are rare, and the offered price will be P^I-$\exists/\forall$. This follows from the decision rule of the seller: search again as long as the expected benefit from searching exceeds the cost. Which gives the strategic partner the optimal policy of setting the price right at this cutoff point.[6] The haircut is intuitive: if there is a 50% chance of finding an informed partner then the price concession can be twice as big.

The analysis of the decision rule is identical if the offer is made by an uninformed partner. The uninformed partner will not set the haircut so as to deter the seller from searching for an informed trader because they do not know P^I (and the optimal informed offer price). To attempt this would lead them to increase their price, which they will not do, since it would result in them being the victim of adverse selection. But they are strategic in discounting the price by $\exists$.

This generates the expected price to the seller of: $(1-\forall)*(P^{LO}-\exists) + (\forall)*(P^I-\exists/\forall)$. We can see that having informed investors mitigates the lemons problem, up to a point. The smaller $\forall$, the larger $\exists$, and the smaller P^I, the more likely that the benefit from finding an informed trader does not meet the cost and both types of partners (informed and uninformed) will offer the same P^{LO}-$\exists$ price.

The preceding assumes that the seller is small relative to the market. Now allow the seller's impact relative to the market to vary. We do this by assuming that 8 is the probability of receiving a sell shock. The amount of selling therefore becomes important if 8 is big - so that many investors are receiving the shock. To see this consider the probability of finding an informed trader, which ex ante is $\forall$. But if each

[6] The decision rule is to search again if [expected (benefits) > costs]. If the investor searches again then the probability of improving is $\forall$ which generates benefits of (P^I-P^{OI}) where P^{OI} is the price offered by the informed partner, with probability (1-$\forall$) the investor is worse off and will revert back to P^{OI}, the previous offer. In this case the investor will execute the same decision rule on whether to search again, facing the same costs and benefits. Along this branch of the tree, then, there is $\forall$ probability of benefit (P^I-P^{OI}) and (1-$\forall$) of continuing. Due to this structure, regardless of whether it is viewed as a multi-period or one-period problem, the solution for maximising P^{OI} for the strategic partner is the same: $P^{OI} = P^I-\exists/\forall$.

informed trader has a probability 8 of being a seller too, then it becomes more difficult to find a trade, now equal to ∀*(1-8) instead of simply ∀. If 8 > ½ it follows that not all trades can be filled at the informed price. Some must be executed at the lemons price.[7] Bonds where there are more informed traders always have a smaller lemons premium, but there is always a 8 such that no trading occurs and the uninformed price is offered by everyone.

Furthermore, once a trade occurs, if the price is observable, investors can update their prices. Uninformed investors can infer P^I from a trade not at the low price and update their information to offer the informed price. In this case ∀ is equivalent to one, all investors are informed. Conversely, if a trade is executed at the low price then informed investors will infer that they can extract additional rents from a desperate partner and so will update their information to offer the lemons price. In which case the price is not informative and trading dries up, except for the most desperate sellers. In this case ∀ is equivalent to zero, all investors are uninformed and a lemons market results. The model offers no dynamics, but it is intuitive that trades at the low price will lead investors to lower the offer price even more.

Let the number of potential traders (ie market participants) be N. Assume a minimum holding size of M (for instance, $1 million). (Alternatively, we can assume that holdings are diffuse but only those holding the largest positions are informed.) Then the number of holders of a given security is H = G/M where G is the amount issued. The number of holders of a close substitute is R = O/M where O is the amount of closely substitutable debt that is traded (think of other debt issued by the same company within recent history). Thus ∀ = (R+H)/N. The point is that ∀ is constructed to depend on R and H - the size of the issue and the amount of other debt the firm has recently outstanding. Later we offer extensions so that ∀ depends on the amount of trading.

This model can generate a loss of liquidity as a result of large price declines

In this model a loss of liquidity is not arbitrarily assumed, rather it is generated by large price declines. Large price declines increase 8 as investors are forced to sell to eliminate losing positions (or meet margin calls) and/or to meet redemptions. This reduces liquidity. Similarly, price declines reduce dealers' willingness to make a one-sided market since they want to reduce not build inventory (and also since hedging costs have increased), which can have a large effect on liquidity since, if they pull back, the probability of a trade falls from 1 to ∀*(1-8).[8]

Extensions

There are two additional intuitive predictions which could be generated from this framework. The first is to show how shocks can be transmitted from one asset to another as sellers (and dealers pulling back) drain liquidity from each market in turn - since a seller will rather sell a different bond than be forced to sell at the lemons price. If the selling is strong enough, the lemons price (which could be different) will result in each market.

Second, additional insight into liquidity can come from a richer view of investor type. "Mark to market" investors (hedge funds and mutual funds) are subject to "sell" shocks when prices fall (but not when they rise). Hedge funds suffer a "sell" shock when prices fall since they must mark to market and meet margin calls. Mutual funds are assumed to be unlevered, but face redemptions. "Buy and hold" investors (insurance companies and pension funds) do not face sell shocks. They never sell, but they are not informed, therefore as they accumulate bond share the liquidity of that bond dries up. Thus, liquidity for a bond diminishes over time as buy-and-hold investors accumulate share and reduce trading.

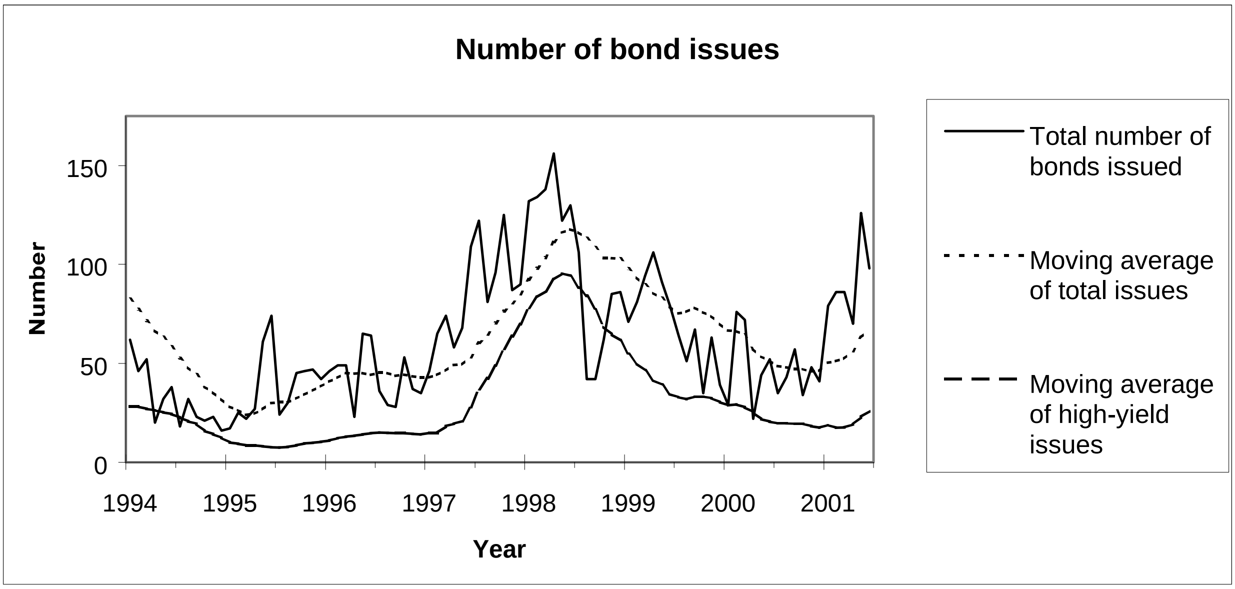
---

7 This would be similar if a seller has to move a particularly large amount of bonds. Or, if the penalty is increasing in the quantity, then it would be more likely to get a trade done at the lemons price.

8 Price increases reduce 8 and so increase liquidity. If positive "buy" shocks were also possible the resulting symmetrical illiquidity of "too much" buying is eliminated by dealers' willingness to stay in the market (as opposed to on the downside) and by their willingness to bring a fresh supply of new bonds. Unfortunately, when the market needs to sell, the issuers have not typically entered the market to retire their debt. Of course, that probably would be an optimal outcome, if the firm had cash on hand.

Exhibit 1

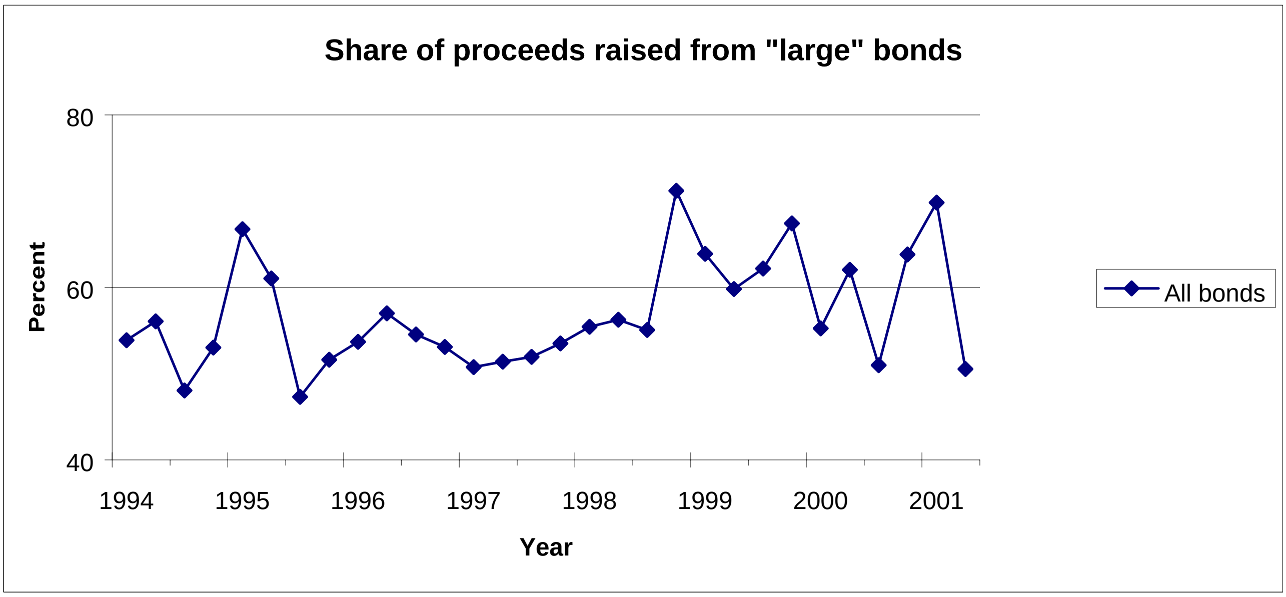
The effect of the LTCM crisis on amount of US non-financial bond issuance



Note: Data are author's calculation from SDC issuance data. US dollar bonds only, issued by US domiciled firms (ie excluding euros and yankees). Non-financial firms only, excluding asset-backed, mortgage-related, and issuance from MTN (medium-term note) programmes.

Exhibit 2

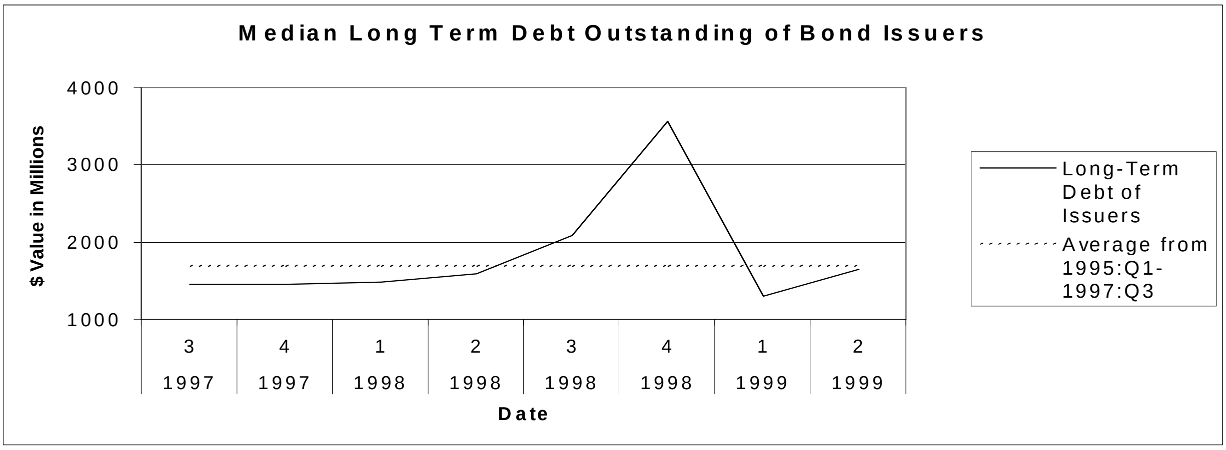
The effect of the LTCM crisis on the amount of "large" bond issuance



Note: Bond sample as in Exhibit 1. "Large" is defined as the upper size quartile as determined by the prior year of issuance.

Exhibit 3

The effect of the LTCM crisis on the amount of "name" bond issuance



Note: Bond sample as in Exhibit 1. The amount of long-term debt outstanding of the issuer is taken from Compustat for the quarter of the bond issue.

Exhibit 4

Impact of issue size and other indicators of liquidity, as well as various controls, on bond pricing

Dependent variable = spread to Treasuries	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
On-the-run premium (versus synthetic)	1.30** [2.31]	1.02* [1.89]	1.32** [2.34]	0.81 [1.18]	0.80 [1.16]
Treasury yield (10-year constant)	13.85*** [3.80]	13.77*** [3.93]	13.75*** [3.78]	5.00 [1.07]	5.27 [1.13]
Yield curve premium (30-year minus 5-year)	–42.85*** [8.37]	–42.20*** [8.55]	–42.73*** [8.35]	–48.91*** [7.07]	–48.21*** [6.97]
Credit spread (BBB - AAA)	112.44*** [16.35]	120.43*** [18.10]	113.06*** [16.43]	112.27*** [11.66]	111.36*** [11.56]
Liquidity spread (AAA-T)	48.91*** [5.25]	54.00*** [6.03]	48.86*** [5.25]	25.98* [1.90]	27.40** [2.01]
Rating grade (AAA=1, CCC=20)	21.70*** [46.20]	22.86*** [48.17]	21.59*** [45.69]	17.08*** [25.10]	17.03*** [25.01]
First issue? (1 if “yes”, 0 if “no”)	**13.62*** [3.70]**	**14.13*** [3.56]**	**10.65*** [2.65]**	**17.83*** [3.58]**	**22.27*** [4.05]**
Multiple issues (same day)? (1 if “yes”, 0 if “no”)	**–14.26*** [3.52]**	**–11.91*** [3.48]**	**–15.17*** [4.27]**	**–18.07*** [4.02]**	**–16.79*** [3.70]**
Time since previous issue(years)		**2.51* [1.71]**			**3.65* [1.90]**
Issue in previous year? (1 if “yes”, 0 if “no”)			**–7.51* [1.88]**		
Amount issued ($ millions)	**–0.034*** [5.38]**	**–0.037*** [6.02]**	**–0.035*** [5.43]**	**–0.011 [1.36]**	**–0.010 [1.27]**
Maturity of issue (years)	0.789*** [4.49]	0.950*** [5.63]	0.801*** [4.56]	0.850*** [3.75]	0.841*** [3.71]
Put option? (1 if “yes”, 0 if “no”)	–43.96*** [6.09]	–45.77*** [6.97]	–43.81*** [6.07]	–49.89*** [5.44]	–49.22*** [5.37]
Call option? (1 if “yes”, 0 if “no”)	7.41* [1.93]	2.81 [0.76]	6.70* [1.74]	3.13 [0.63]	2.84 [0.57]
Private company? (1 if “yes”, 0 if “no”)	61.52*** [6.67]	61.88*** [6.97]	61.44*** [6.67]	na	na
Subordinated issue? (1 if “yes”, 0 if “no”)	86.43*** [13.15]	82.88*** [12.94]	86.29*** [13.13]	103.09*** [9.97]	102.13*** [9.87]
144a issue? (1 if “yes”, 0 if “no”)	64.95*** [8.95]	54.78*** [7.75]	63.20*** [8.64]	24.53** [2.33]	22.50** [2.13]
Industry dummies	Yes***	Yes***	Yes***	Yes***	Yes***
Leverage (debt/assets)				61.25*** [3.93]	62.78*** [4.03]
Coverage (intx/oibd)				1.06* [1.95]	1.06* [1.95]
Long-term debt out ($ millions) (x100)				**–0.026 [0.86]**	**–0.022 [0.73]**
Number of observations	2,639	2,639	2,639	1,185	1,185
Adjusted R-square	.73	.73	.73	.68	.68

Note: Dependent variable is spread to Treasuries on issued bonds. Data are SDC newly issued bonds from 1994 to 2001, excluding financial companies, yankees, euros, asset-backed, pass-throughs, lease-related, mortgage-related, equipment trust-related, MTN programmes, and bonds with guarantees. Straight debt only. Constant term is significant but not reported. T-stats under the coefficients. ***, ** and * are significance at 1%, 5%, and 10% respectively.

Exhibit 5

Determinants of the size of a bond issue

Dependent variable = size of bond issue	Investment grade firms			High-yield firms	
	(1)	(2)		(3)	(4)
Treasury yield (10-year constant)	59.57** [2.38]	84.74** [2.39]		81.30** [2.58]	93.65* [1.67]
Yield curve premium (30-year minus 5-year)	46.81 [0.94]	360.79*** [4.62]		–66.16 [1.13]	–203.54* [1.82]
Credit spread (BBB - AAA)	186.74*** [3.46]	320.18*** [4.05]		–18.32 [1.05]	–21.05 [0.55]
Liquidity spread (AAA-T)	**57.82 [0.83]**	**–261.16** [2.37]**		**320.81*** [3.90]**	**509.32*** [2.75]**
Rating grade (AAA=1, CCC=20)	9.39*** [4.43]	8.46** [2.55]		–18.45*** [4.55]	–13.58 [1.42]
Maturity of issue	0.576 [1.01]	1.29 [1.48]		–1.51 [0.89]	0.73 [0.95]
144a issue? (1 if “yes”, 0 if “no”)	98.44*** [3.13]	42.95 [0.83]		–95.34*** [3.54]	–49.72 [1.04]
Industry dummies	Yes***	Yes***		Yes***	Yes***
Year dummies	Yes***	Yes***		Yes***	Yes***
Leverage (debt/assets)		14.55 [0.20]			–22.46 [0.27]
Coverage (intx/oibd)		–25.93 [0.60]			–0.30 [0.22]
Long-term debt out ($ millions) (x100)		0.83*** [7.17]			2.57*** [4.65]
Number of observations	2,026	983		612	190
Adjusted R-square	.26	.28		.22	.42

Note: Dependent variable is the size of the bond, measured in millions of dollars. SDC issuance data from 1994-2001, as in Exhibit 5. T-stats in brackets under coefficients. ***, ** and * are significance at 1%, 5%, and 10%, respectively.

References

Akerloff, George (1970): “The market for ‘lemons’: quality uncertainty and the market mechanism”, *Quarterly Journal of Economics*, 84, 488-500.

Bank for International Settlements (1999): “A review of financial market events in Autumn 1998”, Committee on the Global Financial System, Bank for International Settlements, Basel, Switzerland.

Blackwell, David W and David S Kidwell (1988): “An investigation of cost differences between public sales and private placements of debt”, *Journal of Financial Economics*, 22, 253-78.

Campbell, John Y, Andrew W Lo and A Craig MacKinlay (1997): *The Econometrics of Financial Markets*, Princeton University Press, Princeton.

Crabbe, Leland E and Christopher M Turner (1995): “Does the liquidity of a debt issue increase with its size? Evidence from the corporate bond and medium-term note markets”, *Journal of Finance*, 50, 1719-34.

Demsetz, Harold (1968): “The cost of transacting”, *Quarterly Journal of Economics*, 82, 33-53.

Fenn, George W (2000): “Speed of issuance and the adequacy of disclosure in the 144A high-yield debt market”, *Journal of Financial Economics*, 56, 383-405.

Fleming, Michael J (2001): “Measuring treasury market liquidity”, *Working Paper*, Federal Reserve Bank of New York.

George, T, G Kaul and M Nimalendran (1991): “Estimation of the bid-ask spread and its components: a new approach”, *Review of Financial Studies*, 4, 23-56.

Grossman, S J and M H Miller (1988): “Liquidity and market structure,” *Journal of Finance*, 43, 617-33.

Hall, George and John Rust (2001): “Middle men versus market-makers: a theory of competitive exchange”, *Working Paper*, Yale University.

Harrison, Paul (2002): “Bond size, liquidity, and the price of liquidity in the corporate bond market”, *Working Paper*, mimeo.

Hausman, J, A Lo and C MacKinlay (1992): “An ordered probit analysis of transaction stock prices”, *Journal of Financial Economics*, 31, 319-79.

Ho, T and H Stoll (1981): “Optimal dealer pricing under transactions and return uncertainty”, *Journal of Financial Economics*, 9, 47-73.

Huang, R and H Stoll (1997): “The components of the bid-ask spread: a general approach”, *Review of Financial Studies*, 10.

Lo, A and A C MacKinlay (1990): “An econometric analysis of nonsynchronous-trading”, *Journal of Econometrics*, 45, 181-212.

O’Hara, Maureen (1995): *Market microstructure theory*, Blackwell Publishers, Cambridge, MA.

Routledge, Bryan R and Stanley E Zin (2001): “Model uncertainty and liquidity”, *Working Paper*, GSIA, Carnegie Mellon University.

Sarig, Oded and Warga, Arthur (1989): “Bond price data and bond market liquidity”, *Journal of Financial and Quantitative Analysis*, 24, 367-78.

Spulber, Daniel F (1996): “Market making by price-setting firms”, *Review of Economic Studies*, 63, 559-80.

Stoll, H (1989): “Inferring the components of the bid-ask spread: theory and empirical tests”, *Journal of Finance*, 44, 115-34.

Wall Street Journal (1998a): “Bonds finish higher on weakness in stocks; big swings, thin trading volume mark session”, 14 October.

——— (1998b): “Illiquidity is crippling bond world”, 19 October.

Liquidity of the Hong Kong stock market since the Asian financial crisis

Jim Wong and Laurence Fung,[1]
Market Research Division, Research Department
Hong Kong Monetary Authority

Abstract

This paper looks into how the liquidity of the Hong Kong stock market has evolved since the Asian financial crisis, and examines the determinants of changes in liquidity. Various conventional liquidity indicators are constructed for the study period from 1997 to June 2001. They show that, having deteriorated during the Asian financial crisis and the Russian crisis, market liquidity has mostly recovered to the pre-crisis level in the more recent period. However, these conventional liquidity indicators have the drawback of not being able to capture fully the dynamics of liquidity. Thus, a GARCH model is developed for five selected stocks to relate the sensitivity of their price movements to net order flows, using a unique set of 30-second tick-by-tick data of the Hong Kong Stock Exchange. Empirical results from our model illustrate clearly a sharp deterioration of market liquidity during the crises, followed by an apparent recovery in the post-crisis period. Based on a simple OLS regression estimation, we also analyse the determinants of the time-variation of market liquidity. It is found that financial crises exerted their influence on local liquidity mainly through their effect on domestic interest rates and price volatility, while global liquidity and risk conditions also played a significant role.

1. Introduction

The liquidity of financial markets stood out as a critical issue in both the Asian financial crisis and the Russia/Long-Term Capital Management (LTCM) crisis. Being one of the most liquid markets in the world, the Hong Kong stock market often served as a hedging tool for emerging markets in the region in periods of heightened uncertainty. As a result, Hong Kong's stock market is extremely sensitive to external factors. The turbulence in the 1997 and 1998 financial crises had placed tremendous pressure on liquidity and the efficient functioning of Hong Kong's stock market, and tested Hong Kong's ability as an international financial centre in withstanding the shocks.

Numerous studies on the dynamics and determinants of market liquidity have been initiated by policymakers and academics. While some studies indicated that the liquidity conditions in Hong Kong's markets have generally improved from the lows reached during the region-specific shocks,[2] local market sentiment remains fragile. Market sources suggested that market participants remained concerned about liquidity, as investors and traders have become more risk averse, and various players have withdrawn from active trading.

Liquidity of the stock market is a good barometer for the proper functioning of a market as it measures the degree of easiness with which stocks can be traded. A mature stock market should be an efficient discounting mechanism and an effective exchange for channelling invested capital to the real economy. From a financial stability perspective, it is important to monitor liquidity during normal times

---

[1] The views expressed in this paper are solely our own and not necessarily those of the Hong Kong Monetary Authority (HKMA). We are grateful to Stefan Gerlach and Grace Lau of the HKMA, Prof Win-lin Chou of the Chinese University of Hong Kong and internal seminar participants for useful comments, and to Polly Lai for excellent secretarial assistance. All remaining errors are ours.

[2] BIS (2001).

and at times of stress, and to promote structural changes that will enhance the liquidity of the stock markets.

To facilitate this process, this paper examines mainly two issues: (i) it looks into how the liquidity of Hong Kong stock market has evolved since the Asian financial crisis, and (ii) it examines the determinants of changes in liquidity. For the first issue, various conventional indicators are constructed to gauge market liquidity during the study period (covering 1997 to June 2001), by assessing mainly market depth. In particular, the paper assesses whether liquidity conditions have recovered to the pre-crisis level. To supplement the conventional liquidity indicators, using a unique set of 30-second tick-by-tick data of the Hong Kong Stock Exchange, a regression model which relates the sensitivity of stock prices to the prevailing order book conditions is built to examine the changes in market depth during the period. For the second issue, results of the above regression analysis are utilised to construct a model to assess the determinants of liquidity. It is found that financial crises exerted their influence on local liquidity mainly through their effect on domestic interest rates and price volatility, while global liquidity and risk conditions also had a significant impact on domestic liquidity.

2. Definitions and measures of liquidity

Market liquidity is difficult to define, given its multifaceted nature. Broadly speaking, there are mainly three possible dimensions of market liquidity: tightness, depth and resiliency. Tightness measures how far the bid or ask prices diverge from the mid-market prices. It is important to market players as it measures the costs incurred. Of the various indicators, the bid-ask spread is one of the most frequently used. Depth refers to the volume of trades possible without moving prevailing market prices. Conventionally, it can be measured either by the order amount on the order books, or by the fluctuation in bid-ask spreads as a result of market impact from order executions. The greater the relative imbalance of buy or sell orders, the farther the market price must diverge from the standard bid or ask prices to clear the imbalance. The relative sensitivity of market prices to a unit of imbalance of order flows may also reflect the relative depth of the market. Resiliency measures the speed with which price fluctuations resulting from trades reconverge, or the speed with which imbalances in order flows are dissipated.[3] Market resiliency gives us a picture of potential market depth, which cannot be observed from prevailing order flows.[4] There is no clear-cut approach to measure resiliency, and one approach is to examine the speed with which the bid-ask spread and order volume are restored to normal market conditions after trades.[5]

Other measures of market liquidity include price volatility,[6] the number and volume of trades, trade frequency and turnover ratio. Among these, price volatility is the most widely used measure, and is closely related to the market depth indicators (it is in fact sometimes treated as one of the depth indicators).

Given the trading system in Hong Kong, where the spread varies predeterminedly according to a set of price ranges for all stocks, market tightness cannot be readily measured from changes in the observed bid-ask spreads.[7] In this paper, we therefore focus mainly on the depth dimension of market liquidity as well as the price volatility indicators.

---

3 Another commonly used concept is immediacy, which is defined as the time necessary to execute a trade of a certain size within a certain price range. Because immediacy incorporates elements of all three of the above dimensions, it is not considered as a separate dimension.

4 Engle and Lange (1997).

5 Muranaga and Shimizu in BIS (1999a).

6 If one assumes a constant level of "true" (ie fundamentals-based) prices, then volatility in observed prices could reflect the bid-ask spread, the market impact of trades, and/or the degree of resiliency. Cohen in BIS (1999a) uses this concept to examine the liquidity of short-term money markets. Specifically, he investigates the linkages between the volatility of various short-term interest rates under different monetary policy operating regimes for nine developed countries.

7 A brief note on the trading system in Hong Kong is given in Annex A.

3. Variations of market liquidity since the Asian financial crisis

3.1 Conventional liquidity indicators

To assess how market liquidity in Hong Kong's stock market interacted and evolved, the following market-wide indicators measuring market depth and volatility, as discussed in Section 2, are constructed based on the daily closing trading statistics of the 33 constituent stocks of the Hang Seng Index (HSI). As these 33 stocks accounted for almost 80% of Hong Kong's stock market capitalisation during the study period (see below), their aggregate liquidity condition should be representative of the overall market.

3.1.1 The indicators

a. Market depth

Traditionally, market depth is measured by a variety of trading activity variables. One measure is the average turnover in a given time interval (such as a day or a week), which is an indicator for normal order flow. A more sophisticated measure of market depth would be to measure the effective supply and demand, which is the sum of actual trades by market participants and potential trades as a result of portfolio adjustments.[8] Other proxies for market depth are the size of trades that market-makers can accommodate[9] and the volume per trade. In this paper, trading volume and turnover value are used to reflect the market depth and they are constructed also as a ratio to both interday and intraday volatility.[10]

b. Price volatility

A widely used measure for price volatility is the interday price volatility, which is readily available from the daily closing price. However, as this volatility measure is not able to reflect within-day price fluctuations, the intraday price volatility is also considered.

To summarise, the following indicators are constructed for the market-wide analysis:

Market depth measures:

Volume:	Total number of shares traded during the day
Value:	Total turnover value (in Hong Kong dollars) during the day
Depth I (III):	Trading volume (or value) per unit of interday volatility
Depth II (IV):	Trading volume (or value) per unit of intraday volatility

Volatility measures:

Interday volatility:	Defined as the square of the daily percentage changes in closing prices, market capitalisation-weighted
Intraday volatility	Defined as (Day High–Day Low)/[(Day High+Day Low)/2]*100%

---

8 Though there are few examples of research to-date in this area, partly because information on order flows is difficult to obtain, Muranaga and Shimizu in BIS (1999a) investigate the dynamics of market depth by constructing simulated markets. Muranaga studies market impact by examining high-frequency data on transactions involving individual stocks listed on the Tokyo Stock Exchange.

9 BIS (1999a).

10 Trading volumes and values by themselves are inadequate measures for market depth. For example, an absence of transactions or low turnover does not necessarily imply the market is illiquid, as investors may wait for their "best" bid-ask quote to trade. On the other hand, high turnover may not mean the market is deep enough if stock price variation is high, which may lead to a widening of spreads. They should therefore be measured against the prevailing price volatility.

3.1.2 *Study period*

The analysis in this section covers the entire period from January 1997 to June 2001. To facilitate comparative analysis of liquidity during the normal and crisis periods, the study period is further divided into the following five sub-periods:

Pre-crisis period:	January 1997-19 October 1997
Asian financial crisis period:	20 October 1997-April 1998
Russia/LTCM crisis period:	May 1998-28 September 1998
Post-crisis period:[11]	29 September 1998-end December 2000
Recent period:	Jan 2001-June 2001

The above division of crisis periods follows largely that of the report of the Committee on the Global Financial System,[12] but some modifications are made to reflect Hong Kong's unique situations. Specifically, while the beginning of the Asian financial crisis is defined as 2 July 1997 in the BIS study, when the Thai government devalued the Thai baht, we define the start of the crisis as 20 October 1997, as the financial market turbulence in Hong Kong only clearly emerged after that day, with the pressure on the Hong Kong dollar and the equity market intensifying.

As for the Russia/LTCM crisis period, it is worth noting that the Russian crisis[13] started on 17 August 1998 when the Russian government effectively defaulted on its sovereign debt and devalued its currency, which largely coincided with the Hong Kong government's operations in the stock market, from 14 August to 28 August, to restore financial market stability.[14] As a result, large turnovers were recorded during this period, along with the rise in stock prices, as shown in Chart 1. Due to this, throughout this paper, other than in Charts 1 to 3, where no exclusions were made, the Russia/LTCM crisis period is defined to exclude the period from 14 August to 28 August, in order to eliminate the distortion caused by the government operation.

3.1.3 *Empirical results and analysis*

The conventional liquidity indicators for different periods are summarised in Table 1 and Charts 1 to 3. As shown in Table 1, market liquidity by all measures deteriorated sharply in the Asian financial crisis, and most of them fell further through the Russia/LTCM crisis. During the crisis periods, the fall in depth was dramatic. For instance, during the Asian financial crisis, market depth measured as the ratio of trading volume to intraday volatility fell by 28%, while in terms of trading value to intraday volatility, it dropped by 43% from the pre-crisis level, reflecting a much shallower market. The sharp falls in depth and rising price volatility all pointed to a rapid evaporation of liquidity in the market during the crisis.

During the post-crisis period, there were distinct trends of a pickup in market liquidity, with market depth improving, and volatility significantly reduced. By the first half of 2001, most market liquidity indicators appeared to have returned to their pre-crisis levels, with some even surpassing them.

---

[11] The post-crisis period is further divided into three sub-periods based on the tightening and easing of interest rate policy by the US Federal Reserve. Period I from 29 September 1998 to 29 June 1999 refers to the round of US interest rate cuts after the financial crises; period II from 30 June 1999 to 15 May 2000 refers to the round of US interest rate hikes; and period III from 16 May 2000 to end-December 2000 refers to the sustained high interest rate era.

[12] BIS (1999b).

[13] The financial trouble regarding Long-Term Capital Management (LTCM) started in early July, but only intensified after massive losses by the company were reported after the Russian default in August. The US Federal Reserve was involved to recapitalise the company on 23 September 1998 in order to prevent a domino effect on other financial institutions.

[14] It was estimated that the Hong Kong government purchased HK$ 118 billion worth of stocks in its attempt to restore financial market stability.

3.2 Sensitivity of stock prices to order imbalances

However, the above analysis suffers from a major deficiency in the use of daily closing data to measure market liquidity, which is changing constantly throughout a trading day. In particular, large and more frequent intraday variations are likely to occur in times of market turbulence. Thus, for an indicator to fully reflect liquidity conditions, statistics capturing changes during the day are needed.

Moreover, most of the conventional indicators characterise the depth of a market as the trade volume or the trade value cleared by a one unit change in prices (also known as liquidity ratios). It is, however, argued that prices change in response to the net disequilibrium in buys and sells, not to total trading volume.[15] Furthermore, the use of liquidity ratios as a measure of market liquidity has its limitations. And they seldom distinguish the sources of price volatility (or price changes). Grossman and Miller (1988) point out that liquidity ratios fail to answer the critical question of how a sudden arrival of a larger than average order would affect price movements. A market's liquidity conditions should thus be measured by its ability to absorb order imbalances without large price changes.

3.2.1 Previous research

Numerous studies have focused on order imbalances and their relationship with market liquidity and other market variables. Chordia et al (2001a) outline two reasons why order imbalances should be more important to stock returns and liquidity than trading volume. First, they argue that "order imbalances sometimes signal private information, which should reduce liquidity at least temporarily and could also move the market price permanently". Second, a large order imbalance exacerbates the inventory risk faced by market-makers, who may respond by widening the bid-ask spread in order to compensate for taking the risk, which in turn further worsens liquidity conditions. Following the same lines of reasoning, a number of studies have emerged to analyse order imbalances. For example, Brown et al (1997) study the interaction between imbalance of bid and ask orders and stock returns in the Australian market. They find that imbalance in terms of number of orders can explain current returns, while imbalance in terms of dollar value can explain both current and future returns. Chordia et al (2001a) examine the relation between S&P 500 returns and order imbalances. They find that there is a strong contemporaneous association between stock returns and order imbalance, and that a contemporaneous order imbalance exerts significant impacts on market returns. These empirical studies indicate that order imbalances affect price movements. Their relationship may thus provide a better measure of market liquidity than the conventional liquidity ratios, such as the ratio of trading volume to price volatility.

However, many of the earlier studies measure the order imbalance based on traded (executed) buy and sell volumes. Furthermore, previous studies often use the number, instead of size, of orders and transactions as a measure of order imbalance, motivated by findings by Jones et al (1994) that the number of transactions is a major determinant of price volatility. The use of traded (executed) buy and sell volumes may be partly driven by the more readily available transaction data from the authorised exchanges. However, with the rising importance of order-driven market structures and the information available from electronic limit order books, attention has rapidly shifted to liquidity provisions in an order-driven market.

The attention to limit orders as the main source of liquidity has been documented by Demsetz (1968). Basically, limit orders can be perceived as a supply of liquidity. Limit orders represent ex ante precommitments to provide liquidity to market orders which may arrive sometime in the future. Thus, following the traditional reasoning regarding liquidity, a liquid limit order market can be characterised as having a large volume of buy and sell limit orders, waiting to be executed at their corresponding bid and ask prices, if and when market orders arrive. To go further, a deep limit order market can be viewed as the ability of a market to absorb a large pool of limit orders without significant impacts on price movements, and the ability to restore the limit order book after a market order is submitted and executed.

As for Hong Kong, a number of empirical studies of its stock market regarding the issue of limit order and order-driven mechanism have been conducted over the past few years. Chan and Hwang (1998) study the impact of tick size on market quality. Ahn and Cheung (1999) and Brockman and Chung

[15] Kempf and Korn (1997), Engle and Lange (1997).

(1998) study the liquidity pattern of the Hong Kong stock market. Brockman and Chung (1999) investigate the intertemporal and cross-sectional depth pattern in an electronic, order-driven environment and find an inverted U-shaped pattern at the weekly, daily and trading session level. They also demonstrate that market depth at cross-sectional, corporate level is negatively related to information asymmetry. Brockman and Chung (2001) find commonality in spreads and depth across all sizes of firms. Ahn et al (2000) investigate the relation between market depth and transitory volatility. However, few have investigated the dynamic relation between price movements and order imbalance as a measure of market depth.

3.2.2 The model

To supplement the conventional market depth indicators, and to remedy some of their drawbacks, using a unique set of 30-second tick-by-tick data of the Hong Kong Stock Exchange, the following model is built to examine the general relationship between the changes in stock prices and the net position of order books:

$$\Delta \ln(P_t) = \alpha + \beta \ln(BSI_t) + \varepsilon_t \qquad (1)$$

where P_t is the share price at time t, BSI_t is the net buying/selling pressure at time t, and ε is the error term. α is the constant term, while the parameter β measures the short-term sensitivity of the changes in stock prices to the contemporaneous order imbalance.

In the equation, $\Delta \ln(P_t)$ is thus the change in share price at time t over time t–1, while BSI_t is the net position of the order book, which is derived by subtracting the total selling orders (of the first five selling queues) at each 30-second tick from the total buying orders (of the first five buying queues),[16] as follows:

BSI_t = the net buying/selling pressure at time t

$$= \sum_{i=1}^{5}(BuyingQueue_i) - \sum_{i=1}^{5}(SellingQueue_i) \text{ in number of shares, at time } t$$

As order imbalance is likely to have a lagged impact on stock prices, lagged variables of $\Delta \ln(BSI_t)$ are introduced into the model. Furthermore, as the 30-second changes of stock prices are likely to exhibit serial correlation, lagged variables of $\Delta \ln(P_t)$ are included in the right-hand side to control for autocorrelation in short-term stock price fluctuations. The basic model (1) is thus extended to be as follows:

$$\Delta \ln(P_t) = \alpha + \beta \ln(BSI_t) + \sum_{i=0}^{m} \gamma_i \Delta \ln(BSI_{t-i}) + \sum_{j=1}^{n} \theta_j \Delta \ln(P_{t-j}) + \varepsilon_t \qquad (2)$$

where m and n are the lag lengths for $\Delta \ln(BSI_t)$ and $\Delta \ln(P_t)$ respectively.

The lag structure of the $\Delta \ln(BSI_t)$ and the $\Delta \ln(P_t)$ variables in the right-hand side is then determined with reference to the Akaike Information Criterion (AIC). The proper lag structure is found to be $m = 8$ and $n = 12$.

Unit root test is performed on the dependent and explanatory variables to check for stationarity. Like many other time series of high-frequency financial data, our data also exhibit the autoregressive conditional heteroscedasticity (ARCH) effects. To capture these, our model is estimated under the GARCH estimation procedure, instead of the traditional Ordinary Least Square (OLS) estimation.

Five constituent stocks from the Hang Seng Index are selected for the analysis. Together, they account for 25% of the total Hong Kong stock market capitalisation.[17] Our analysis will focus on the

[16] Our micro, stock-level study utilises the intraday Bid and Ask Record obtained from the Stock Exchange of Hong Kong. For each 30-second tick, the intraday Bid and Ask Record contains information on limit-order prices and order quantities, including the nominal price of a stock, as well as the number of shares quoted in the first five queues for both buying and selling orders at their respective bid and ask prices.

[17] These stocks are Hang Seng Bank and Bank of East Asia from the finance sector, Cheung Kong Holdings and Sun Hung Kai Properties from the property sector and Hutchison Whampoa from the commerce and industry sector.

coefficient β, which measures the depth of the market. β should have a positive sign. A higher coefficient indicates lower liquidity and vice versa.

3.2.3 Study period

Similar to Section 3.1.2, the models are estimated for the period from 1997 to June 2001, which is divided into five sub-periods. However, as 30-second tick-by-tick data are collected, which involved a huge amount of data per day and substantial downloading and processing efforts, only data for the key months (instead of working out the data for the entire study period) are collected for the analysis. Specifically, the following months during each of the sub-periods are included in this section's analysis:

Pre-crisis period:	May-August 1997
Asian financial crisis period:	20 October 1997-November 1997
Russia/LTCM crisis period:	May 1998-13 August 1998
Post-crisis period:[18]	November 1998-October 2000
Recent period:	Jan-June 2001

3.2.4 Empirical results and analysis

GARCH estimation results of five selected stocks are summarised in Tables 2 to 6 and Charts 4 to 8. As shown in the tables, the estimated parameter β in all cases has the expected positive sign and is statistically significant. The positive relationship between the *BSI* variable and changes in stock prices shows that a net buying pressure drives up stock prices, whereas a net selling pressure pulls down stock values. The magnitude of the estimated value for β measures the sensitivity of changes in stock prices to the net buying/selling pressure, which in turn reflects liquidity conditions of the stock market.

As shown in the charts, the estimated parameter β for all stocks rose during crisis periods from the pre-crisis period. These results demonstrate the worsening of market liquidity during crises. While the worsening of liquidity conditions during the Asian financial crisis seemed to be more severe than during the Russian crisis for three of the five selected stocks, it appeared to be less severe for the other two stocks. As for the post-crisis period, the estimated parameter β declined in general, as the market calmed down and cuts in interest rates improved the liquidity condition from the Russian crisis period. Market liquidity then fluctuated within a narrow range, and for most of the selected stocks it has returned to the pre-crisis level in the recent period.

4. Determinants of market liquidity

Knowledge about what factors determine market liquidity is essential to the understanding of how financial crises exert their impact on market liquidity. Existing market microstructure theories on market liquidity are represented by the "inventory control" and "asymmetric information" models.[19] In general, these models suggest that the willingness of market-makers and investors to trade and invest, which determines market liquidity, is largely dependent on cost and risk factors. Market liquidity is expected to be negatively correlated with the cost and risk level. Thus a decrease in interest rates

---

[18] Similar to Section 3.1.2, the post-crisis period is further divided into three sub-periods based on the interest rate policy of the US Federal Reserve. However, the exact months included in this section are different from that of Section 3.1.2, with only data for key months collected. In this section, period I from November 1998 to March 1999 refers to the round of US interest rate cuts after the financial crises; period II from July 1999 to December 1999 refers to the round of US interest rate hikes; and period III from June 2000 to October 2000 refers to the sustained high interest rate era.

[19] Under the "inventory control" models, bid-ask spread is negatively related to trading volume, but positively related to price volatility. The "asymmetric information" models argue that the widening of bid-ask spread compensates market participants for taking the adverse selection risk, the risk of trading with other market participants with superior information. Contrary to the "inventory control" models, unusually high trading volume is positively related to the bid-ask spread under the "asymmetric information" models.

may stimulate trading interest and enhance market liquidity, while a volatile market would influence liquidity through an increase in inventory and short-term speculative risks.

4.1 Previous research

Based on the theoretical framework, a number of studies have attempted to explain market liquidity by cost and volatility. While based on 30 stocks in the Dow Jones Industrial Average, Hasbrouck and Seppi (2001) do not find conclusive evidence of economically significant common factors in explaining their liquidity proxies. Using data of 240 shares traded in the New York Stock Exchange, and focusing on four traditional proxies of liquidity, Huberman and Halka (2001) show that the temporal variations in their liquidity proxies are positively correlated with return and negatively correlated with volatility. Using a similar set of data, Chordia et al (2000) find quoted spreads, depths and trading activity respond to short-term interest rates, the term spread, equity market returns and recent market volatility. In a recent study, using daily closing data, Chordia et al (2001b) show that lagged market returns, lagged interest rates, the lagged bid-ask spread and lagged volume are strong predictors of the bid-ask spread and volume changes in both the stock and bond markets in the United States.

4.2 The model

To facilitate our regression analysis on the determinants of market liquidity, we utilise the same GARCH model in equation (2) and estimate the model on a monthly basis for the same selected periods as in Section 3.2.3 to obtain a series of monthly estimations of β. Charts 9 to 13 present the monthly movements of estimated β values for the five selected stocks.

For the examination of the determinants of stock market liquidity in Hong Kong, a model is specified to relate β (representing market liquidity) to cost and risk variables. In addition, given Hong Kong's role as a financial centre, the liquidity of the Hong Kong stock market should be affected by fund flows and the global liquidity trend. Market liquidity is therefore a function of the following factors:

$$\beta_t = f(I_t, ID_t, VHK_t, VUSA_t, MLUSA_t, D_{1t}, D_{2t}) \qquad (3)$$

where the dependent variable β_t is the liquidity level in the Hong Kong market at time t, which is proxied by the $\hat{\beta}$ presented in Charts 9 to 13. I_t is the Hong Kong three-month interbank rate (monthly average), representing the cost of investing and trading stocks. ID_t is the interest rate differential between the Hong Kong overnight interbank offered rates and the London interbank offered rates. Other things being equal, a positive ID_t should attract capital into Hong Kong and is positive to liquidity conditions. VHK_t is the intraday volatility of HSI while $VUSA_t$ is the intraday volatility of US stocks, measured by the volatility of the Dow Jones Industrial Average and the Nasdaq Composite Index, market capitalisation-weighted.[20] These two variables represent the domestic and global risk factors respectively. $MLUSA_t$ is the liquidity level of the US market, specified as the ratio of daily turnover of US stocks to the price volatility of the Dow Jones Industrial Average and the Nasdaq Composite Index, market capitalisation-weighted, which is used as a proxy to global liquidity. D_{1t} and D_{2t} are the dummy variables for the Asian financial crisis and the Russian crisis, respectively.

4.3 Empirical results and analysis[21]

OLS technique is used to perform the estimation for equation (3). Models of various specifications (with different combinations of the above explanatory variables) are estimated. The results are summarised in Table 7; it is found that:

---

[20] Defined as (day high–day low)/[(day high+day low)/2] * 100%.

[21] One should note that the variance of the disturbance term in the regression estimations is expected to be large, as the estimation error of the dependent variables β is incorporated in the disturbance term as well. Even though this should cause no problem for the estimation, as long as we model the disturbance term correctly, one should interpret the estimation results and the significance of the estimated parameters with caution.

(i) As expected, domestic interest rates (I_t) is significant and has the correct sign for five stocks in 12 estimations.[22] This indicates that a rise in domestic interest rates would lead to a deterioration of local market liquidity.

(ii) ID_t is found to be highly correlated to I_t (correlation coefficient of 0.80), as the differential between Hong Kong and US interest rates is largely determined by the fluctuation in Hong Kong rates, particularly during the crisis periods. If both of them are included in the regression equation, their estimated coefficients yield wrong signs due to multicollinearity. Furthermore, if only ID_t appears in the model, the estimated coefficient for ID_t consistently has a positive sign. This suggests that the inclusion of ID_t in the model fails to capture the impact of an expected influx of funds (which should yield a negative sign for the coefficient) and has instead reflected mainly the movement of local interest rates. As a result, ID_t was therefore dropped from all the models.

(iii) In line with the "inventory control" models, local market volatility (VHK_t) and overseas market volatility ($VUSA_t$) have the expected positive sign and are significant for four stocks in 14 estimations[23] and four stocks in 13 estimations[24] respectively. This indicates that an increase in volatility in either local or global stock markets would lead to a fall in market liquidity, and vice versa. However, when both local and overseas market volatility are included in the model, Hong Kong share price volatility is statistically significant in most cases, while that of the United States is insignificant (regressions 1 and 4) due to multicollinearity.

(iv) The variable $MLUSA_t$ is significant and has a correct sign for three out of the five stocks in 16 estimations,[25] suggesting that a deterioration of global liquidity conditions may have a negative impact on local market liquidity. It also indicates that $MLUSA_t$ is rather stock-specific.

(v) Naturally, D_{1t} and D_{2t} appear to be very powerful in explaining the sharp rise in β during the crises (regressions 7 to 12). However, whenever D_{1t} and D_{2t} are included in the regressions, other independent variables such as I_t and $VUSA_t$ become insignificant. An examination of the relationship between I_t and $VUSA_t$ separately with D_{1t} and D_{2t} shows that the two variables are highly correlated with the dummy variables. This seems to indicate that the impact of the crises on liquidity conditions might largely be effected through the interest rate and risk levels. As we are more interested in the impact of I_t and $VUSA_t$, the D_{1t} and D_{2t} are excluded from some of the models.

5. Conclusion

In this paper we studied the evolution of the Hong Kong stock market's liquidity since the Asian financial crisis and tried to explain the time-variation of market liquidity. Using a unique set of 30-second tick-by-tick data from the Hong Kong Stock Exchange, empirical results from our GARCH model for five selected stocks, which relates the sensitivity of their price movements to net order flows, confirm the sharp deterioration of market liquidity during the crisis periods. Furthermore, they also illustrate that, in the more recent period, the liquidity of most of the selected stocks has returned to the pre-crisis level.

---

22 Regressions 3 and 6 for Cheung Kong Holdings, Hang Seng Bank, Sun Hung Kai Properties and Bank of East Asia, and Regressions 3, 6, 9 and 12 for the Hutchison Whampoa Limited.

23 Regressions 1, 2, 4 and 5 for Cheung Kong Holdings, Hang Seng Bank, and Hutchison Whampoa Limited, and Regressions 2 and 5 for Bank of East Asia.

24 Regressions 3 and 6 for Cheung Kong Holdings, Hang Seng Bank and Hutchison Whampoa Limited, and Regressions 3, 4, 6, 7, 9, 10 and 12 for Sun Hung Kai Properties.

25 Regressions 5, 10, 11 and 12 for Cheung Kong Holdings, and Regressions 4, 5, 6, 10, 11 and 12 for Hang Seng Bank and Bank of East Asia.

This paper also establishes the correlation of stock market liquidity with cost and risk factors. The findings are consistent with the "inventory control" models, which predict that market depth is negatively correlated with price volatility. Largely in line with empirical studies of US market liquidity, which show that liquidity is correlated with lagged short-term interest rates, lagged market returns and market volatility, our OLS regression analysis also shows that financial crises exert their influence on local liquidity mainly through their effect on domestic interest rates and price volatility. Furthermore, given Hong Kong's role as a financial centre, our results indicate that, to a significant extent, global liquidity and risk conditions have an impact on domestic market liquidity as well.

Annex A
The Hong Kong stock market's bid and ask system

The trading system of the Exchange is an order-driven system, and is fully centralised and computerised, via terminals in the trading hall of the Exchange and terminals of the Exchange's members. Investors initiate buying and selling transactions by placing orders through brokers. These orders are consolidated into the Exchange's electronic limit-order book and executed (with some specific exceptions) through an automated trading system. Information regarding the limit-order book is disseminated on a real-time basis and available to all market participants through an electronic screen. The electronic screen displays the best five bid-ask prices, along with the broker identities and the numbers of shares intended to be bought and sold at each of the bid-ask queues. Orders are executed in strict price and time priority. The spreads vary according to a set of predetermined price ranges for all stocks (Table A1). A stock would have different dollar spreads if its price appreciates or drops to the next level of price range, and it would have different % spreads (as a % of the value of the stock) when prices move even within the price ranges.

Table A1

Spread table of stock trading on the Hong Kong stock exchange

	Price range (HK$)			Spread (HK$)	Spread as a % of price		
From	0.01	to	0.25	0.001	10	-	0.4
Over	0.25	to	0.50	0.005	2	-	1
Over	0.50	to	2.00	0.010	2	-	0.5
Over	2.00	to	5.00	0.025	1.25	-	0.5
Over	5.00	to	30.00	0.050	1	-	0.17
Over	30.00	to	50.00	0.100	0.33	-	0.2
Over	50.00	to	100.00	0.250	0.5	-	0.25
Over	100.00	to	200.00	0.500	0.5	-	0.25
Over	200.00	to	1,000.00	1.000	0.5	-	0.1
Over	1,000.00	to	9,995.00	2.500	0.25	-	0

Tables

Table 1

Liquidity indicators[1] of the Hong Kong stock market: pre-crisis, crises and post-crisis

	Pre-crisis[2]	Asian financial crisis[3]	Russia/LTCM crisis[4]	Post-crisis[5]	2001 H1[6]
Depth					
Volume (m shares)	175.2	243.9	189.7	188.1	232.6
Volume/intraday volatility	103.7	74.6	65.9	89.2	140.1
Volume/interday volatility	59.4	15.8	26.4	36.2	104.8
Value (HK$ bn)	5.1	5.7	3.6	3.7	4.8
Value/intraday volatility	3.0	1.7	1.2	1.7	2.9
Value/interday volatility	1.7	0.4	0.5	0.7	2.2
Volatility					
Intraday volatility	1.7	3.3	2.9	2.1	1.7
Interday volatility	3.0	15.4	7.2	5.2	2.2

[1] Weighted by market capitalisation of the 33 constituent stocks of the Hang Seng Index. [2] January 1997 to 19 October 1997. [3] 20 October 1997 to April 1998. [4] May 1998 to 28 September 1998, but excluding 14 August to 28 August 1998. [5] 29 September 1998 to 29 June 1999. [6] January 2001 to June 2001.

Sources: Bloomberg; HKMA staff estimates.

Table 2

Estimation results for Cheung Kong Holdings

$$\text{Model: } \Delta \ln(P_t) = \alpha + \beta \ln(BSI_t) + \sum_{i=0}^{8} \gamma_i \Delta \ln(BSI_{t-i}) + \sum_{j=1}^{12} \theta_j \Delta \ln(P_{t-j}) + \varepsilon_t$$

(Pre-crisis from 1997:05 to 1997:08, Asian crisis from 1997:10:20 to 1997:11, Russian crisis from 1998:05 to 1998:08:13, post-crisis I from 1998:11 to 1999:03, post-crisis II from 1999:07 to 1999:12, post-crisis III from 2000:06 to 2000:10 and recent period from 2001:01 to 2001:06)

	Pre-crisis	**Asian crisis**	**Russian crisis**	**Post-crisis**			**Recent period**
				I	**II**	**III**	
$\hat{\beta}$	0.9* (6.3)	3.5* (2.4)	1.3* (4.6)	1.4* (4.4)	1.0* (4.8)	1.1* (2.9)	2.2* (6.8)
$\hat{\gamma}_0$	14.2* (27.1)	22.0* (7.7)	9.6* (12.2)	25.7* (30.2)	3.7* (7.2)	10.5* (11.2)	24.4* (25.6)
$\hat{\gamma}_1$	2.2 (0.9)	19.2* (2.3)	9.4* (10.3)	14.0* (11.3)	4.7* (4.3)	4.7* (3.3)	10.8* (6.9)
$\hat{\gamma}_2$	0.5* (2.2)	16.0 (1.8)	7.2* (8.1)	16.6 (11.8)	7.0* (6.3)	2.0 (1.3)	13.5* (7.7)
$\hat{\gamma}_3$	2.1 (0.9)	15.8 (1.9)	5.7* (5.4)	9.3* (7.0)	5.1* (4.0)	1.0 (0.7)	9.9* (5.6)
$\hat{\gamma}_4$	6.6* (3.2)	9.3 (1.0)	5.0* (5.0)	9.0* (7.0)	5.0* (4.3)	2.5 (1.3)	6.2* (3.8)
$\hat{\gamma}_5$	3.7 (1.4)	5.9 (0.6)	3.7* (4.0)	4.3* (3.0)	5.0* (4.9)	–0.8 (–0.4)	1.4 (0.8)
$\hat{\gamma}_6$	5.9* (2.9)	2.5 (0.2)	5.5* (5.3)	5.1* (4.3)	2.6* (2.5)	0.6* (0.3)	2.4 (1.3)
$\hat{\gamma}_7$	0.9 (0.4)	4.0 (0.4)	2.2 (1.9)	3.0* (2.2)	2.9* (2.1)	–0.4 (–0.2)	3.2 (1.5)
$\hat{\gamma}_8$	2.2 (0.9)	4.2 (0.4)	2.0* (2.1)	2.6* (2.1)	1.0 (0.9)	1.3 (0.9)	3.2 (1.6)
$\bar{R}^2$	0.057	0.0099	0.0094	0.028	0.018	0.054	0.020
SSR	0.053	0.15	0.088	0.14	0.094	0.099	0.087
N	38,507	14,083	34,765	48,388	58,681	49,342	55,920

Notes: t-statistics in parentheses.

* Denotes significance at the 5% level. The $\ln(BSI_t)$ and $\Delta\ln(BSI_{t-i})$ variables are divided by 10,000. $\bar{R}^2$ is the adjusted R^2. SSR is the sum of squared residual. N is the number of observations.

Source: HKMA staff estimates.

Table 3

Estimation results for Hang Seng Bank

Model: $\Delta \ln(P_t) = \alpha + \beta \ln(BSI_t) + \sum_{i=0}^{8} \gamma_i \Delta \ln(BSI_{t-i}) + \sum_{j=1}^{12} \theta_j \Delta \ln(P_{t-j}) + \varepsilon_t$

(Pre-crisis from 1997:05 to 1997:08, Asian crisis from 1997:10:20 to 1997:11, Russian crisis from 1998:05 to 1998:08:13, post-crisis I from 1998:11 to 1999:03, post-crisis II from 1999:07 to 1999:12, post-crisis III from 2000:06 to 2000:10 and recent period from 2001:01 to 2001:06)

	Pre-crisis	**Asian crisis**	**Russian crisis**	**Post-crisis**			**Recent period**
				I	**II**	**III**	
$\hat{\beta}$	1.7* (4.7)	3.2* (3.7)	2.0* (5.8)	1.0* (2.0)	0.7* (4.2)	0.8* (4.2)	1.5* (9.3)
$\hat{\gamma}_0$	15.6* (33.9)	0.5 (0.04)	17.6* (17.5)	13.3* (11.9)	3.4* (10.5)	6.3* (10.5)	29.1* (37.2)
$\hat{\gamma}_1$	4.9* (2.9)	7.4* (2.0)	9.4* (6.5)	9.3* (6.2)	3.2* (3.7)	8.3* (9.9)	14.3* (10.3)
$\hat{\gamma}_2$	8.3* (4.5)	4.4 (1.4)	10.5* (7.9)	7.1* (5.3)	2.2* (2.5)	5.3* (5.3)	9.6* (7.4)
$\hat{\gamma}_3$	6.3* (3.9)	7.7* (2.4)	10.7* (6.7)	2.5 (1.3)	3.3* (4.0)	7.1* (7.8)	9.9* (6.3)
$\hat{\gamma}_4$	8.5* (4.4)	4.0 (1.4)	6.4* (4.5)	4.7* (2.7)	2.4* (2.3)	2.0 (1.7)	7.0* (4.9)
$\hat{\gamma}_5$	4.8* (2.5)	10.7* (4.1)	9.6* (6.8)	5.0* (2.5)	0.0 (0.03)	1.8 (1.9)	4.8* (3.4)
$\hat{\gamma}_6$	5.6* (2.7)	7.3* (2.8)	10.9* (6.8)	4.0* (2.3)	3.2* (3.1)	3.8* (4.0)	4.6* (3.0)
$\hat{\gamma}_7$	3.2 (1.4)	2.5 (1.0)	7.9* (4.7)	1.3 (0.7)	2.0* (2.1)	3.2* (3.1)	6.9* (5.3)
$\hat{\gamma}_8$	3.1 (1.5)	4.8 (1.8)	3.1 (1.7)	–0.3 (–0.2)	0.9 (1.0)	2.4* (2.2)	2.9 (1.8)
$\overline{R}^2$	0.038	0.0091	0.0019	0.013	0.014	0.024	0.033
SSR	0.095	0.16	0.082	0.10	0.065	0.067	0.094
N	38,526	14,071	31,144	48,218	58,729	47,807	56,381

Notes: t-statistics in parentheses.

* denotes significance at the 5% level. The $\ln(BSI_t)$ and $\Delta\ln(BSI_{t-i})$ variables are divided by 10,000. $\overline{R}^2$ is the adjusted R^2. SSR is the sum of squared residual. N is the number of observations.

Source: HKMA staff estimates.

Table 4

Estimation results for Hutchison Whampoa Limited

$$\text{Model: } \Delta\ln(P_t) = \alpha + \beta\ln(BSI_t) + \sum_{i=0}^{8}\gamma_i\Delta\ln(BSI_{t-i}) + \sum_{j=1}^{12}\theta_j\Delta\ln(P_{t-j}) + \varepsilon_t$$

(Pre-crisis from 1997:05 to 1997:08, Asian crisis from 1997:10:20 to 1997:11, Russian crisis from 1998:05 to 1998:08:13, post-crisis I from 1998:11 to 1999:03, post-crisis II from 1999:07 to 1999:12, post-crisis III from 2000:06 to 2000:10 and recent period from 2001:01 to 2001:06)

	Pre-crisis	**Asian crisis**	**Russian crisis**	**Post-crisis**			**Recent period**
				I	**II**	**III**	
$\hat{\beta}$	1.5* (2.5)	11.2* (9.4)	2.1* (4.1)	1.5* (4.4)	0.6* (2.3)	1.8* (3.5)	1.6* (7.0)
$\hat{\gamma}_0$	31.0* (24.1)	–3.7* (–2.0)	10.3* (11.2)	31.5* (48.2)	–3.6* (–8.4)	24.7* (15.5)	31.9* (50.3)
$\hat{\gamma}_1$	18.6* (6.0)	7.8 (1.1)	9.7* (7.7)	19.2* (11.2)	10.7* (10.3)	18.4* (6.6)	10.6* (8.8)
$\hat{\gamma}_2$	17.8* (6.0)	16.0* (2.4)	10.7* (9.1)	14.9* (8.8)	7.0* (6.4)	31.8* (12.5)	9.2* (7.9)
$\hat{\gamma}_3$	8.3* (2.6)	6.8 (0.9)	7.5* (5.9)	11.2* (5.8)	5.2* (4.2)	24.6* (9.4)	0.9 (0.8)
$\hat{\gamma}_4$	15.5* (5.4)	–2.8 (–0.44)	8.1* (6.8)	12.0* (7.2)	2.4* (2.2)	29.4* (11.5)	8.1* (7.3)
$\hat{\gamma}_5$	9.2* (3.2)	–0.5 (–0.08)	9.6* (8.3)	8.6* (4.3)	4.8* (3.6)	6.0* (2.2)	9.3* (6.7)
$\hat{\gamma}_6$	12.0* (4.4)	–0.7 (–0.1)	5.4* (4.7)	6.6* (3.2)	4.4* (3.5)	16.2* (5.5)	11.7* (8.7)
$\hat{\gamma}_7$	7.3* (2.4)	1.3 (0.2)	4.9* (4.1)	10.1* (5.6)	2.3 (1.8)	15.4* (5.8)	9.9* (7.2)
$\hat{\gamma}_8$	8.2* (2.5)	–4.4 (–0.7)	–0.6 (–0.5)	9.7* (5.3)	3.8* (2.6)	6.9* (2.3)	6.5* (4.5)
$\bar{R}^2$	0.075	0.0036	0.0020	0.014	0.025	0.077	0.044
SSR	0.078	0.14	0.090	0.13	0.11	0.14	0.098
N	38,517	14,077	34,760	48,386	58,723	49,316	56,379

Notes: t-statistics in parentheses.

* denotes significance at the 5% level. The $\ln(BSI_t)$ and $\Delta\ln(BSI_{t-i})$ variables are divided by 10,000. $\bar{R}^2$ is the adjusted R^2. SSR is the Sum of Squared Residual. N is the number of observations.

Source: HKMA staff estimates.

Table 5

Estimation results for Sun Hung Kai Properties

Model: $\Delta \ln(P_t) = \alpha + \beta \ln(BSI_t) + \sum_{i=0}^{8} \gamma_i \Delta \ln(BSI_{t-i}) + \sum_{j=1}^{12} \theta_j \Delta \ln(P_{t-j}) + \varepsilon_t$

(Pre-crisis from 1997:05 to 1997:08, Asian crisis from 1997:10:20 to 1997:11, Russian crisis from 1998:05 to 1998:08:13, post-crisis I from 1998:11 to 1999:03, post-crisis II from 1999:07 to 1999:12, post-crisis III from 2000:06 to 2000:10 and recent period from 2001:01 to 2001:06)

	Pre-crisis	**Asian crisis**	**Russian crisis**	**Post-crisis**			**Recent period**
				I	**II**	**III**	
$\hat{\beta}$	1.6* (4.6)	4.5* (2.6)	2.6* (6.4)	1.2* (4.6)	2.7* (10.8)	1.5* (6.7)	2.9* (10.5)
$\hat{\gamma}_0$	13.2* (16.8)	8.0* (10.6)	14.6* (15.0)	13.2* (19.2)	14.4* (28.7)	24.2* (38.6)	15.9* (44.3)
$\hat{\gamma}_1$	5.2* (3.1)	5.6 (1.4)	8.9* (7.4)	9.5* (10.6)	12.6* (12.2)	18.4* (23.7)	20.7* (19.5)
$\hat{\gamma}_2$	6.9* (4.7)	10.9* (2.6)	4.9* (3.2)	11.3* (11.7)	12.3* (8.0)	14.9* (19.3)	16.9* (12.8)
$\hat{\gamma}_3$	4.9* (3.3)	1.0 (0.2)	5.4* (3.3)	6.2* (6.0)	6.7* (4.3)	5.9* (5.2)	14.4* (11.4)
$\hat{\gamma}_4$	5.3* (3.7)	3.4 (0.7)	8.1* (5.1)	–1.9 (–1.8)	4.3* (2.4)	4.0* (3.4)	9.7* (8.4)
$\hat{\gamma}_5$	3.1* (2.0)	3.5 (0.7)	6.3* (4.0)	2.1* (2.1)	5.9* (3.8)	1.2 (1.0)	6.6* (5.3)
$\hat{\gamma}_6$	4.1* (2.5)	–1.8 (–0.2)	0.4 (0.3)	–0.4 (–0.3)	3.9* (2.1)	2.5* (2.2)	4.9* (3.5)
$\hat{\gamma}_7$	2.4 (1.4)	–3.3 (–0.5)	4.5* (2.9)	1.4 (1.5)	0.4 (0.2)	3.1* (2.5)	7.9* (5.4)
$\hat{\gamma}_8$	5.0 (3.4)	–3.0 (–0.4)	–2.3 (–1.4)	–4.3* (–6.1)	5.2* (2.9)	1.6 (1.4)	2.6 (1.9)
$\overline{R}^2$	0.020	0.0076	0.0034	0.016	0.017	0.0052	0.012
SSR	0.058	0.14	0.083	0.12	0.11	0.12	0.10
N	38,524	14,075	34,751	48,203	58,718	49,325	55,429

Notes: t-statistics in parentheses.

* denotes significance at the 5% level. The $\ln(BSI_t)$ and $\Delta\ln(BSI_{t-i})$ variables are divided by 10,000. $\overline{R}^2$ is the adjusted R^2. SSR is the sum of squared residual. *N* is the number of observations.

Source: HKMA staff estimates.

Table 6

Estimation results for Bank of East Asia

Model: $\Delta\ln(P_t) = \alpha + \beta\ln(BSI_t) + \sum_{i=0}^{8}\gamma_i\Delta\ln(BSI_{t-i}) + \sum_{j=1}^{12}\theta_j\Delta\ln(P_{t-j}) + \varepsilon_t$

(Pre-crisis from 1997:05 to 1997:08, Asian crisis from 1997:10:20 to 1997:11, Russian crisis from 1998:05 to 1998:08:13, post-crisis I from 1998:11 to 1999:03, post-crisis II from 1999:07 to 1999:12, post-crisis III from 2000:06 to 2000:10 and recent period from 2001:01 to 2001:06)

	Pre-crisis	**Asian crisis**	**Russian crisis**	**Post-crisis**			**Recent period**
				I	**II**	**III**	
$\hat{\beta}$	2.3* (6.3)	12.5* (3.5)	2.9* (7.8)	1.0* (3.2)	1.2* (4.5)	1.0* (5.7)	1.5* (6.6)
$\hat{\gamma}_0$	14.2* (27.1)	–2.0 (–0.8)	39.5* (29.3)	6.0* (5.1)	8.3* (13.7)	3.0* (5.4)	17.6* (20.0)
$\hat{\gamma}_1$	2.2 (0.9)	–2.3 (–0.2)	15.7* (6.8)	2.0 (1.2)	3.3* (2.3)	1.5 (1.9)	13.4* (10.5)
$\hat{\gamma}_2$	0.5* (2.2)	14.9 (1.5)	8.7* (3.6)	1.4 (0.8)	4.1* (3.1)	1.6 (1.9)	12.7* (9.1)
$\hat{\gamma}_3$	2.1 (0.9)	17.6 (1.7)	2.2 (0.9)	1.6 (1.0)	7.8* (7.1)	1.0 (1.2)	12.9* (9.3)
$\hat{\gamma}_4$	6.6* (3.2)	–0.4 (–0.04)	16.1* (7.4)	6.1* (3.5)	2.4* (2.0)	2.0* (2.2)	7.3* (4.9)
$\hat{\gamma}_5$	3.7 (1.4)	–9.5 (–0.8)	21.0* (9.6)	5.5* (3.2)	4.0* (3.2)	0.3 (0.4)	9.6* (6.5)
$\hat{\gamma}_6$	5.9* (2.9)	4.9 (0.4)	–10.4* (–4.8)	4.0* (2.3)	1.3 (0.9)	1.7* (2.1)	2.5* (2.1)
$\hat{\gamma}_7$	0.9 (0.4)	–9.9 (–1.0)	–1.9 (–1.0)	4.9* (3.0)	–3.4* (–2.6)	0.4 (0.5)	7.5* (5.5)
$\hat{\gamma}_8$	2.2 (0.9)	–5.4 (–0.5)	14.2* (6.3)	2.1 (1.3)	2.2 (1.6)	2.5* (3.1)	1.1 (0.8)
$\overline{R}^2$	0.057	0.0088	0.00059	0.0040	0.0097	0.024	0.021
SSR	0.053	0.080	0.14	0.13	0.091	0.087	0.10
N	38,507	14,064	34,767	48,369	58,706	49,319	56,377

Notes: t-statistics in parentheses.

* denotes significance at the 5% level. The $\ln(BSI_t)$ and $\Delta\ln(BSI_{t-i})$ variables are divided by 10,000. $\overline{R}^2$ is the adjusted R^2. SSR is the sum of squared residual. N is the number of observations.

Source: HKMA staff estimates.

Table 7

Determinants of market liquidity

$$\beta_t = f(I_t, ID_t, VHK_t, VUSA_t, MLUSA_t, D_{1t}, D_{2t})$$

	Regression no	Constant	I_t (x 10^{-4})	VHK_t (x 10^{-4})	$VUSA_t$ (x 10^{-4})	$MLUSA_t$ (x 10^{-4})	D_{1t} (x 10^{-4})	D_{2t} (x 10^{-4})	$\overline{R}^2$	N
Cheung Kong Holdings	1	–0.0002 (–1.3)	–0.4 (–1.6)	2.7* (4.3)	1.1 (1.7)	–	–	–	0.72	32
	2	0.00001 (0.1)	–0.6* (–2.4)	3.3* (6.0)	–	–	–	–	0.70	32
	3	–0.0006* (–2.1)	0.6* (2.4)	–	2.7* (2.8)	–	–	–	0.56	32
	4	–0.000006 (–0.0)	–0.6* (–2.0)	2.8* (4.6)	1.1 (1.7)	–0.1 (–1.8)	–	–	0.75	32
	5	0.0002 (1.6)	–0.8* (–3.0)	3.3* (5.2)	–	–0.1* (–2.2)	–	–	0.73	32
	6	–0.0004 (–1.4)	0.6* (2.0)	–	2.7* (2.8)	–0.1 (–1.5)	–	–	0.56	32
	7	0.0001 (0.9)	–0.2 (–0.7)	0.5 (0.7)	0.7 1.1)	–	11.7* (3.4)	0.8 (0.7)	0.81	32
	8	0.0003* (2.5)	–0.3 (–1.1)	0.7 (0.9)	–	–	12.7* (3.9)	1.1 (0.9)	0.80	32
	9	0.0002 (1.0)	–0.05 (–0.3)	–	0.7 (1.3)	–	13.4* (5.7)	1.0 (0.9)	0.81	32
	10	0.0003 (1.8)	–0.3 (–1.1)	0.6 (0.9)	0.6 (1.1)	–0.1* (–2.2)	11.4* (3.6)	0.6 (0.5)	0.83	32
	11	0.0005* (3.5)	–0.4 (–1.6)	0.8 (1.1)	–	–0.1* (–2.3)	12.4* (4.0)	0.9 (0.8)	0.83	32
	12	0.0003 (1.9)	–0.1 (–0.8)	–	0.7 (1.3)	–0.1* (–2.2)	13.4* (6.1)	0.9 (0.8)	0.83	32

Notes: t-statistics in parentheses.

* denotes significance at the 5% level. – denotes corresponding variable not included in the respective model. Estimation period as specified in Section 3.2.3 of the paper. Standard errors are obtained by the heteroscedasticity consistent estimator of White (1980) when necessary. Data are monthly averages. $\overline{R}^2$ is the adjusted R^2. N is the number of observations.

Source: HKMA staff estimates.

Table 7 (cont)

	Regression no	Constant	I_t (x 10^{-4})	VHK_t (x 10^{-4})	$VUSA_t$ (x 10^{-4})	$MLUSA_t$ (x 10^{-4})	D_{1t} (x 10^{-4})	D_{2t} (x 10^{-4})	$\overline{R}^2$	N
Hang Seng Bank	1	–0.0002	–0.05	1.4*	0.6	–	–	–	0.69	32
		(–1.8)	(–0.2)	(3.2)	(1.3)					
	2	–0.00008	–0.2	1.8*	–	–	–	–	0.68	32
		(–1.3)	(–0.8)	(4.5)						
	3	–0.0004*	0.5*	–	1.4*	–	–	–	0.59	32
		(–2.6)	(3.4)		(2.5)					
	4	–0.00002	–0.2	1.5*	0.6	–0.1*	–	–	0.74	32
		(–0.2)	(–0.9)	(3.6)	(1.3)	(–2.5)				
	5	0.00008	–0.3	1.8*	–	–0.1*	–	–	0.73	32
		(0.9)	(–1.5)	(5.0)		(–2.5)				
	6	–0.0003	0.5*	–	1.4*	–0.1*	–	–	0.62	32
		(–1.4)	(2.8)		(2.5)	(–2.5)				
	7	0.0001	–0.01	–0.04	0.2	–	9.3*	2.2*	0.78	32
		(1.1)	(–0.1)	(–0.1)	(0.4)		(3.7)	(2.5)		
	8	0.0002*	–0.04	0.003	–	–	9.6*	2.3*	0.79	32
		(2.1)	(–0.2)	(0.0)			(4.1)	(2.6)		
	9	0.0001	–0.02	–	0.2	–	9.2*	2.2*	0.79	32
		(1.1)	(–0.1)		(0.5)		(5.4)	(2.6)		
	10	0.0003*	–0.1	0.05	0.1	–0.1*	9.0*	2.0*	0.83	32
		(2.3)	(–0.6)	(0.1)	(0.4)	(–2.9)	(4.1)	(2.6)		
	11	0.0003*	–0.1	0.08	–	–0.1*	9.3*	2.1*	0.84	32
		(3.6)	(–0.8)	(0.2)		(–3.0)	(4.5)	(2.7)		
	12	0.0003*	–0.1	–	0.2	–0.1*	9.2*	2.0*	0.84	32
		(2.4)	(–0.8)		(0.4)	(–3.0)	(6.2)	(2.7)		

Notes: t-statistics in parentheses.

* denotes significance at the 5% level. – denotes corresponding variable not included in the respective model. Estimation period as specified in Section 3.2.3 of the paper. Standard errors are obtained by the heteroscedasticity consistent estimator of White (1980) when necessary. Data are monthly averages. $\overline{R}^2$ is the adjusted R^2. N is the number of observations.

Source: HKMA staff estimates.

Table 7 (cont)

	Regression no	Constant	I_t (x 10^{-4})	VHK_t (x 10^{-4})	$VUSA_t$ (x 10^{-4})	$MLUSA_t$ (x 10^{-4})	D_{1t} (x 10^{-4})	D_{2t} (x 10^{-4})	$\overline{R}^2$	N
Hutchison Whampoa	1	–0.0005* (–3.3)	1.39 (1.0)	2.0* (3.1)	1.1 (1.6)	–	–	–	0.80	32
	2	–0.0003* (–3.1)	0.1 (0.4)	2.6* (4.1)	–	–	–	–	0.79	32
	3	–0.0008* (–3.3)	1.1* (4.7)	–	2.0* (2.7)	–	–	–	0.74	32
	4	–0.0006* (–3.0)	0.3 (1.1)	2.0* (3.0)	1.1 (1.6)	0.05 (0.5)	–	–	0.79	32
	5	–0.0004* (–2.5)	0.1 (0.5)	2.6* (4.5)	–	0.04 (0.5)	–	–	0.78	32
	6	–0.0009* (–3.2)	1.2* (4.7)	–	2.2* (2.7)	0.06 (0.8)	–	–	0.73	32
	7	–0.0001 (–0.7)	0.5 (2.0)	–0.2 (–0.2)	0.6 (0.9)	–	12.3* (3.4)	1.2 (0.9)	0.86	32
	8	–0.000009 (–0.1)	0.4 (1.8)	–0.05 (–0.1)	–	–	13.2* (3.8)	1.4 (1.1)	0.86	32
	9	–0.0001 (–0.8)	0.5* (2.6)	–	0.5 (0.9)	–	11.7* (4.8)	1.1 (0.9)	0.86	32
	10	–0.0002 (–1.0)	0.6 (2.0)	–0.2 (–0.3)	0.6 (0.9)	0.05 (0.7)	12.4* (3.4)	1.3 (1.0)	0.85	32
	11	–0.00007 (–0.4)	0.5 (1.8)	–0.09 (1.8)	–	0.05 (0.7)	13.3* (3.8)	1.5 (1.2)	0.85	32
	12	–0.0002 (–1.0)	0.5* (2.7)	–	0.5 (0.9)	0.05 (0.7)	11.7* (4.8)	1.2 (1.0)	0.86	32

Notes: t-statistics in parentheses.

* denotes significance at the 5% level. – denotes corresponding variable not included in the respective model. Estimation period as specified in Section 3.2.3 of the paper. Standard errors are obtained by the heteroscedasticity consistent estimator of White (1980) when necessary. Data are monthly averages. $\overline{R}^2$ is the adjusted R^2. N is the number of observations.

Source: HKMA staff estimates.

Table 7 (cont)

	Regression no	Constant	I_t (x 10^{-4})	VHK_t (x 10^{-4})	$VUSA_t$ (x 10^{-4})	$MLUSA_t$ (x 10^{-4})	D_{1t} (x 10^{-4})	D_{2t} (x 10^{-4})	$\overline{R}^2$	N
Sun Hung Kai Properties	1	–0.0002	0.3	0.06	1.9	–	–	–	0.40	32
		(–1.0)	(0.5)	(0.1)	(2.5)					
	2	0.00008	–0.03	1.0	–	–	–	–	0.26	32
		(0.5)	(–0.1)	(1.2)						
	3	–0.0003*	0.3*	–	1.9*	–	–	–	0.42	32
		(–2.3)	(2.8)		(3.5)					
	4	–0.0002	0.3	0.07	1.8*	–0.03	–	–	0.38	32
		(–0.8)	(0.5)	(0.1)	(2.5)	(–0.7)				
	5	0.0001	–0.06	1.0	–	–0.04	–	–	0.24	32
		(0.8)	(–0.1)	(1.2)		(–0.8)				
	6	–0.0002	0.3*	–	1.9*	–0.03	–	–	0.41	32
		(–1.4)	(2.3)		(3.4)	(–0.4)				
	7	0.00003	–0.06	–0.1	1.6*	–	4.0*	4.3*	0.58	32
		(0.2)	(–0.2)	(–0.1)	(2.8)		(2.2)	(3.2)		
	8	0.0004*	–0.3	0.2	–	–	6.5*	4.9*	0.47	32
		(3.8)	(–1.0)	(0.3)			(2.5)	(2.3)		
	9	0.00003	–0.08	–	1.6*	–	3.7	4.2*	0.59	32
		(0.2)	(–0.4)		(2.8)		(1.6)	(3.7)		
	10	0.00005	–0.07	–0.1	1.6*	–0.01	4.0	4.3*	0.56	32
		(0.3)	(–0.3)	(–0.1)	(2.7)	(–0.2)	(1.1)	(3.4)		
	11	0.0004*	–0.3	0.3	–	–0.02	6.4*	4.9*	0.45	32
		(3.3)	(–1.1)	(0.3)		(–0.4)	(2.4)	(2.2)		
	12	0.00005	–0.1	–	1.6*	–0.02*	3.7	4.2*	0.58	32
		(0.2)	(–0.5)		(2.7)	(–0.4)	(1.8)	(2.5)		

Notes: t-statistics in parentheses.

* denotes significance at the 5% level. – denotes corresponding variable not included in the respective model. Estimation period as specified in Section 3.2.3 of the paper. Standard errors are obtained by the heteroscedasticity consistent estimator of White (1980) when necessary. Data are monthly averages. $\overline{R}^2$ is the adjusted R^2. N is the number of observations.

Source: HKMA staff estimates.

Table 7 (cont)

	Regression no	Constant	I_t (x 10^{-4})	VHK_t (x 10^{-4})	$VUSA_t$ (x 10^{-4})	$MLUSA_t$ (x 10^{-4})	D_{1t} (x 10^{-4})	D_{2t} (x 10^{-4})	$\overline{R}^2$	N
Bank of East Asia	1	–0.0001 (–1.0)	0.1 (0.4)	0.9 (1.4)	0.8 (1.2)	–	–	–	0.42	32
	2	–0.000007 (–0.1)	–0.03 (–0.1)	1.3* (2.4)	–	–	–	–	0.42	32
	3	–0.0003* (–2.3)	0.5* (3.4)	–	1.3 (1.7)	–	–	–	0.41	32
	4	0.0001 (0.8)	–0.08 (–0.3)	1.0 (1.7)	0.7 (1.2)	–0.2* (–3.0)	–	–	0.55	32
	5	0.0003* (2.2)	–0.2 (–0.9)	1.4* (2.8)	–	–0.2* (–3.0)	–	–	0.54	32
	6	–0.00002 (–0.2)	0.3* (2.3)	–	1.3 (1.9)	–0.2* (–3.6)	–	–	0.52	32
	7	–0.0002 (–0.7)	0.3 (0.9)	0.4 (0.4)	0.8 (1.1)	–	1.4 (0.3)	–1.2 (–0.8)	0.41	32
	8	0.00001 (0.1)	0.2 (0.5)	0.6 (0.6)	–	–	2.6 (0.6)	–0.9 (–0.6)	0.41	32
	9	–0.0001 (–0.7)	0.4 (1.6)	–	0.8 (0.9)	–	2.8 (1.0)	–1.1 (–0.9)	0.43	32
	10	0.0001 (0.5)	0.1 (0.4)	0.6 (0.7)	0.7 (1.1)	–0.2* (–3.1)	1.0 (0.3)	–1.5 (–1.2)	0.56	32
	11	0.0003 (1.8)	–0.0005 (–0.0)	0.7 (0.9)	–	–0.2* (–3.1)	2.1 (0.6)	–1.3 (–1.0)	0.55	32
	12	0.0001 (0.6)	0.3 (1.3)	–	0.8 (1.3)	–0.2* (–3.1)	2.8 (1.1)	–1.3 (–1.0)	0.57	32

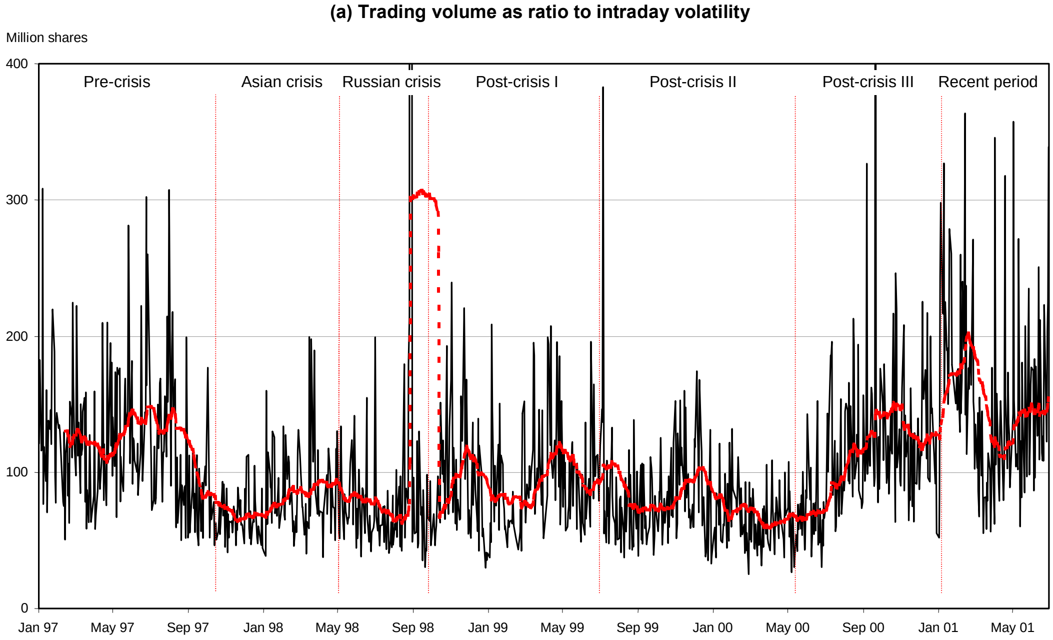
Notes: t-statistics in parentheses.

* denotes significance at the 5% level. – denotes corresponding variable not included in the respective model. Estimation period as specified in Section 3.2.3 of the paper. Standard errors are obtained by the heteroscedasticity consistent estimator of White (1980) when necessary. Data are monthly averages. $\overline{R}^2$ is the adjusted R^2. N is the number of observations.

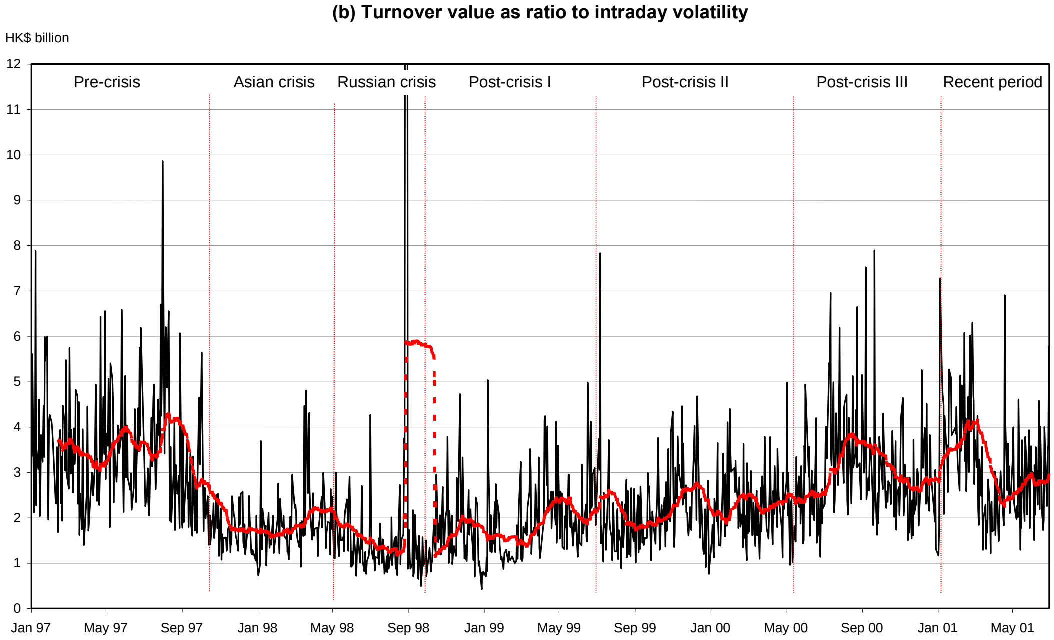
Source: HKMA staff estimates.

Chart 1

Market depth indicators of the Hong Kong stock market



Note: ----- line is a 30-day moving average.

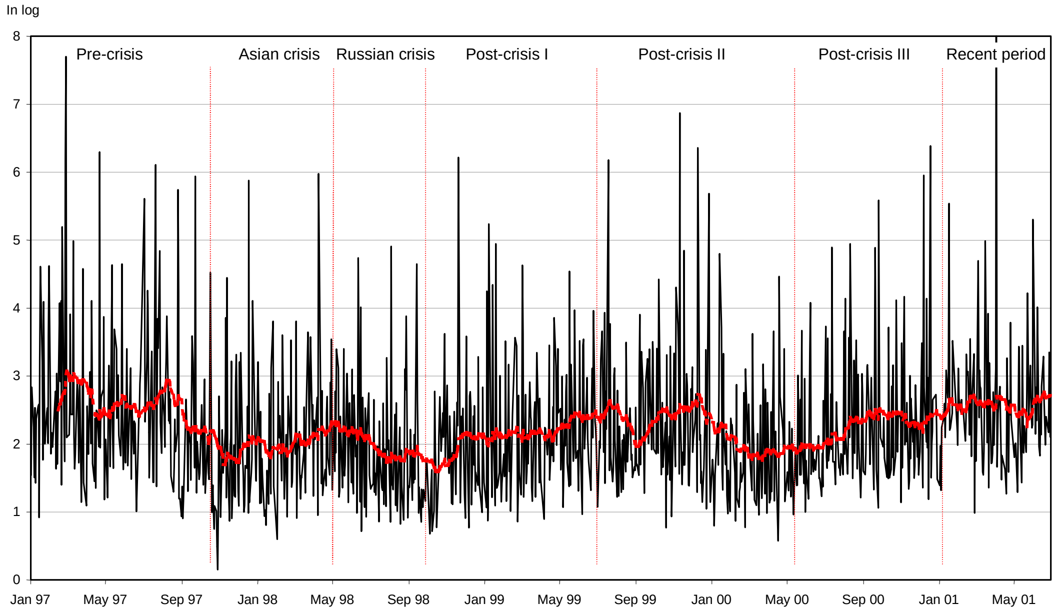


Note: ----- line is a 30-day moving average.

Chart 2

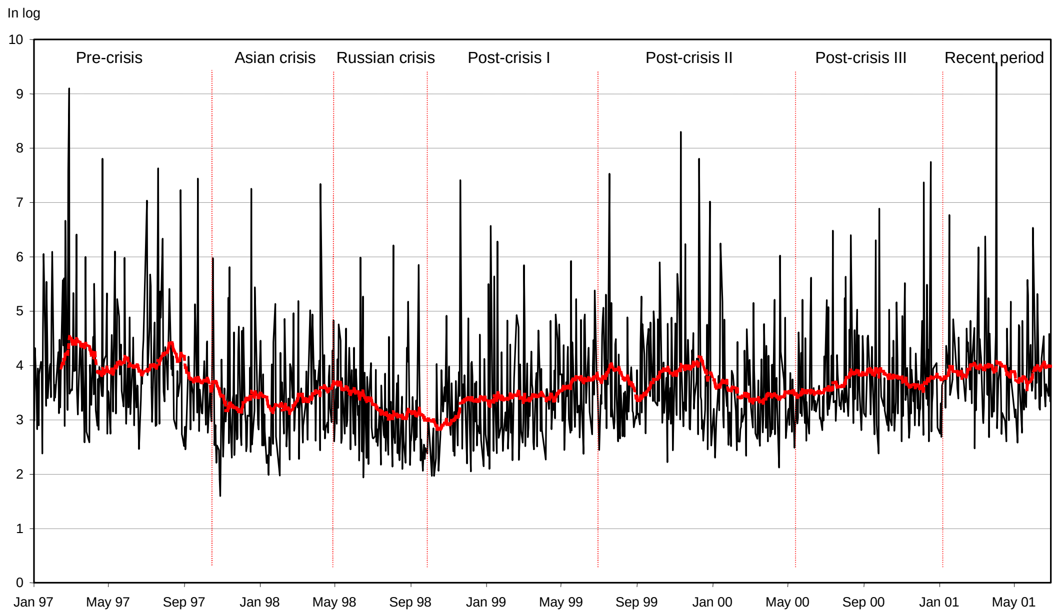
Market depth indicators of the Hong Kong stock market

(a) Trading volume as ratio to interday volatility



Note: ----- line is a 30-day moving average.

(b) Turnover value as ratio to interday volatility

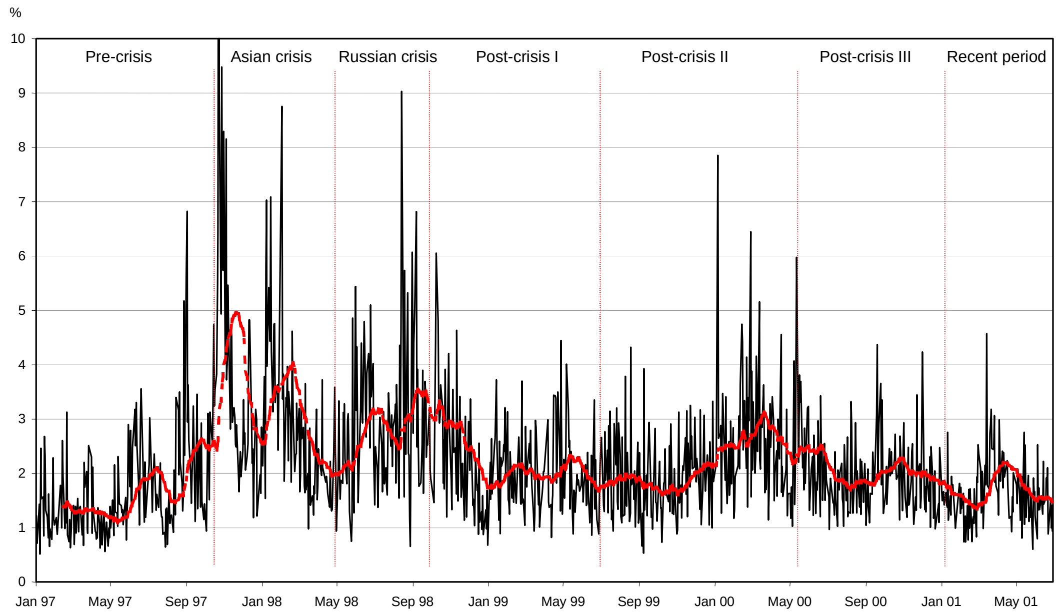


Note: ----- line is a 30-day moving average.

Chart 3

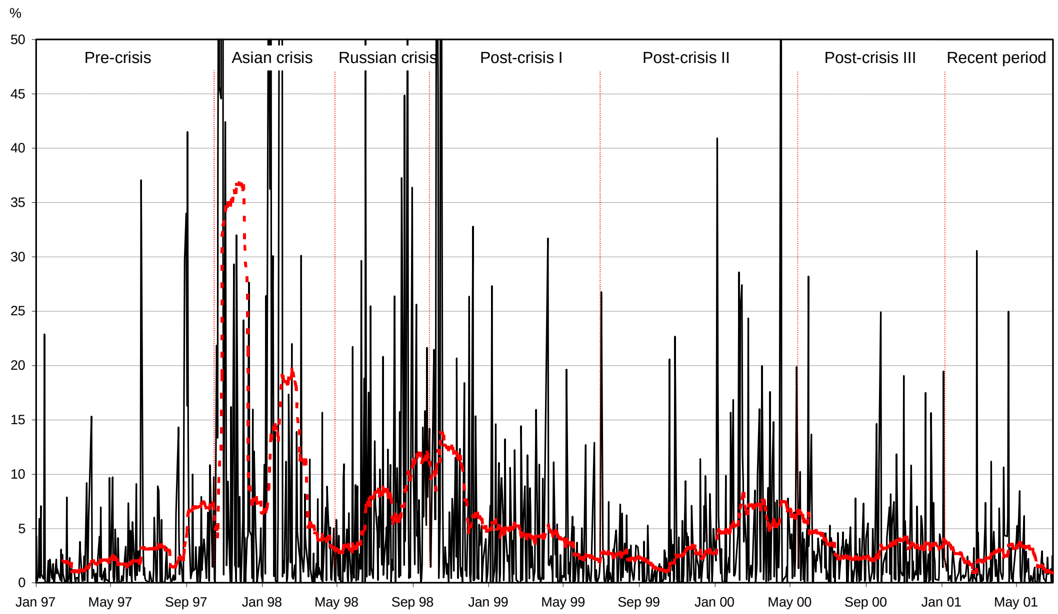
Price volatility of the Hong Kong stock market

(a) Intraday price volatility



Note: ----- line is a 30-day moving average.

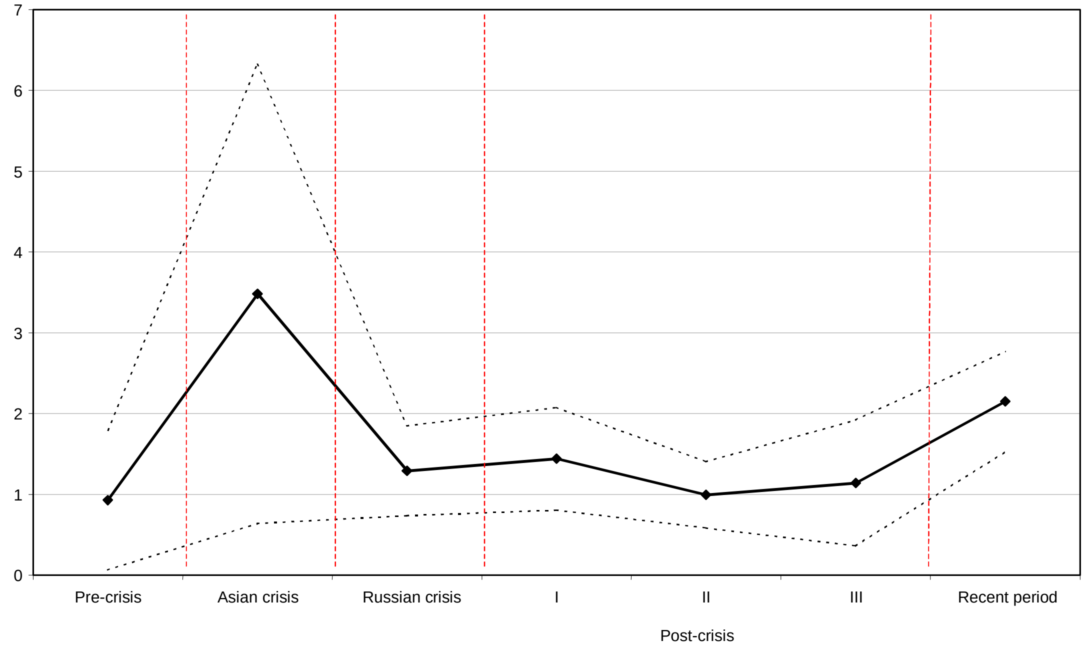
(b) Interday price volatility



Note: ----- line is a 30-day moving average.

Chart 4

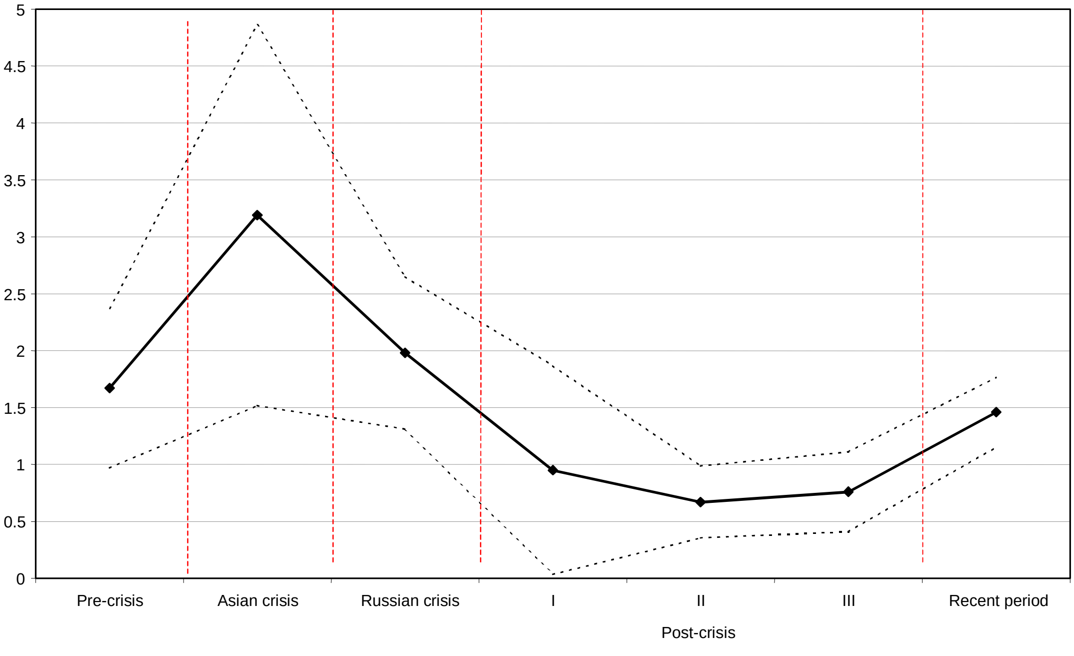
Estimated β coefficient for Cheung Kong Holdings



Note: ----- lines represent confidence interval of 95%.

Chart 5

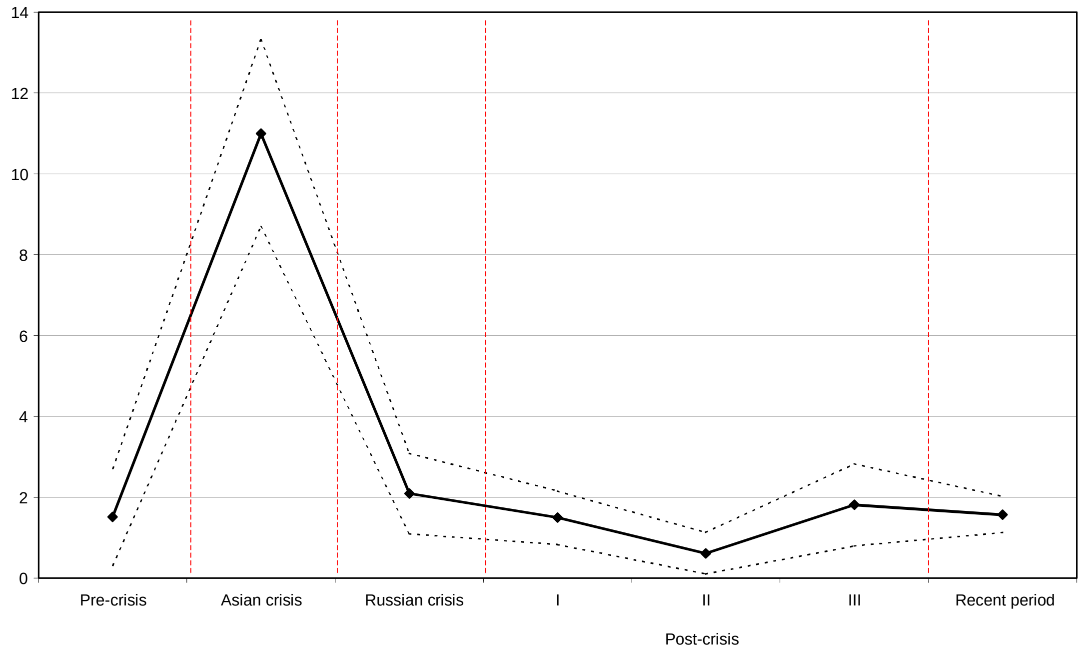
Estimated β coefficient for Hang Seng Bank



Note: ----- lines represent confidence interval of 95%.

Chart 6

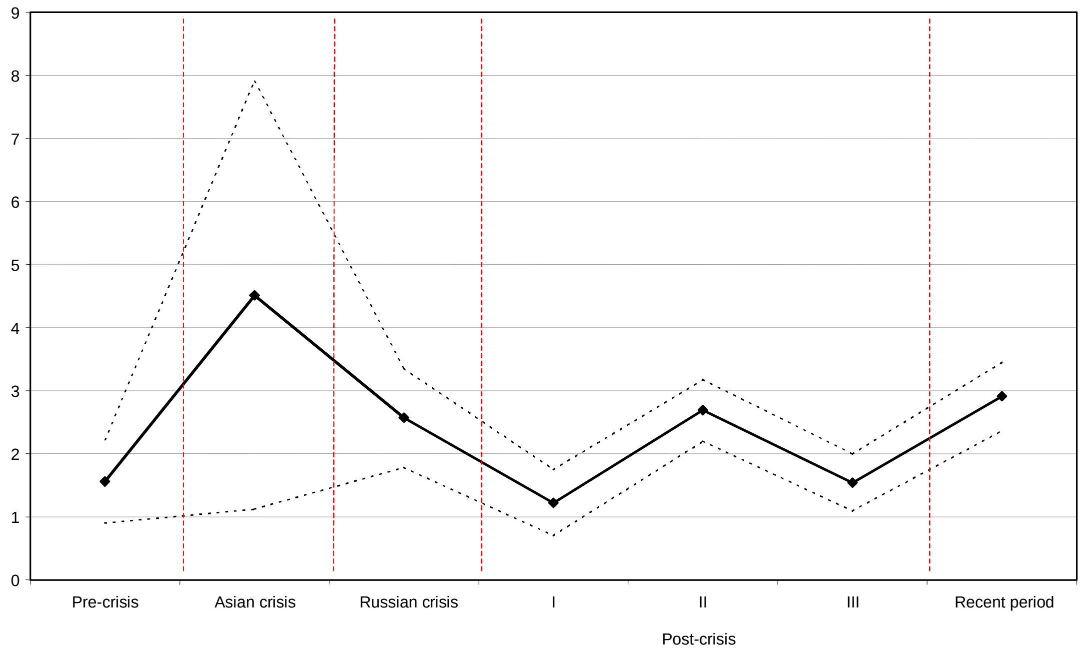
Estimated β coefficient for Hutchison Whampoa Limited



Note: ----- lines represent confidence interval of 95%.

Chart 7

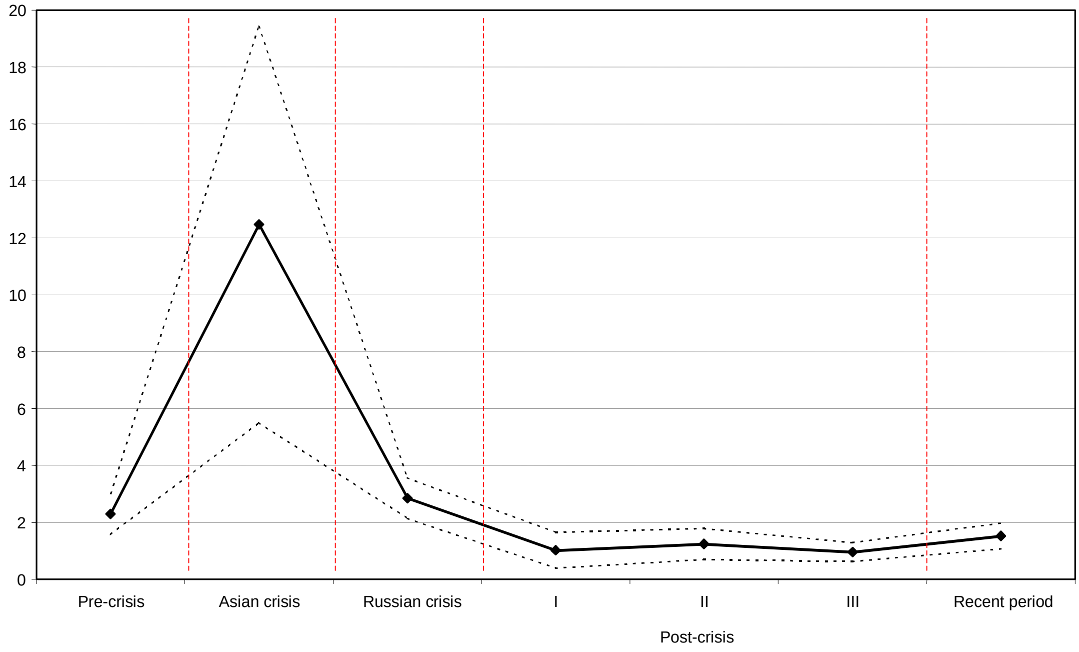
Estimated β coefficient for Sun Hung Kai Properties



Note: ----- lines represent confidence interval of 95%.

Chart 8

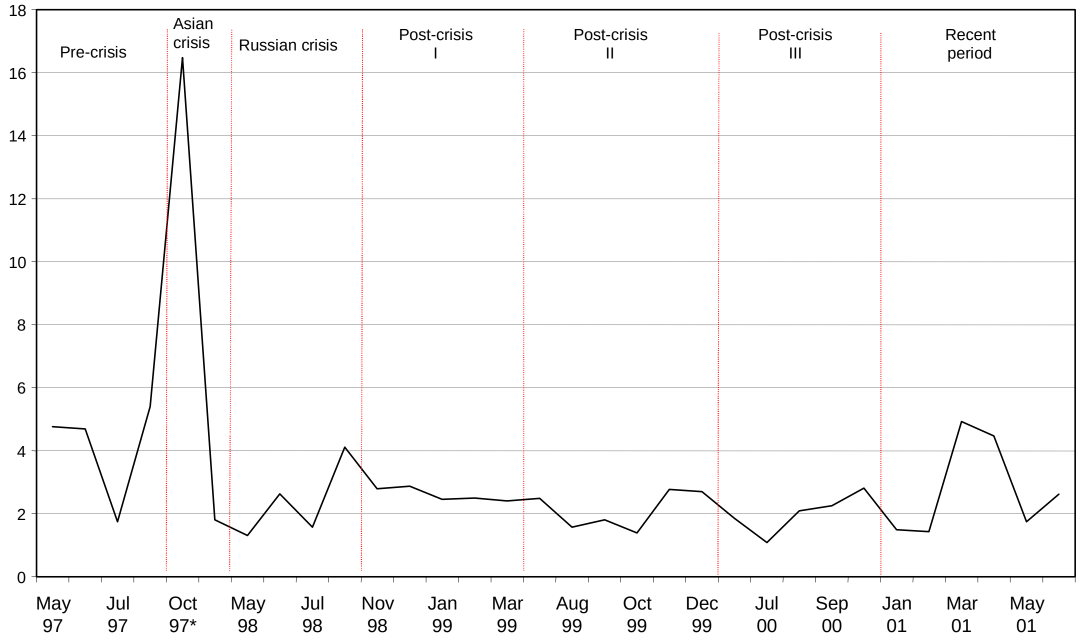
Estimated β coefficient for Bank of East Asia



Note: ----- lines represent confidence interval of 95%.

Chart 9

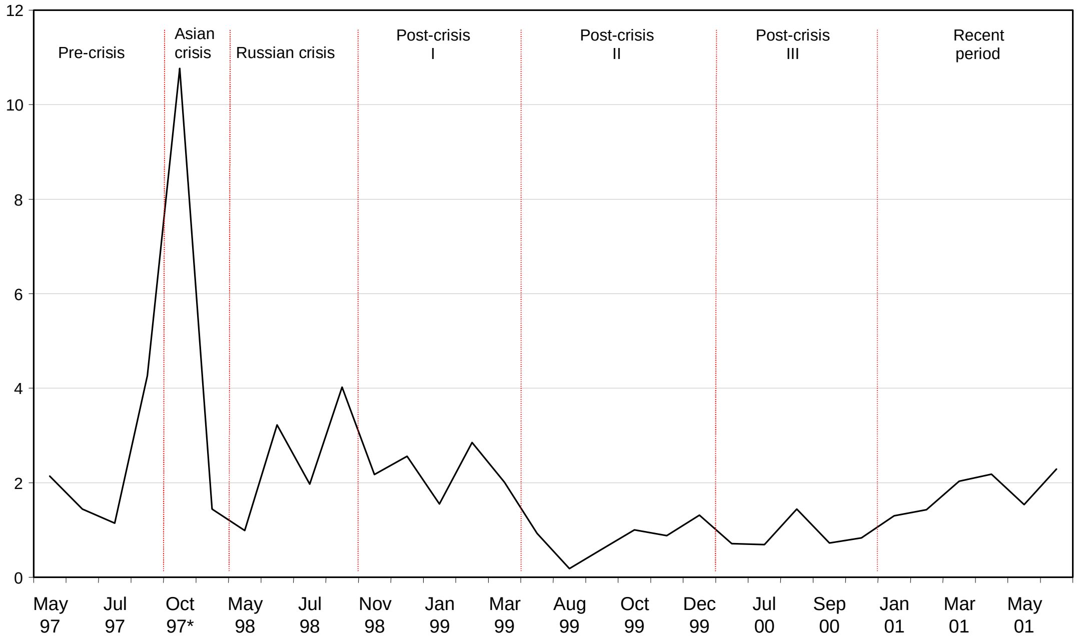
Monthly movement of estimated β coefficient for Cheung Kong Holdings



Note: All estimated βs are significant at the 5% level.

Chart 10

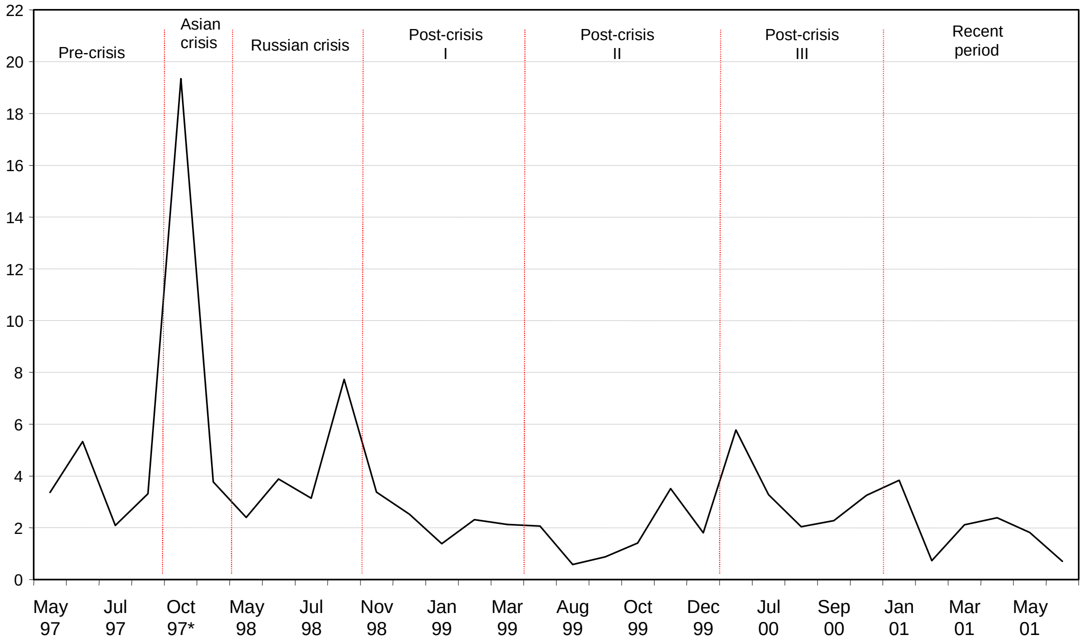
Monthly movement of estimated β coefficient for Hang Seng Bank



Note: All estimated βs are significant at the 5% level.

Chart 11

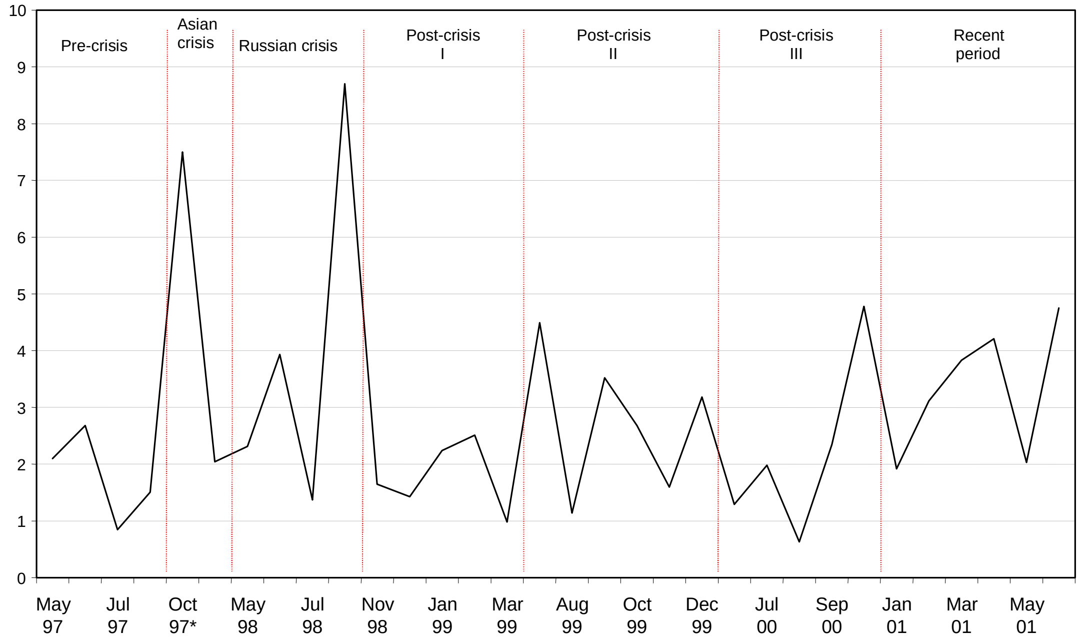
Monthly movement of estimated β coefficient for Hutchison Whampoa Ltd



Note: All estimated βs are significant at the 5% level.

Chart 12

Monthly movement of estimated β coefficient for Sun Hung Kai Properties



Note: All estimated βs are significant at the 5% level.

Chart 13

Monthly movement of estimated β coefficient for Bank of East Asia

Note: All estimated βs are significant at the 5% level.

References

Amihud, Y and H Mendelson (1980): "Dealership market: market-making with inventory", *Journal of Financial Economics* 8, 31-53.

Ahn, H J and Y L Cheung (1999): "The intraday patterns of the spread and depth in a market without market makers: the stock exchange of Hong Kong", *Pacific-Basin Finance Journal* 7, 539-56.

Ahn, H J, K H Bae and K Chan (2000): "Limited orders, depth, and volatility: evidence from the stock exchange of Hong Kong", *Journal of Finance*, forthcoming.

Bank for International Settlements (1999a): *Market liquidity: research findings and selected policy implications*, report of a study group established by the Committee on the Global Financial System of the central banks of the Group of Ten countries, May.

——— (1999b): *A review of financial market events in autumn 1998*, Committee on the Global Financial System, October.

——— (2001): *Structural aspects of market liquidity from a financial stability perspective*, discussion note prepared by the Committee on the Global Financial System for the March 2001 meeting of the Financial Stability Forum (FSF), June.

Brockman, P and D Y Chung (1998): "Inter- and intra-day liquidity patterns on the stock exchange of Hong Kong", *Journal of International Financial Markets, Institutions and Money* 8, 279-300.

——— (1999): "An analysis of depth behavior in an electronic, order-driven environment", *Journal of Banking and Finance* 23, 1861-86.

——— (2001): "Commonality in liquidity: evidence from an order-driven market structure", *The Journal of Financial Research*, forthcoming.

Brown, P, D Walsh and A Yuen (1997): "The interaction between order imbalance and stock price", *Pacific-Basin Finance Journal* 5, 539-57.

Chan, K and W Fong (2000): "Trade size, order imbalance, and the volatility-volume relation", *Journal of Financial Economics* 57, 247-73.

Chan, K C and C Y Hwang (1998): "The impact of tick size on market quality: an empirical investigation of the stock exchange of Hong Kong", Hong Kong University of Science and Technology, *Working Paper Series.*

Chordia, T, R Roll and A Subrahmanyam (2000): "Market liquidity and trading activity", Emory University and University of California, Los Angeles, *Working Paper Series*.

——— (2001a): "Order imbalance, liquidity, and market returns", *Journal of Financial Economics*, forthcoming.

Chordia, T, A Sarkar and A Subrahmanyam (2001b): "Common determinants of bond and stock market liquidity: the impact of financial crises, monetary policy, and mutual fund flows", Emory University, Federal Reserve Bank of New York and University of California, Los Angeles, *Working Paper Series*.

Demsetz, H (1968): "The costs of transacting", *Quarterly Journal of Economics* 82, 33-53.

Easley, D and M O'Hara (1992): "Time and the process of security price adjustment", *Journal of Finance* 47, 577-605.

Engle, Robert F and Joe Lange (1997): "Measuring, forecasting and explaining time varying liquidity in the stock market", *National Bureau of Economic Research Working Paper* 6129*.*

Grossman, S and M Miller (1988): "Liquidity and market structure", National Bureau of Economic Research, *Working Paper* 2641.

Hasbrouck, J and D J Seppi (2001): "Common factors in prices, order flows and liquidity", *Journal of Financial Economics* 59, 383-411.

Ho, T and H R Stoll (1980): "On dealer markets under competition", *Journal of Finance* 35, 2, 259-67.

——— (1981): "Optimal dealer pricing under transactions and return uncertainty", *Journal of Financial Economics* 9, 47-73.

Huberman, G and D Halka (2001): "Systematic liquidity", *The Journal of Financial Research* 24, 2, 161-78.

Jones, C, G Kaul and M Lipson (1994): "Transactions, volume, and volatility", *Review of Financial Studies* 7, 631-51.

Kempf, A and O Korn (1997): "Market depth and order size", University of Mannheim, *Working Paper Series*.

Kyle, A (1985): "Continuous auctions and insider trading", *Econometric* 53, 6, 1315-35.

O'Hara, M (1995): *Market microstructure theory*, Blackwell, Cambridge MA & Oxford.

O'Hara, M and G Oldfield (1986): "The microeconomics of market marking", *Journal of Financial and Quantitative Analysis*, 21, 361-76, December.

Roll, R (1984): "A simple measure of the bid-ask spread in an efficient market", *Journal of Finance* 39, 1127-39.

Stoll, H (1989): "Inferring the components of the bid-ask spread", *Journal of Finance* 44, 115-34.

Upper, Christian (2000): "How safe was the 'safe haven'? Financial market liquidity during the 1998 turbulences", Deutsche Bundesbank, *Discussion paper* 1/00, February.

White, H (1980): "A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity", *Econometrica* 48, 817-38.

Session 5

Risk measurement

Modelling and forecasting realised volatility[1]

Torben G Andersen,[2] Tim Bollerslev,[3] Francis X Diebold[4] and Paul Labys[5]

This paper provides a general framework for integration of high-frequency intraday data into the measurement, modelling and forecasting of daily and lower-frequency volatility and return distributions. Most procedures for modelling and forecasting financial asset return volatilities, correlations and distributions rely on restrictive and complicated parametric multivariate ARCH or stochastic volatility models, which often perform poorly at intraday frequencies. Use of realised volatility constructed from high-frequency intraday returns, in contrast, permits the use of traditional time series procedures for modelling and forecasting. Building on the theory of continuous-time arbitrage-free price processes and the theory of quadratic variation, we formally develop the links between the conditional covariance matrix and the concept of realised volatility. Next, using continuously recorded observations for the Deutsche mark/dollar and yen/dollar spot exchange rates covering more than a decade, we find that forecasts from a simple long-memory Gaussian vector autoregression for the logarithmic daily realised volatilities perform admirably compared to popular daily ARCH and related models. Moreover, the vector autoregressive volatility forecast, coupled with a parametric lognormal-normal mixture distribution implied by the theoretically and empirically grounded assumption of normally distributed standardised returns, gives rise to well calibrated density forecasts of future returns and correspondingly accurate quantile estimates. Our results hold promise for practical modelling and forecasting of the large covariance matrices relevant in asset pricing, asset allocation and financial risk management applications.

---

[1] This paper supersedes the earlier manuscript “Forecasting volatility: a VAR for VaR”. The work reported in the paper was supported by the National Science Foundation. We are grateful to Olsen and Associates, who generously made available their intraday exchange rate quotation data. For insightful suggestions and comments we thank Rob Engle, Atsushi Inoue, Neil Shephard, Clara Vega, Sean Campbell and seminar participants at Chicago, Michigan, Montreal/CIRANO, NYU, Rice, and the June 2000 Meeting of the Western Finance Association.

[2] Department of Finance, Kellogg School of Management, Northwestern University, Evanston, IL 60208, and NBER, phone: 847-467-1285, e-mail: t-andersen@kellogg.northwestern.edu.

[3] Department of Economics, Duke University, Durham, NC 27708, and NBER, phone: 919-660-1846, e-mail: boller@econ.duke.edu.

[4] Department of Economics, University of Pennsylvania, Philadelphia, PA 19104, and NBER, phone: 215-898-1507, e-mail: fdiebold@sas.upenn.edu.

[5] Graduate Group in Economics, University of Pennsylvania, 3718 Locust Walk, Philadelphia, PA 19104, phone: 435-753-9671, e-mail: labys@ssc.sas.upenn.edu.

Comparative analyses of expected shortfall and value-at-risk under market stress[1]

Yasuhiro Yamai and Toshinao Yoshiba,
Bank of Japan

Abstract

In this paper, we compare value-at-risk (VaR) and expected shortfall under market stress. Assuming that the multivariate extreme value distribution represents asset returns under market stress, we simulate asset returns with this distribution. With these simulated asset returns, we examine whether market stress affects the properties of VaR and expected shortfall.

Our findings are as follows. First, VaR and expected shortfall may underestimate the risk of securities with fat-tailed properties and a high potential for large losses. Second, VaR and expected shortfall may both disregard the tail dependence of asset returns. Third, expected shortfall has less of a problem in disregarding the fat tails and the tail dependence than VaR does.

1. Introduction

It is a well known fact that value-at-risk[2] (VaR) models do not work under market stress. VaR models are usually based on normal asset returns and do not work under extreme price fluctuations. The case in point is the financial market crisis of autumn 1998. Concerning this crisis, CGFS (1999) notes that "a large majority of interviewees admitted that last autumn's events were in the "tails" of distributions and that VaR models were useless for measuring and monitoring market risk". Our question is this: Is this a problem of the estimation methods, or of VaR as a risk measure?

The estimation methods used for standard VaR models have problems for measuring extreme price movements. They assume that the asset returns follow a normal distribution. So they disregard the fat-tailed properties of actual returns, and underestimate the likelihood of extreme price movements.

On the other hand, the concept of VaR as a risk measure has problems for measuring extreme price movements. By definition, VaR only measures the distribution quantile, and disregards extreme loss beyond the VaR level. Thus, VaR may ignore important information regarding the tails of the underlying distributions. CGFS (2000) identifies this problem as *tail risk*.

To alleviate the problems inherent in VaR, Artzner et al (1997, 1999) propose the use of expected shortfall. Expected shortfall is the conditional expectation of loss given that the loss is beyond the VaR level. [3] Thus, by definition, expected shortfall considers loss beyond the VaR level. Yamai and Yoshiba (2002c) show that expected shortfall has no tail risk under more lenient conditions than VaR.

---

1 The views expressed here are those of the authors and do not reflect those of the Bank of Japan. (E-mail: yasuhiro.yamai@boj.or.jp; toshinao.yoshiba@boj.or.jp.) This paper is a revised version of the paper presented at the Third Joint Central Bank Research Conference on Risk Measurement and Systemic Risk on 7-8 March 2002 in Basel. The content of this paper is the same as Yamai, Y and T Yoshiba, "Comparative analyses of expected shortfall and value-at-risk (3): their validity under market stress", *IMES Discussion Paper* No 2002-E-2, Bank of Japan, 2002.

2 VaR at the 100(1-α)% confidence level is the upper 100α percentile of the loss distribution. We denote the VaR at the 100(1–α)% confidence level as $VaR_\alpha(Z)$, where Z is the random variable of loss.

3 When the distributions of loss Z are continuous, expected shortfall at the 100(1–α)% confidence level ($ES_\alpha(Z)$) is defined by the following equation:

$$ES_\alpha(Z) = E[Z | Z \geq VaR_\alpha(Z)] .$$

The existing research implies that the tail risk of VaR and expected shortfall may be more significant under market stress than under normal market conditions. The loss under market stress is larger and less frequent than that under normal conditions. According to Yamai and Yoshiba (2002a), the tail risk is significant when asset losses are infrequent and large.[4]

In this paper, we examine whether the tail risk of VaR and expected shortfall is actually significant under market stress. We assume that the multivariate extreme value distributions represent the asset returns under market stress. With this assumption, we simulate asset returns with those distributions, and compare VaR and expected shortfall.[5,6]

Our assumption of the multivariate extreme value distributions is based on the theoretical results of extreme value theory. This theory states that the multivariate exceedances over a high threshold asymptotically follow the multivariate extreme value distributions. As extremely large fluctuations characterise asset returns under market stress, we assume that the asset returns under market stress follow the multivariate extreme value distributions.

Following this Introduction, Section 2 introduces the concepts and definitions of the tail risk of VaR and expected shortfall based on Yamai and Yoshiba (2002a, 2002c). Section 3 provides a general introduction to multivariate extreme value theory. Section 4 adopts univariate extreme value distributions to examine how the fat-tailed properties of these distributions result in the problems of VaR and expected shortfall. Section 5 adopts simulations with multivariate extreme value distributions[7] to examine how tail dependence results in the tail risk of VaR and expected shortfall. Section 6 presents empirical analyses to examine whether past financial crisis have resulted in the tail risk of VaR and expected shortfall. Finally, Section 7 presents the conclusions and implications of this paper.

2. Tail risk of VaR and expected shortfall

A. The definition and concept of the tail risk of VaR

In this paper, we say that *VaR has tail risk* when VaR fails to summarise the relative choice between portfolios as a result of its underestimation of the risk of portfolios with fat-tailed properties and a high potential for large losses.[8,9] The tail risk of VaR emerges since it measures only a single quantile of the profit/loss distributions and disregards any loss beyond the VaR level. This may lead one to think that securities with a higher potential for large losses are less risky than securities with a lower potential for large losses.

For example, suppose that the VaR at the 99% confidence level of portfolio A is 10 million and that of portfolio B is 15 million. Given these numbers, one may conclude that portfolio B is more risky than portfolio A. However, the investor does not know how much may be lost outside of the confidence

---

When the underlying distributions are discontinuous, see Definition 2 of Acerbi and Tasche (2001).

4 Jorion (2000) makes the following comment in analysing the failure of Long-Term Capital Management (LTCM): "The payoff patterns of the investment strategy [of LTCM] were akin to short positions in options. Even if it had measured its risk correctly, the firm failed to manage its risk properly."

5 Prior comparative analyses of VaR and expected shortfall focus on their sub-additivity. For example, Artzner et al (1997, 1999) show that expected shortfall is sub-additive, while VaR is not. Acerbi et al (2001) prove that expected shortfall is sub-additive, including the cases where the underlying profit/loss distributions are discontinuous. Rockafeller and Uryasev (2000) utilise the sub-additivity of the expected shortfall to find an efficient algorithm for optimising expected shortfall.

6 The other important aspect of the comparative analyses of VaR and expected shortfall is their estimation errors. Yamai and Yoshiba (2002b) show that expected shortfall needs a larger size sample than VaR for the same level of accuracy.

7 For other financial applications of multivariate extreme value theory, see Longin and Solnik (2001), Embrechts et al (2000) and Hartmann et al (2000).

8 We only consider whether VaR and expected shortfall are effective for the relative choice of portfolios. We do not consider the issue of the absolute level of risk, such as whether VaR is appropriate as a benchmark of risk capital.

9 For details regarding the general concept and definition of the tail risk of risk measures, see Yamai and Yoshiba (2002c).

interval. When the maximum loss of portfolio A is 1 trillion and that of B is 16 million, portfolio A should be considered more risky since it loses much more than portfolio B under the worst case. In this case, VaR has tail risk since VaR fails to summarise the choice between portfolios A and B as a result of its disregard of the tail of profit/loss distributions.

We further illustrate the concept of the tail risk of VaR with two examples.

Example 1: Option portfolio (Danielsson (2001))

Danielsson (2001) shows that VaR is conducive to manipulation since it measures only a single quantile. We introduce his illustration as a typical example of the tail risk of VaR.

The solid line in Figure 1 depicts the distribution function of the profit/loss of a given security. The VaR of this security is VaR_0 as it is the lower quantile of the profit/loss distribution.

One is able to decrease this VaR to an arbitrary level by selling and buying options of this security. Suppose the desired VaR level is VaR_D. One way to achieve this is to write a put with a strike price right below VaR_0 and buy a put with a strike price just above VaR_D. The dotted line in Figure 1 depicts the distribution function of the profit/loss after buying and selling the options. The VaR is decreased from VaR_0 to VaR_D. This trading strategy increases the potential for large loss. The right end of Figure 1 shows that the probability of large loss is increased.

This example shows that the tail risk of VaR can be significant with simple option trading. One is able to manipulate VaR by buying and selling options. As a result of this manipulation, the potential for large loss is increased. VaR fails to consider this perverse effect since it disregards any loss beyond the confidence level.

Example 2: Credit portfolio (Lucas et al (2001))

The next example demonstrates the tail risk of VaR in a credit portfolio, using the result of Lucas et al (2001).

Lucas et al (2001) derive an analytic approximation to the credit loss distribution of large portfolios. To illustrate their general result, they provide a simple example of credit loss calculation.[10] They consider a bond portfolio where the amount of credit exposure for individual bonds is identical and the default is triggered by a single factor. For simplicity, they assume that the loss is recognised in the default mode and that the factor sensitivities of the latent variables and default probabilities are homogeneous.[11] They show that the credit loss of the bond portfolio converges almost surely to C, as defined in the following equation, when the number of bonds approaches infinity (Lucas et al (2001, p 1643, equation (14)).

$$C \approx \Phi\left(\frac{s - \rho Y}{\sqrt{1-\rho^2}}\right) \tag{1}$$

Φ :The distribution function of the standard normal distribution

Y :Random variable following the standard normal distribution

s :The value of $\Phi^{-1}(p)$ when the default rate is p, and Φ^{-1} is the inverse of Φ.

ρ :Correlation coefficient among the latent variables

Based on this result, we calculate the distribution functions of the limiting credit loss C for $\rho = 0.7$ and 0.9, and plot them in Figure 2.

The results show that VaR has tail risk. The bond portfolio is more concentrated when $\rho = 0.9$ than when $\rho = 0.7$. The tail of the credit loss distribution is fatter when $\rho = 0.9$ than when $\rho = 0.7$. Thus, the

[10] Lucas et al (2001) also develop more general analyses in their paper.

[11] The total exposure of the bond portfolio is 1.

bond portfolio is more risky when ρ = 0.9 than when ρ = 0.7. However, the VaR at the 95% confidence interval is higher when ρ = 0.7 than when ρ = 0.9. This shows that VaR fails to consider credit concentration since it disregards the loss beyond the confidence level.

The preceding examples show that VaR has tail risk when the loss distributions intersect beyond the confidence level. In such cases, one is able to decrease VaR by manipulating the tails of the loss distributions. This manipulation of the distribution tails increases the potential for extreme losses, and may lead to a failure of risk management. This problem is significant when the portfolio profit/loss is non-linear and the distribution function of the profit/loss is discontinuous.[12]

B. The tail risk of expected shortfall

We define the tail risk of expected shortfall in the same way as the tail risk of VaR. In this paper, we say that *expected shortfall has tail risk* when expected shortfall fails to summarise the relative choice between portfolios as a result of its underestimation of the risk of portfolios with fat-tailed properties and a high potential for large losses.

To illustrate our definition of the tail risk of expected shortfall, we present an example from Yamai and Yoshiba (2002c). Table 1 shows the payoff and profit/loss of two sample portfolios A and B. The expected payoff and the initial investment amount of both portfolios are equal at 97.05.

In most of the cases, both portfolios A and B do not incur large losses. The probability that the loss is less than 10 is about 99% for both portfolios.

The magnitude of extreme loss is different. Portfolio A never loses more than half of its value while Portfolio B may lose three quarters of its value. Thus, portfolio B is more risky than Portfolio A when one is worried about extreme loss.

Table 2 shows the VaR and expected shortfall of the two portfolios at the 99% confidence level. Both VaR and expected shortfall are higher for Portfolio A, which has a lower magnitude of extreme loss. Thus, expected shortfall has tail risk since it chooses the more risky portfolio as a result of its disregard of extreme losses.

The example above shows that expected shortfall may have tail risk. However, the tail risk of expected shortfall is less significant than that of VaR. Yamai and Yoshiba (2002c) show that expected shortfall has no tail risk under more lenient conditions than VaR. This is because VaR completely disregards any loss beyond the confidence level while expected shortfall takes this into account as a conditional expectation.

3. Multivariate extreme value theory

In this section, we give a brief introduction to multivariate extreme value theory.[13] We use this theory to represent asset returns under market stress in the following sections.

Multivariate extreme value theory consists of two modelling aspects: the tails of the marginal distributions and the dependence structure among extreme values.

We restrict our attention to the bivariate case in this paper.

---

[12] Yamai and Yoshiba (2002c) show that VaR has no tail risk when the loss distributions are of the same type of an elliptical distribution.

[13] For detailed explanations of extreme value theory, see Coles (2001), Embrechts et al (1997), Kotz and Nadarajah (2000) and Resnick (1987).

A. Univariate extreme value theory

Let Z denote a random variable and F the distribution function of Z. We consider extreme values in terms of exceedances with a threshold θ ($\theta > 0$). The exceedances are defined as $m_\theta(Z) = \max(Z, \theta)$. Z is larger than θ with probability p, and smaller than θ with probability $1 - p$. Then, by the definition of exceedances, $p = 1 - F(\theta)$. We call p tail probability.

The conditional distribution F_θ defined below gives the stochastic behaviour of extreme values.

$$F_\theta(x) = \Pr\{Z - \theta \le x \mid Z > \theta\} = \frac{F(x) - F(\theta)}{1 - F(\theta)}, \ \theta \le x. \tag{2}$$

This is the distribution function of $(Z - \theta)$ given that Z exceeds θ. F_θ is not known precisely unless F is known.

The extreme value theory tells us the approximation to F_θ that is applicable for high values of threshold θ. The Pickands-Balkema-de Haan theorem shows that as the value of θ tends to the right end point of F, F_θ converges to a generalised Pareto distribution. The generalised Pareto distribution is represented as follows:[14, 15]

$$G_{\xi,\sigma}(x) = 1 - (1 + \xi \cdot \frac{x}{\sigma})^{-1/\xi}, \ x \ge 0. \tag{3}$$

With equations (1) and (2), when the value of θ is sufficiently large, the distribution function of exceedances $m_\theta(Z)$, denoted by $F_m(x)$, is approximated as follows:

$$F_m(x) \approx (1 - F(\theta))G_{\xi,\sigma}(x - \theta) + F(\theta) = 1 - p(1 + \xi \cdot \frac{x - \theta}{\sigma})^{-1/\xi}, \ x \ge \theta. \tag{4}$$

In this paper, we call $F_m(x)$ the *distribution of exceedances*.

The distribution of exceedances is described by three parameters: the tail index ξ, the scale parameter σ, and the tail probability p. The tail index ξ represents how fat the tail of the distribution is, so the tail is fat when ξ is large (see Figure 3). The scale parameter σ represents how dispersed the distribution is, so the distribution is dispersed when σ is large (see Figure 4). The tail probability p determines the threshold θ as $F_m(\theta) \approx 1 - p$.

When the confidence level of VaR and expected shortfall is less than p, the distribution of exceedances is used to calculate VaR and expected shortfall. (See Section 4 for the specific calculations.)

B. Copula

As a preliminary to the dependence modelling of extreme values, we provide a simple explanation of copula.[16]

Suppose we have two-dimensional random variables (Z_1, Z_2). Their joint distribution function $F(x_1, x_2) = P[Z_1 \le x_1, Z_2 \le x_2]$ fully describes their marginal behaviour and dependence structure. The main idea of copula is that we separate this joint distribution into the part that describes the dependence structure and the part that describes the marginal behaviour.

Let $(F_1(x_1), F_2(x_2))$ denote the marginal distribution functions of (Z_1, Z_2). Suppose we transform (Z_1, Z_2) to have standard uniform marginal distributions.[17] This is done by $(Z_1, Z_2) \mapsto (F_1(Z_1), F_2(Z_2))$. The joint

---

14 See Coles (2001) and Embrechts et al (1997) for a detailed explanation of this theorem.

15 In this paper we assume that $\xi \ne 0$.

16 For the precise definition of copula and proofs of the theorems adopted here, see eg Embrechts et al (2002), Joe (1997), Nelsen (1999) and Frees and Valdez (1998).

17 The standard uniform distribution is the uniform distribution over the interval [0,1].

distribution function C of the random variable ($F_1(Z_1)$,$F_2(Z_2)$) is called the copula of the random vector (Z_1,Z_2). It follows that:

$$F(x_1,x_2)=P[Z_1 \le x_1, Z_2 \le x_2]=C(F_1(x_1),F_2(x_2)). \tag{5}$$

Sklar's theorem shows that (4) holds with any F for some copula C and that C is unique when $F_1(x_1)$ and $F_2(x_2)$ are continuous.

In general, the copula is defined as the distribution function of a random vector with standard uniform marginal distributions. In other words, the distribution function C is a copula function for the two random variables U_1, U_2 that follow the standard uniform distribution.

$$C(u_1,u_2)=\Pr[U_1 \le u_1, U_2 \le u_2]. \tag{6}$$

One of the most important properties of the copula is its invariance property. This property says that a copula is invariant under increasing and continuous transformations of the marginals. That is, when the copula of (Z_1,Z_2) is $C(u_1,u_2)$ and $h_1(\bullet),h_2(\bullet)$ are increasing continuous functions, the copula of ($h_1(Z_1)$,$h_2(Z_2)$) is also $C(u_1,u_2)$.

The invariance property and Sklar's theorem show that a copula is interpreted as the dependence structure of random variables. The copula represents the part that is not described by the marginals, and is invariant under the transformation of the marginals.

C. Multivariate extreme value theory

We give a brief illustration of the bivariate exceedances approach as a model for the dependence structure of extreme values.[18]

Let $Z=(Z_1,Z_2)$ denote the two-dimensional vector of random variables and $F(Z_1,Z_2)$ the distribution function of Z. The bivariate exceedances of Z correspond to the vector of univariate exceedances defined with a two-dimensional vector of threshold $\theta=(\theta_1,\theta_2)$ (see Figure 5). These exceedances are defined as follows:

$$m_{(\theta_1,\theta_2)}(Z_1,Z_2)=(\max(Z_1,\theta_1),\max(Z_2,\theta_2)). \tag{7}$$

The marginal distributions of the bivariate exceedances defined in (6) converge to the distribution of exceedances introduced in Section 3.A when the thresholds tend to the right end points of the marginal distributions. This is because the bivariate exceedance is the vector of univariate exceedances whose distribution converges to a generalised Pareto distribution.

The copula of bivariate exceedances also converges to a class of copula that satisfies several conditions. Ledford and Tawn (1996) show that this class is represented by the following equation (see Appendix A for details):

$$C(u_1,u_2)=\exp\{-V(-\frac{1}{\log u_1},-\frac{1}{\log u_2})\}, \tag{8}$$

where

$$V(z_1,z_2)=\int_0^1 \max\{sz_1^{-1},(1-s)z_2^{-1}\}dH(s), \tag{9}$$

and H is a non-negative measure on [0,1] satisfying the following condition:

$$\int_0^1 s\,dH(s)=\int_0^1 (1-s)dH(s)=1. \tag{10}$$

Following Hefferman (2000), we call this type of copula the *bivariate extreme value copula* or the *extreme value copula*.

18 For more detailed explanations of multivariate extreme value theory, see Coles (2001) Ch 8 , Kotz and Nadarajah (2000) Ch 3, McNeil (2000), Resnick (1987) Ch 5, etc.

The class of the extreme value copula is wide, being constrained only by (9). We have an infinite number of parameterised extreme value copulas. In practice, we choose a parametric family of copula that satisfies (9), and use the copula for the analysis of bivariate extreme values.

One standard type of bivariate extreme value copula is the Gumbel copula. The Gumbel copula is the most frequently used extreme value copula for applied statistics, engineering and finance (Gumbel (1960), Tawn (1988), Embrechts et al (2002), McNeil (2000), Longin and Solnik (2001)). The Gumbel copula is expressed by:

$$C(u_1,u_2) = \exp\{-[(-\log u_1)^{\alpha} + (-\log u_2)^{\alpha}]^{1/\alpha}\}, \tag{11}$$

for a parameter $\alpha \in [1,\infty]$. We obtain (10) by defining V in (8) as follows:

$$V(z_1,z_2) = (z_1^{-\alpha} + z_2^{-\alpha})^{1/\alpha}. \tag{12}$$

The dependence parameter α controls the level of dependence between random variables. α = 1 corresponds to full dependence and $\alpha = \infty$ corresponds to independence.

The Gumbel copula has several advantages over other parameterised extreme value copulas.[19] It includes the special cases of independence and full dependence, and only one parameter is needed to model the dependence structure. The Gumbel copula is tractable, which facilitates simulations and maximum likelihood estimations. Given these advantages, we adopt the Gumbel copula as the extreme value copula.

To summarise, extreme value theory shows that the bivariate exceedances asymptotically follow a joint distribution whose marginals are the distributions of exceedances and whose copula is the extreme value copula.

D. Tail dependence

We introduce the concept of tail dependence between random variables. Suppose that a random vector (Z_1,Z_2) has a joint distribution function $F(Z_1,Z_2)$ with marginals $F_1(x_1),F_2(x_2)$.

Assume that marginals are equal. We define a dependence measure χ as follows:

$$\chi \equiv \lim_{z\to z^+} \Pr\{Z_1 > z \mid Z_2 > z\}, \tag{13}$$

where z^+ is the right end point of F.

χ measures the asymptotic survival probability over one value to be large given that the other is also large. When $\chi = 0$, we say Z_1 and Z_2 are *asymptotically independent*. When $\chi > 0$, we say Z_1 and Z_2 are *asymptotically dependent*. χ increases with the strength of dependence within the class of asymptotically dependent variables.

When F has different marginals F_{Z_1} and F_{Z_2}, χ is defined as follows:

$$\chi \equiv \lim_{u\to 1} \Pr\{F_{Z_1}(Z_1) > u \mid F_{Z_2}(Z_2) > u\}. \tag{14}$$

Further defining the other dependence measure $\chi(u)$ as in (14), the relationship $\chi = \lim_{u\to 1}\chi(u)$ holds (Coles et al (1999)).

$$\chi(u) \equiv 2 - \frac{\log \Pr\{F_{Z_1}(Z_1) < u, F_{Z_2}(Z_2) < u\}}{\log \Pr\{F_{Z_1}(Z_1) < u\}}, \text{ for } 0 \le u \le 1. \tag{15}$$

[19] For other parameterised extreme value copulas, see, for example, Joe (1997) and Kotz and Nadarajah (2000).

Although χ measures dependence when random variables are asymptotically dependent, it fails to do so when random variables are asymptotically independent. When random variables are asymptotically independent, $\chi = 0$ by definition and χ is unable to provide dependence information.

The class of asymptotically independent copulas includes important copulas such as the Gaussian copula and the Frank copula, which are introduced in the next section. Ledford and Tawn (1996, 1997) and Coles et al (1999) say that the asymptotically independent case is important in the analysis of multivariate extreme values.

To counter this shortcoming of the dependence measure χ, Coles et al (1999) propose a new dependence measure $\bar{\chi}$ as defined below.

$$\bar{\chi} \equiv \lim_{u \to 1} \bar{\chi}(u) \tag{16}$$

where $$\bar{\chi}(u) \equiv \frac{2 \log \Pr\{F_{Z_1}(Z_1) > u\}}{\log \Pr\{F_{Z_1}(Z_1) > u, F_{Z_2}(Z_2) > u\}} - 1 \tag{17}$$

$\bar{\chi}$ measures dependence within the class of asymptotically independent variables. For asymptotically independent random variables, $-1 < \bar{\chi} < 1$. For asymptotically dependent random variables, $\bar{\chi} = 1$.

Thus, the combination $(\chi, \bar{\chi})$ measures tail dependence for both asymptotically dependent and independent case (see Table 3). For asymptotically dependent random variables, $\bar{\chi} = 1$ and χ measures tail dependence. For asymptotically independent random variables, $\chi = 0$ and $\bar{\chi}$ measures tail dependence.

E. Copula and tail dependence

With some calculations, it is shown that $\chi(u)$ is constant for the bivariate extreme value copula as follows:

$$\chi(u) = \chi = 2 - V(1,1). \text{ for all } 0 \le u \le 1. \tag{18}$$

For the Gumbel copula, this becomes $\chi = 2 - 2^{1/\alpha}$ ($\alpha \ge 1$) (see Table 4). Thus, for the bivariate extreme value copula, random variables are either independent or asymptotically dependent. In other words, the bivariate extreme copula is unable to represent the dependence structure when random variables are asymptotically independent.

Ledford and Tawn (1996, 1997) and Coles (2001) say that multivariate exceedances may be asymptotically independent and that modelling multivariate exceedances with the extreme value copula is likely to lead to misleading results in this case. They say that the use of asymptotically independent copulas is effective when the multivariate exceedances are asymptotically independent. Hefferman (2000) provides a list of asymptotically independent copulas that are useful for modelling multivariate extreme values.

In this paper, we adopt the Gaussian copula and the Frank copula as asymptotically independent copulas. These are defined as follows (see Table 4).

Gaussian copula

$$C(u,v) = \Phi_\rho(\Phi^{-1}(u), \Phi^{-1}(v)) \tag{19}$$

where Φ_ρ is the distribution function of a bivariate standard normal distribution with a correlation coefficient ρ, and Φ^{-1} is the inverse function of the distribution function for the univariate standard normal distribution.

Frank copula[20]

$$C(u,v) = -\frac{1}{\delta} \ln\left(\frac{1-e^{-\delta}-(1-e^{-\delta u})(1-e^{-\delta v})}{1-e^{-\delta}}\right). \tag{20}$$

The dependence parameters ρ and δ control the level of dependence between random variables. For the Gaussian copula, $\rho = \pm 1$ corresponds to full dependence and $\rho = 0$ corresponds to independence. For the Frank copula, $\delta = \pm\infty$ corresponds to full dependence and $\delta = 0$ corresponds to independence.

For both of these copulas, random variables are asymptotically independent. For the Gaussian copula with $-1 < \rho < 1$, $\chi = 0$ and $\overline{\chi} = \rho$. For the Frank copula, $\chi = \overline{\chi} = 0$.[21] The latter shows that the Frank copula has very weak tail dependence.

The use of asymptotically independent copula for modelling multivariate exceedances may bring some doubt since extreme value theory shows that the asymptotic copula of exceedances is the extreme value copula. However, the rate of convergence of marginals may be higher than that of the copula. In this case, the generalised Pareto distribution well approximates the marginals of exceedances while the extreme value copula does not approximate the dependence structure of exceedances. Thus, in some cases, it is valid to assume that marginals are modelled by the generalised Pareto distribution while dependence is modelled by asymptotically independent copula.

4. The tail risk under univariate extreme value distribution

In this section, we examine whether VaR and expected shortfall have tail risk when asset returns are described by the univariate extreme value distribution. We use (4) to calculate the VaR and expected shortfall of two securities with different tail fatness, and examine whether VaR and expected shortfall underestimate the risk of securities with fat-tailed properties and a high potential for large loss.

Suppose Z_1 and Z_2 are random variables denoting the loss of two securities. Using the univariate extreme value theory introduced in Section 3.A, with high thresholds, the exceedances of Z_1 and Z_2 follow the distributions below:

$$F_{m(Z_1)}(x) = 1 - p_1(1+\xi_1 \cdot \frac{x-\theta_1}{\sigma_1})^{-1/\xi_1}, \tag{21}$$

$$F_{m(Z_2)}(x) = 1 - p_2(1+\xi_2 \cdot \frac{x-\theta_2}{\sigma_2})^{-1/\xi_2}. \tag{22}$$

As an example of the tail risk of VaR, we set the parameter values as follows: the tail probability is $p_1 = p_2 = 0.1$; the threshold value is $\theta_1 = \theta_2 = 0.05$; the tail indices are $\xi_1 = 0.1$ and $\xi_2 = 0.5$; and the scale parameters are $\sigma_1 = 0.05$ and $\sigma_2 = 0.035$. Figure 6 plots (21) and (22) with this parameter setting.

Figure 6 shows that VaR has tail risk in this example. Given $\xi_2 > \xi_1$, Z_2 has a fatter tail than Z_1 (see Section 3.A). Thus, Z_2 has a higher potential for large loss than Z_1. However, Figure 6 shows that the VaR at the 95% confidence level is higher for Z_1 than for Z_2. Thus, VaR indicates that Z_1 is more risky than Z_2. As in the two examples in Section 2.A, VaR has tail risk as the distribution functions intersect beyond the VaR confidence level.

---

20 This definition of the Frank copula follows Joe (1997).

21 See Ledford and Tawn (1996, 1997), Coles et al (1999) and Hefferman (2000) for the definition and concepts of tail dependence, including the derivations of χ and $\overline{\chi}$ for each copula.

We derive the conditions for the tail risk of VaR. Following McNeil (2000), we calculate the VaR from (21) and (22). Let $VaR_\alpha(Z)$ denote the VaR of Z at the (1 – α) confidence level. Since VaR is the upper (1 – α) quantile of the loss distribution, the following holds:

$$1-\alpha \approx 1-p\left(1+\xi\cdot\frac{VaR_\alpha(Z)-\theta}{\sigma}\right)^{-1/\xi}. \tag{23}$$

We then solve (23) to obtain the following:

$$VaR_\alpha(Z) \approx \theta+\frac{\sigma}{\xi}\left(\left(\frac{p}{\alpha}\right)^{\xi}-1\right). \tag{24}$$

With (24), we derive the condition of the tail risk of VaR as follows. Without the loss of generality, we assume $\xi_2 > \xi_1$, or that the tail of Z_2 is fatter than the tail of Z_1. In other words, Z_2 has higher potential for extreme loss than Z_1. VaR has tail risk when the VaR of Z_2 is smaller than that of Z_1, or when the following inequality holds:

$$VaR_\alpha(Z_1) > VaR_\alpha(Z_2). \tag{25}$$

Assuming $\theta_1 = \theta_2$ and $p_1 = p_2 = p$ for simplification, we obtain the following condition from (24) and (25):

$$\frac{\sigma_1}{\sigma_2} > \overline{\kappa}_{VaR}, \text{ where } \overline{\kappa}_{VaR} = \frac{\xi_1}{\xi_2}\left(\frac{(p/\alpha)^{\xi_2}-1}{(p/\alpha)^{\xi_1}-1}\right). \tag{26}$$

The value $\overline{\kappa}_{VaR}$ indicates how strict the condition for the tail risk of VaR is. When $\overline{\kappa}_{VaR}$ is small, a small difference between the scale parameters σ_1 and σ_2 brings about tail risk of VaR. When $\overline{\kappa}_{VaR}$ is large, a large difference between σ_1 and σ_2 is needed to bring about tail risk of VaR.

Table 5 shows the value of $\overline{\kappa}_{VaR}$ with varying (ξ_1, ξ_2) for VaR at the 95% and 99% confidence levels, when p is 0.05 and 0.1.[22] This table shows two aspects of this condition.

First, the scale parameter of the thin-tailed distribution σ_1 must be larger than the scale parameter of the fat-tailed distribution σ_2. This is because $\overline{\kappa}_{VaR} > 1$ for all combinations of (ξ_1, ξ_2).

Figure 7 illustrates this point. The figure plots the distribution of exceedance with parameter values $\xi_1 = 0.5$, $\sigma_1 = 1$. The figure also plots the distribution of exceedances with parameter values $\xi_2 = 0.1$ and $\sigma_2 = 1$, 1.5 and 2. Here, we denote the VaR for $\xi_1 = 0.5$, $\sigma_1 = 1$ as $VaR(\xi_1 = 0.5, \sigma_1 = 1)$ and that for $\xi_2 = 0.1$, $\sigma_2 = \sigma$ as $VaR(\xi_2 = 0.1, \sigma_2 = \sigma)$. The distribution with $\xi_1 = 0.5$ has a fatter tail and higher potential for large loss than the distribution with $\xi_2 = 0.1$. Thus, VaR has tail risk if $VaR(\xi_1 = 0.5, \sigma_1 = 1) < VaR(\xi_2 = 0.1, \sigma_2 = \sigma)$.

From the figure, we find $VaR(\xi_1 = 0.5, \sigma_1 = 1) < VaR(\xi_2 = 0.1, \sigma_2 = 2)$ with a confidence level below 99%, and $VaR(\xi_1 = 0.5, \sigma_1 = 1) < VaR(\xi_2 = 0.1, \sigma_2 = 1.5)$ with a confidence level below 98%. On the other hand, $VaR(\xi_1 = 0.5, \sigma_1 = 1) > VaR(\xi_2 = 0.1, \sigma_2 = 1)$ with a confidence level above 95%. Therefore, VaR has tail risk with a high confidence level when the difference between the scale parameters is large.

Second, the smaller the difference between the tail indices ξ_1 and ξ_2, the more lenient the conditions for the tail risk of VaR. This is because $\overline{\kappa}_{VaR}$ is small when the difference between the tail indices is small.

Figure 8 illustrates this point. The figure plots the distribution of exceedances with parameter values $\xi_1 = 0.1$, $\sigma_1 = 1$. The figure also plots the distribution of exceedances with parameter values $\sigma_2 = 0.75$

[22] When the tail probability is $p = 0.05$, the VaR at the confidence level of 95% is not beyond the threshold, so we do not calculate VaR at the confidence level of 95% when $p = 0.05$.

and $\xi_2 = 0.3,\ 0.5,\ 0.9$. Here, we denote the VaR for $\xi_1 = 0.1,\ \sigma_1 = 1$ as $VaR(\xi_1 = 0.1, \sigma_1 = 1)$ and that for $\xi_2 = \xi,\ \sigma_2 = 0.75$ as $VaR(\xi_2 = \xi, \sigma_2 = 0.75)$. As the distribution tail is fatter with $\xi_2 = \xi,\ \sigma_2 = 0.75$ than with $\xi_1 = 0.1,\ \sigma_1 = 1$, VaR has tail risk if $VaR(\xi_1 = 0.1, \sigma_1 = 1) > VaR(\xi_2 = \xi, \sigma_2 = 0.75)$. We find $VaR(\xi_1 = 0.1, \sigma_1 = 1) > VaR(\xi_2 = 0.3, \sigma_2 = 0.75)$ with a confidence level below 99%, and $VaR(\xi_1 = 0.1, \sigma_1 = 1) > VaR(\xi_2 = 0.5, \sigma_2 = 0.75)$ with a confidence level below 97%. On the other hand, $VaR(\xi_1 = 0.1, \sigma_1 = 1) < VaR(\xi_2 = 0.9, \sigma_2 = 0.75)$ with a confidence level above 95%. Therefore, VaR has tail risk with a high confidence level when the difference between the tail indices is small.

We analyse the condition for the tail risk of expected shortfall as we analysed that of VaR. Following McNeil (2000), we can calculate the expected shortfall of Z at the $(1 - \alpha)$ confidence level (denoted by $ES_\alpha(Z)$) from (24).[23]

$$\begin{aligned}
ES_\alpha(Z) &= E[Z \mid Z \geq VaR_\alpha(Z)] \\
&= VaR_\alpha(Z) + E[Z - \theta) - (VaR_\alpha(Z) - \theta) \mid Z - \theta \geq VaR_\alpha(Z) - \theta] \\
&= VaR_\alpha(Z) + \frac{\sigma + \xi \cdot (VaR_\alpha(Z) - \theta)}{1 - \xi} \\
&= \frac{\sigma - \xi\theta}{1 - \xi} + \frac{VaR_\alpha(Z)}{1 - \xi} \approx \theta + \frac{\sigma}{1 - \xi}\left\{1 + \frac{1}{\xi}\left(\left(\frac{p}{\alpha}\right)^{\xi} - 1\right)\right\}.
\end{aligned} \tag{27}$$

Given $\xi_2 > \xi_1$, expected shortfall has tail risk when the following inequality holds:

$$ES_\alpha(Z_1) > ES_\alpha(Z_2). \tag{28}$$

Assuming $\theta_1 = \theta_2$ and $p_1 = p_2 = p$ for simplification, we obtain the following condition from (27) and (28):

$$\frac{\sigma_1}{\sigma_2} > \bar{\kappa}_{ES}, \text{ where } \bar{\kappa}_{ES} = \frac{1 - \xi_1}{1 - \xi_2}\left(\frac{1 + \left((p/\alpha)^{\xi_2} - 1\right)/\xi_2}{1 + \left((p/\alpha)^{\xi_1} - 1\right)/\xi_1}\right). \tag{29}$$

Table 6 shows the value of $\bar{\kappa}_{ES}$ with varying (ξ_1, ξ_2) for expected shortfall at the 95% and 99% confidence levels, when p is 0.05 and 0.1.[24] This table shows that the conditions for the tail risk of expected shortfall are stricter than those for the tail risk of VaR. This confirms the result of Yamai and Yoshiba (2002c) that expected shortfall has no tail risk under more lenient conditions than VaR.

To summarise, VaR and expected shortfall may underestimate the risk of securities with fat-tailed properties and a high potential for large loss. The condition for tail risk to emerge depends on the parameters of the loss distribution and the confidence level.

5. The tail risk under multivariate extreme value distribution

The use of risk measures may lead to a failure of risk management when they fail to consider the change in dependence between asset returns. The credit portfolio example in Section 2.A shows that VaR disregards the increase in default correlation and thus fails to note the high potential for extreme loss in concentrated credit portfolios. In this case, the use of VaR for credit portfolios may lead to credit concentration.

In this section, we examine whether VaR and expected shortfall disregard the changes in dependence under a multivariate extreme value distribution. As the multivariate extreme value distribution, we use the joint distribution of exceedances introduced in Section 3.C. The marginal of this distribution is the

[23] The third equality is based on Embrechts et al (1997), Theorem 3.4.13 (e).

[24] We do not calculate expected shortfall at the confidence level of 95% when p = 0.05 (see footnote 22).

generalised Pareto and its copula is the Gumbel copula. We also use the Gaussian and Frank copulas for the copulas of exceedances for the case where the exceedances are asymptotically independent.

A. The difficulty of applying multivariate extreme value distribution to risk measurement

The application of multivariate extreme value distribution to financial risk measurement has some problems that the univariate application does not. In the univariate case, the model for exceedances enables us to calculate VaR and expected shortfall as in Section 4. This is because the VaR and expected shortfall of exceedances are equal to the VaR and expected shortfall of the original loss data. However, in the multivariate case, the model for exceedances is not sufficient to calculate VaR and expected shortfall. This is because, in the multivariate case, the sum of exceedances is not necessarily equal to the exceedances of the sum. To calculate VaR and expected shortfall, we need the exceedances of the sum, which are unavailable from the model for exceedances alone.[25, 26, 27]

A simple example illustrates this point (Figure 9). Let (U_1,U_2) denote a vector of independent standard uniform random variables. With a threshold value of $(\theta_1,\theta_2) = (0.9,0.9)$, the exceedances of (U_1,U_2) are $(m_{0.9}(U_1), m_{0.9}(U_2)) = (\max(U_1,0.9), \max(U_2,0.9))$. With the convolution theorem, the 95% upper quantile of $U_1 + U_2$ is calculated to be 1.68, while that of $m_{\theta_1}(U_1) + m_{\theta_2}(U_2)$ is calculated to be 1.88.[28] Thus, the sum of exceedances is larger than the exceedances of the sum.

This example shows that, to calculate VaR and expected shortfall in the multivariate case, we need a model for non-exceedances as well as one for exceedances.

In this paper, we assume that the marginal distribution of the non-exceedances is the standard normal distribution as we interpret the non-exceedances as asset loss under normal market conditions. That is, we assume that the marginal distribution is expressed by (30) below (Figure 10):[29]

---

25 This is also a problem when the model for maxima is used for calculating VaR and expected shortfall. This is because the sums of maxima are not necessarily equal to the maxima of sums. Hauksson et al (2000) and Bouyé (2001) propose the use of multivariate generalised extreme value distributions for financial risk measurement, but they do not address this problem.

26 The quantile of the sum of exceedances is equal to that of the original data when the underlying random variables are fully dependent.

27 McNeil (2000) says that multivariate extreme value modelling has the problem of "the curse of dimensionality". He notes that, when the number of dimension is more than two, the estimation of copula is not tractable.

28 The upper 95% quantile of $U_1 + U_2$ is calculated as follows. Denote the distribution function of $U_1 + U_2$ as $G(x)$. Clearly, the upper 95% quantile of $U_1 + U_2$ is greater than 1. So assuming $x > 1$, $G(x)$ is calculated by the convolution theorem as follows:

$$G(x) = \int_0^1 \Pr[U_1 \le x-u]du = -\frac{1}{2}(x-2)^2 + 1$$

The upper 95% quantile is x that satisfies $G(x) = 0.95$, which is calculated as $x \approx 1.6838$.

The upper 95% quantile of the sum of the exceedances is calculated as follows. Define $H(x) \equiv \Pr[\max(U_1,0.9) + \max(U_2,0.9) \le x]$. Using the convolution theorem, this is restated as follows:

$$H(x) = \int_0^1 \Pr[\max(U_1,0.9) \le x-u] \cdot \Pr[\max(U_2,0.9) = u]du = \begin{cases} x^2/2 - 0.81 & (x \le 1.9) \\ -(x-2)^2/2 + 1 & (x > 1.9) \end{cases}$$

The upper 95% quantile is x that satisfies $G(x) = 0.95$, which is calculated as $x \approx 1.8761$.

29 A different assumption might be that the marginal distribution of exceedances is a non-standard normal distribution, a *t*-distribution, a generalised Pareto distribution, or an empirical distribution produced from actual data. Assuming a non-standard normal distribution, a *t*-distribution, and a generalised Pareto distribution, we simulated asset loss as in sections B and C of this chapter, and found the same result as in those sections. Furthermore, under the assumption of a generalised Pareto distribution, the convolution theorem is applied to obtain the analytics of the tail risk of VaR (see Appendix B for the details).

$$F(x)=\begin{cases}\Phi(x) & (x<\Phi^{-1}(1-p)),\\ 1-p(1+\xi\cdot\dfrac{x-\Phi^{-1}(1-p)}{\sigma})^{-1/\xi} & (x\geq\Phi^{-1}(1-p)).\end{cases} \tag{30}$$

Φ :the distribution function of the standard normal

Φ^{-1} :the inverse function of Φ

In the following analysis, we simulate two dependent asset losses to analyse the tail risk of VaR and expected shortfall.[30] In the simulation, we assume that the marginal distribution of asset loss is (30). We also assume that the copula of asset loss is one of three copulas introduced in Section 3.E: Gumbel, Gaussian and Frank. We set the marginal distribution of each asset loss as identical so that we can examine the pure effect of dependence on the tail risk of VaR and expected shortfall. We limit our attention to the cases where the tail index is $0<\xi<1$.[31]

B. One specific copula case

In this section, we assume that the change in the dependence structure of asset loss is represented by the change in the dependence parameters within one specific copula. Under this assumption, we examine whether VaR and expected shortfall consider the change in dependence by taking the following steps. First, we take one of the three copulas introduced in Section 3.E: Gumbel, Gaussian or Frank. Second, we simulate asset losses under the one copula for varied dependence parameter levels (Gumbel: α, Gaussian: ρ, and Frank: δ). Third, we calculate VaR and expected shortfall with the simulated asset losses for each dependence parameter level.

If VaR and expected shortfall do not increase with the rise in the level of dependence, VaR and expected shortfall disregard dependence and thus have tail risk.

Figure 11 shows an example of this analysis. The figure plots the empirical distribution of the sum of two simulated asset losses. These losses are simulated adopting (30) as the marginals and the Gumbel copula as the copula. The parameters of the marginal are set at $\xi=0.5$, $\sigma=1$, $p=0.1$, and the dependence parameter α of the Gumbel copula is set at 1.0, 1.1, 1.5, 2.0 and ∞.[32] For each dependence parameter, we conduct one million simulations.

The result shows that the distribution tail gets fatter as the value of the dependence parameter α increases, or the asset losses are more dependent. Furthermore, the empirical distributions do not intersect with each other. This shows that the portfolio diversification effect works to decrease the risk of the portfolio and that VaR has no tail risk regardless of its confidence level.

Table 7 provides a more general analysis. The figure gives the VaR and expected shortfall under one million simulations for each copula with various dependence parameter levels. Two of the three marginal distribution parameters (ξ, σ, p) are set at σ = 1, p = 0.1, and the tail index ξ is set at 0.1, 0.25, 0.5 and 0.75. One of the copulas (Gumbel, Gaussian and Frank) is adopted. With these marginals and copulas, asset losses are simulated. VaR and expected shortfall are calculated for varied dependence parameter levels (Gumbel: α, Gaussian: ρ, and Frank: δ).

---

30 We use the Mersenne Twister for generating uniform random numbers, and the Box-Müller method for transforming the uniform random numbers into normal random numbers. We follow Frees and Valdez (1998) in simulating the Gumbel copula, and Joe (1997) for simulating the Gaussian and Frank copulas.

31 The generalised Pareto distribution with $\xi>1$ is so fat-tailed that its mean is infinite (Embrechts et al (1997), Theorem 3.4.13 (a)).

The generalised Pareto distribution with $\xi>1$ has several interesting properties. However, it is not considered in this paper because such a fat-tailed distribution is rarely observed in financial data. For details, see Appendix B.

32 Under the Gumbel copula $\chi=2-2^{1/\alpha}$, so the corresponding values of χ become $\chi=0,\ 0.12,\ 0.41,\ 0.59,\ 1$.

Table 7 shows that VaR and expected shortfall consider the change in dependence and have no tail risk in most of the cases. VaR and expected shortfall increase as the value of the dependence parameter rises, except for the Frank copula with extremely high dependence parameter levels.[33]

To summarise, VaR and expected shortfall have no tail risk when the change in dependence is represented by the change in parameters using one specific copula. Thus, if we select portfolios whose dependence structure is nested in one of the three copulas above, we can depend on VaR and expected shortfall for measuring dependent risks.

C. Different copulas case

In the previous section, we assume that the change in the dependence of asset losses is represented by the change in the parameters using one specific copula. However, this assumption has a problem. One specific copula does not represent both asymptotic dependence and asymptotic independence.

Let us consider an example of this problem. Suppose we have two portfolios both composed of two securities. Also suppose that the security returns of one portfolio are asymptotically dependent while those of the other are asymptotically independent. Adopting one specific copula and changing the dependence parameters to describe the change in dependence does not work in this case. This is because one specific copula does not represent the change from asymptotic dependence to asymptotic independence. We need different types of copulas to compare asymptotic dependence with asymptotic independence.

In this section, we assume that the change in dependence is represented by the change in copula. We adopt the Gumbel, Gaussian and Frank copulas introduced in Section 3.E since the Gumbel copula corresponds to asymptotic dependence and the Gaussian and Frank copulas correspond to asymptotic independence. By changing copula from Gumbel to Gaussian and Frank, we can change the dependence structure from asymptotic dependence to asymptotic independence.

In comparing the results with three copulas, we set the values of the dependence parameters of those copulas (Gumbel: α, Gaussian: ρ, and Frank: δ) so that the Spearman's rho (ρ_S) is equal across those copulas.[34,35] By setting the Spearman's rho equal, we can eliminate the effect of global dependence and examine the pure effect of tail dependence since the Spearman's rho is a measure of global dependence.

The upper half of Figure 12 shows the empirical distributions of the sums of two simulated asset losses for the Gumbel, Gaussian and Frank copulas. This is generated from one million simulations for each copula where the parameters are fixed at $\xi = 0.5$, $\sigma = 1$, $\rho_S = 0.5$, $p = 0.1$. The range of the horizontal axis (cumulative probability) is above 99.5%.

The tail shape of the loss distribution for each copula is consistent with the tail dependence of each copula. The empirical loss distribution for the Gumbel copula, which is asymptotically dependent

---

33 In the case of the Frank copula, the VaR at the 95% confidence level when $\delta = \infty$ (full dependence) is smaller than the VaR when $\delta = 9$.

This might be because the Frank copula has low tail dependence ($\chi = \bar{\chi} = 0$) and does not represent tail dependence when δ is large.

34 The Spearman's rho is the linear correlation of the marginals, and is defined by the following equation:

$$\rho_S(Z_1, Z_2) \equiv \frac{Cov(F_{Z_1}(Z_1), F_{Z_2}(Z_2))}{\sqrt{V[F_{Z_1}(Z_1)]V[F_{Z_2}(Z_2)]}}.$$

The Spearman's rho differs from χ and $\bar{\chi}$ in that it measures global dependence while χ and $\bar{\chi}$ measure tail dependence.

The Spearman's rho does not fully represent the dependence structures since the combination of the Spearman's rho and the marginal distribution does not uniquely define the joint distribution. In particular, it does not represent the asymptotic dependence measured by χ and $\bar{\chi}$. Nevertheless, the Spearman's rho is relatively superior as a single measure of global dependence (see Embrechts et al (2002)).

35 We use the calculation in Joe (1997, p 147, Table 5.2) for the values of the dependence parameters that equate the Spearman's rho.

$(\chi > 0, \bar{\chi} = 1)$, has the fattest tail. The empirical loss distribution for the Frank copula, which has the weakest tail dependence $(\chi = 0, \bar{\chi} = 0)$, has the thinnest tail.[36]

This shows that the potential for extreme loss is high when the tail dependence is high. Thus, if we are worried about extreme loss, portfolios with higher tail dependence should be considered more risky than those with lower tail dependence. As for the three copulas adopted here, we should consider the Gumbel copula as the most risky and the Frank copula the least risky in terms of tail risk. In this context, VaR and expected shortfall have tail risk when they do not increase in the order of Frank, Gaussian and Gumbel copulas.

The lower half of Figure 12 shows that VaR has tail risk in this example. The figure shows that the VaR at the 95% confidence level increases in the order of Gumbel, Gaussian and Frank. VaR says that the Gumbel copula is the least risky while the Frank copula is the most risky. This contradicts our observation of the upper tail described above.

Table 8 provides a more general analysis. The table shows the results of VaR and expected shortfall calculations for one million simulations for each copula with the tail index of the marginal distribution of $\xi =$ 0.1, 0.25, 0.5, and 0.75, and Spearman's rho of ρ_S = 0.2, 0.5 and 0.8.

The findings of the analysis are threefold. First, VaR and expected shortfall vary depending on the copula adopted. This means that the type of copula affects the level of VaR and expected shortfall. The difference is large when the tail index and the Spearman's rho are large.

Second, VaR at the 95% confidence level has tail risk when the tail index ξ is 0.25 or higher. For example, when $\xi = 0.5$ and ρ_S = 0.8, the VaR at the 95% confidence level is largest for the Frank copula and smallest for the Gumbel copula. On the other hand, VaR at the 99% and 99.9% confidence level has no tail risk, except when the tail is as fat as $\xi = 0.75$.

Third, expected shortfall has no tail risk at the 95, 99, or 99.9% confidence level, except when the tail is as fat as $\xi = 0.75$. This confirms the result of Yamai and Yoshiba (2002c) that expected shortfall has no tail risk under more lenient conditions than VaR.

D. Different marginals case

In Sections 5.B and 5.C, the marginal distributions are assumed to be identical. In financial data, however, the distributions of asset returns are rarely identical. In this section, we extend our analysis to the different marginals case. We examine whether the conclusions in Sections 5.B and 5.C are still valid when the marginal distributions are different.

1. Independence vs full dependence case

We examine whether the results in Section 5.B (the specific copula case) are still valid when the marginal distributions are different. We compare independence and full dependence, noting the fact that independence and full dependence are nested in the Gumbel, Gaussian and Frank copulas. When the VaR for independence is higher than the VaR for full dependence, VaR has tail risk.

We simulate independent and fully dependent asset losses with all combinations of parameters of the marginal distributions from $\xi_1 = 0.1$, 0.25, 0.5, 0.75, $\xi_2 = 0.1$, 0.25, 0.5, 0.75, $\sigma_1 = 1$, $\sigma_2 = 1.00$, 1.25, 1.5,…, 9.5, 9.75, 10. We set the number of simulations at one million for each parameter combination. We calculate VaR and expected shortfall for both independence and full dependence, and compare them to see whether they have tail risk. We adopt the tail probability of $p = 0.1$.

We found that the VaR for full dependence is never smaller than the VaR for independence.[37] Thus, at least within this framework, VaR captures full dependence and independence when the marginal distributions are different.

---

36 See Figure 7 for the values of χ and $\bar{\chi}$ for each copula.

2. Different copulas case

We next examine whether the results in Section 5.B (the different copulas case) are still valid when the marginal distributions are different. We follow the same steps as in Section 5.C except that we set different parameter levels for two marginal distributions.

Under each one of the three copulas, as in Section 5.B, we simulate asset losses following the same method used in the previous subsection.

We find that VaR at the 95% confidence level may have tail risk even when the distribution tail is not so fat as $\xi = 0.25$.[38] This means that the conditions of the tail risk of VaR are more lenient when the marginals are different than when they are identical. Table 9 shows that, with a tail index of $\xi = 0.1$, VaR at the 95% confidence level has tail risk. VaR is larger for the Gaussian copula than for the Gumbel copula.[39]

On the other hand, at the confidence level of 99%, we find that VaR has tail risk only when the tail is as fat as $\xi = 0.75$.

6. Empirical analyses

In Sections 4 and 5, we examine the tail risk of VaR and expected shortfall under extreme value distributions. We summarise the results as follows.

In the univariate case, VaR and expected shortfall may underestimate the risk of securities with fat-tailed properties and a high potential for large losses. The conditions for this to happen are expressed by a simple analytical inequality.

In the multivariate case, VaR and expected shortfall may both disregard the tail dependence when the tails of the marginal distributions are fat.

In this section, we conduct empirical analyses with exchange rate data to confirm whether VaR and expected shortfall have tail risk in actual financial data. We focus on the following questions.

Do VaR and expected shortfall underestimate the risk of currencies with fat-tailed properties and a high potential for large losses in the univariate case?

Is there asymptotic dependence that may bring the tail risk of VaR and expected shortfall in the multivariate case?

A. Data

The data used for the analyses are the daily logarithmic changes of exchange rates of three industrialised countries and 18 emerging economies.[40,41,42] The raw historical data are the exchange rates per one US dollar from 1 November 1993 to 29 October 2001.

---

37 The results of this simulation are omitted here due to space restrictions.

38 See Footnote 37.

39 This finding was confirmed by running 10 million simulations.

40 The data are sourced from Bloomberg.

41 We set the exchange rate as constant over holidays at the levels of the previous business day. This treatment does not affect our results as we estimate only the tails of distributions.

42 The currencies of developed countries are as follows: Japanese yen, the Deutsche mark and pound sterling. The currencies of emerging economies are as follows: Hong Kong dollar, Indonesian rupiah, Malaysian ringgit, Philippine peso, Singapore dollar, South Korean won, new Taiwan dollar, Thai baht, Czech koruna, Hungarian forint, Polish zloty, Slovakian koruna, Brazilian real, Chilean peso, Colombian peso, Mexican new peso, Peruvian new sol and Venezuelan bolívar.

B. Univariate analyses

We estimate the parameters of the generalised Pareto distribution on the daily exchange rate data.[43] We use the maximum likelihood method described in Embrechts et al (1997), and Coles (2001). We vary the tail probability as 1%, 2%, …, 10%, and estimate the parameters ξ, σ, and θ for each. We then calculate the VaR and expected shortfall at the confidence levels of 95% and 99% using the estimated parameter values.

Table 10 shows the estimation results, and these findings may be summarised as follows. First, the tail indices are higher for the emerging economies (especially those in Asia and South America) than for the developed countries. In other words, the distribution tails are fatter in the emerging economies than in the developed countries.

Second, the scale parameter (σ) is smaller in the emerging economies than in the developed countries. This suggests that the condition for tail risk derived in Section 4 may hold.

Third, VaR has tail risk in comparing the risk of some emerging economies and some developed countries. For example, let us compare the VaR for Japan and those for emerging economies.[44] The VaR at the 95% confidence level for all the emerging economies except for Indonesia and Brazil is smaller than that for Japan. Even the VaR at the 99% confidence level is smaller for 10 emerging economies (Hong Kong, Singapore, Taiwan, Hungary, Poland, Slovakia, Chile, Columbia, Peru and Venezuela) than that for Japan.

Fourth, expected shortfall also has tail risk in comparing the risk of some emerging economies and some developed countries. For example, the expected shortfall at the 99% confidence level is smaller for six emerging economies (Hong Kong, Singapore, Taiwan, Chile, Columbia and Peru) than for Japan.[45]

Fifth, expected shortfall has tail risk in fewer cases than VaR. This is consistent with our findings in Section 4.

C. Bivariate analyses (an example)

We provide an example where VaR has tail risk in actual exchange rate data in the bivariate case. We pick five currencies in Southeast Asian countries: the Indonesian rupiah, the Malaysian ringgit, the Philippine peso, the Singapore dollar and the Thai baht.

First, we estimate the parameters of the bivariate extreme value distribution introduced in Section 3. We adopt the same method as Longin and Solnik (2001). As in the analyses in Sections 4 and 5, we assume that the marginal distributions of bivariate exceedances are approximated by the generalised Pareto distribution (the distribution of exceedance as in (4), to be exact) and that their copula is approximated by the Gumbel copula.[46] Given tail probabilities p_1 and p_2, the joint bivariate distribution of exceedances is described by the following parameters: the tail indices of the marginals (ξ_1 and ξ_2), the scale parameters of the marginals (σ_1 and σ_2), the thresholds (θ_1 and θ_2), and the dependence parameter of the Gumbel copula (α).

We estimate those parameters on the right tails of each pair of Southeast Asian currencies by the maximum likelihood method[47] for the tail probability of 10%. Table 11 shows the results of the estimation.

---

43 The extreme value theory is applicable to a stationary process given that the process satisfies some condition. See Ch 5 of Coles (2001) for details.

44 In the comparison here, we use the averages of the VaRs at the 95% confidence level in the right tail with the tail probabilities from 5% to 10%, and the average of VaRs at the 99% confidence level in the right tail with the tail probabilities from 1% to 10%.

45 In the comparison here, we use the average of the expected shortfalls at the 99% confidence level in the right tail with the tail probabilities from 1% to 10%.

46 Instead of using parametric technique, one is able to use non-parametric estimation techniques. See Capéraà et al (1997) for details.

47 See Longin and Solnik (2001) and Ledford and Tawn (1996) for the construction of the maximum likelihood function.

After the estimation, we examine whether VaR and expected shortfall disregard tail dependence with the estimated parameter levels. We take the same step as in Section 5.C. First, we simulate the logarithm changes in exchange rates with the distribution of exceedances and the Gumbel copula, using the parameter levels estimated here. Second, we also simulate the logarithm changes in exchange rates with the Gaussian and Frank copulas. The dependence parameters for the Gaussian and Frank copulas are set so that the Spearman's rho (ρ_s) is equal to that of Gumbel copula with the dependence parameter α at the estimated level. Third, we calculate the VaR and expected shortfall of the sums of the logarithm changes in two exchange rates. We run ten million simulations for each case.

Table 12 shows the result of those simulations. We find that the VaR at the 95% confidence level has tail risk for each pair of Southeast Asian currencies since the VaRs are larger for the Gaussian copula than for the Gumbel copula. Thus, VaR may disregard tail dependence in actual financial data. On the other hand, the VaR at the 99% confidence level and the expected shortfall at the 95% and 99% confidence levels have no tail risk in this example.

7. Conclusions and implications

This paper shows that VaR and expected shortfall have tail risk under extreme value distributions. In the univariate case, VaR and expected shortfall may underestimate the risk of securities with fat-tailed properties and a high potential for large losses. In the multivariate case, VaR and expected shortfall may disregard the tail dependence.

The tail risk is the result of the interaction among various factors. These include the tail index, the scale parameter, the tail probability, the confidence level and the dependence structure.

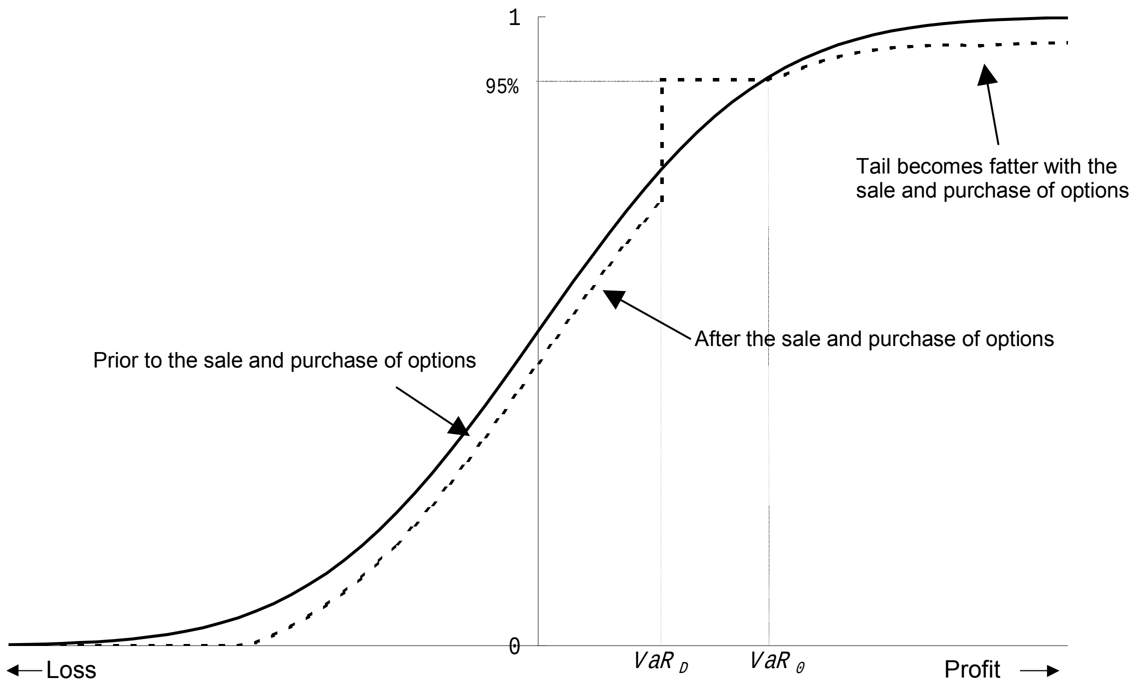
These findings imply that the use of VaR and expected shortfall should not dominate financial risk management. Dependence on a single risk measure has a problem in disregarding important information on the risk of portfolios. To capture the information disregarded by VaR and expected shortfall, it is essential to monitor diverse aspects of the profit/loss distribution, such as tail fatness and asymptotic dependence.

The findings also imply that the widespread use of VaR for risk management could lead to market instability.[48] Basak and Shapiro (2001) show that when investors use VaR for their risk management, their optimising behaviour may result in market positions that are subject to extreme loss because VaR provides misleading information regarding the distribution tail. They also note that such investor behaviour could result in higher volatility in equilibrium security prices. This paper shows that, under extreme value distribution, VaR may provide misleading information regarding the distribution tail.

[48] See Dunbar (2001) for the practitioners' view on this argument.

Figure 1

Tail risk of VaR with option trading

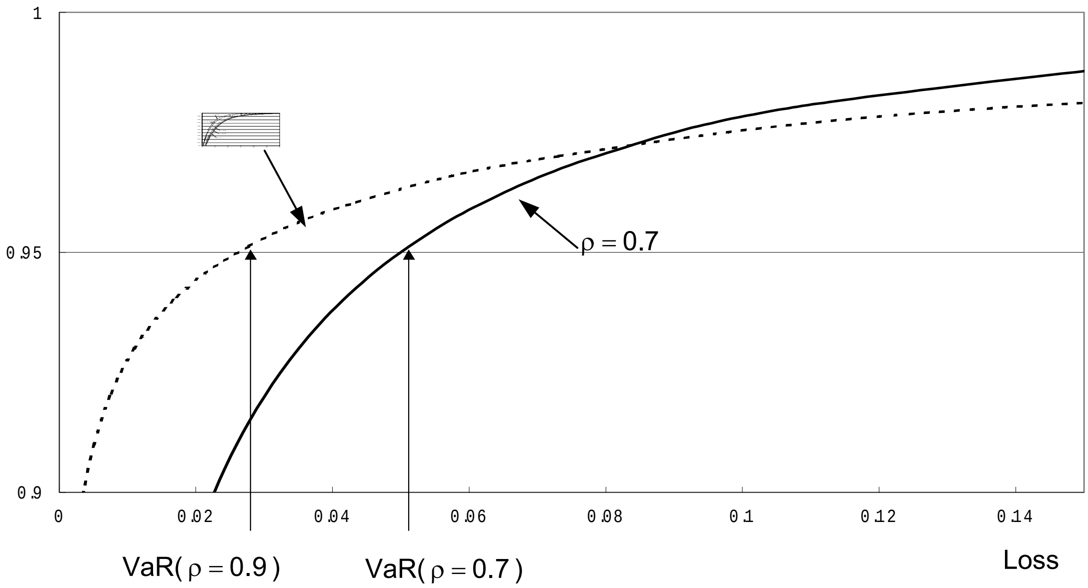


Source: Based on Danielsson (2001), Figure 2.

Figure 2

Tail risk of VaR in a credit portfolio

(Loss distribution of a uniform portfolio with a default rate of 1%)



Source: Calculated from equation (14) in Lucas et al (2001).

Table 1

Sample portfolio payoff

Portfolio A			Portfolio B		
Payoff	**Loss**	**Probability**	**Payoff**	**Loss**	**Probability**
100	– 2.95	50.000%	98	– 0.95	50.000%
95	2.05	49.000%	97	0.05	49.000%
50	47.05	1.000%	90	7.05	0.457%
			20	77.05	0.543%

Note: The probability that Portfolio B has a payoff of 90 or 20 is rounded off, and not precisely expressed. The model is set so that the sum of the probabilities of these payoffs is 1% and the expected payoff is 97.05.

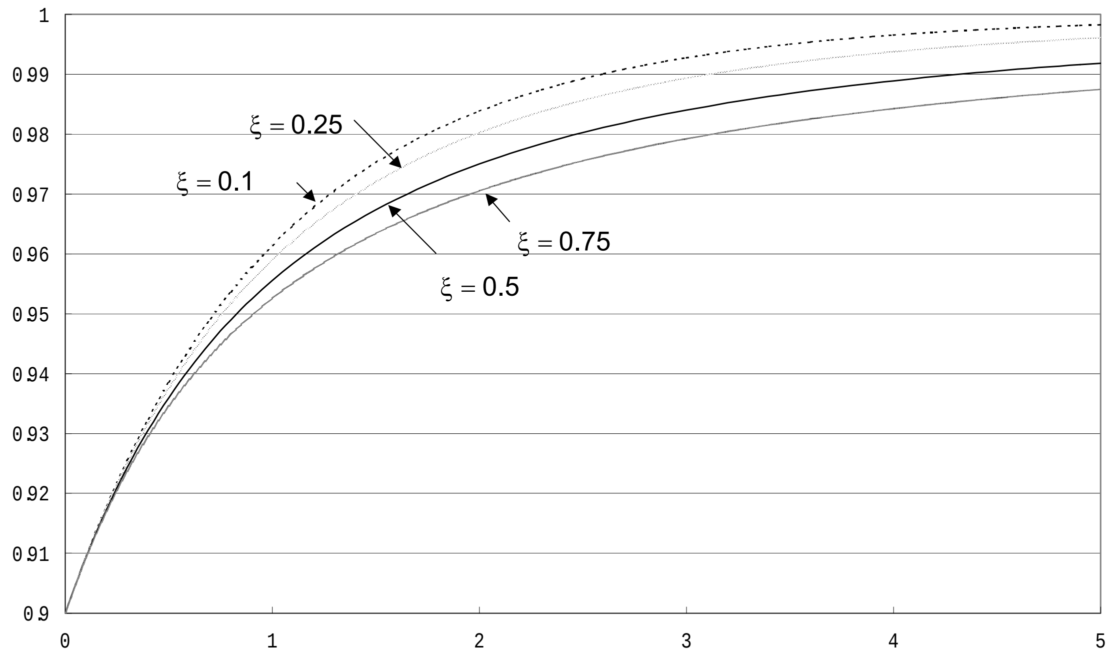
Table 2

Sample portfolio VaR and expected shortfall

	Portfolio A	Portfolio B
Expected payoff	97.05	97.05
VaR (confidence level: 99%)	47.05	7.05
Expected shortfall (confidence level: 99%)	47.05	45.05

Figure 3

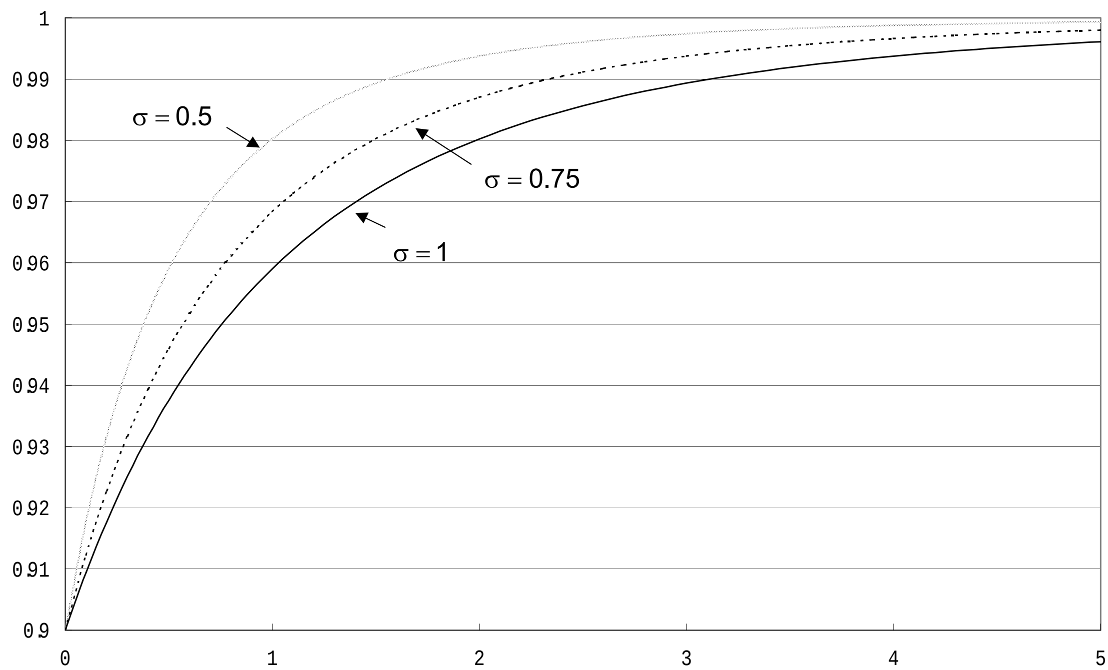
Distribution of exceedances with varied tail indices



Note: Where the tail probability is $p = 0.1$, the threshold value is $\theta = 0$, and the scale parameter is $\sigma = 1$.

Figure 4

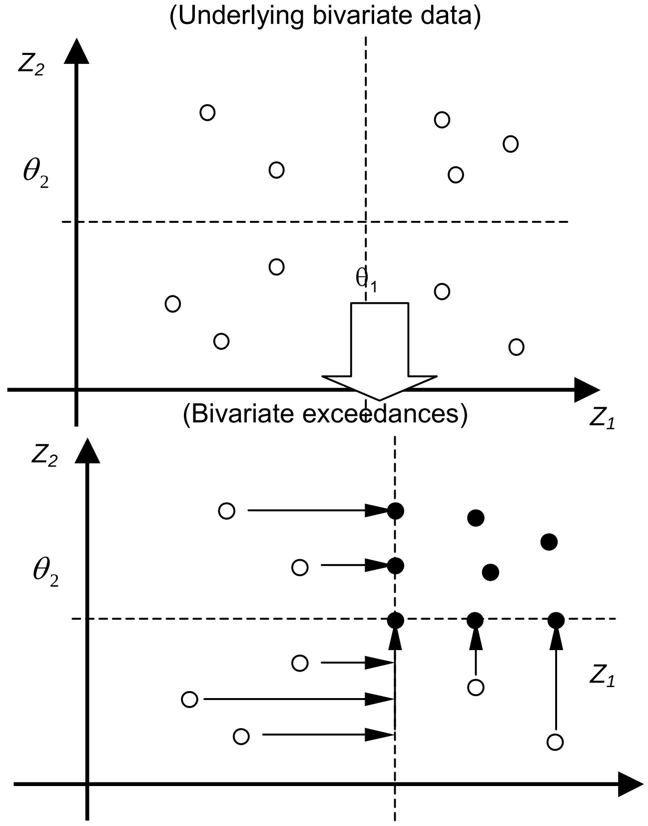
Distribution of exceedances with varied scale parameters



Note: Where the tail probability is $p = 0.1$, the threshold value is $\theta = 0$, and the tail index is $\xi = 0.25$.

Figure 5

Image diagram of bivariate exceedances



Note: The white circles represent the values of the underlying bivariate data and the black circles represent their exceedances.
Source: Based on Reiss and Thomas (2000), Figure 10.1.

Table 3

Asymptotic dependence and dependence measures χ and $\bar{\chi}$

	Independent	**Asymptotically independent**	**Asymptotically dependent**
χ	$\chi = 0$	$\chi = 0$	$0 < \chi \leq 1$
$\bar{\chi}$	$\bar{\chi} = 0$	$-1 < \bar{\chi} < 1$	$\bar{\chi} = 1$
Reference	Represented by the extreme value copula	Not represented by the extreme value copula	Represented by the extreme value copula

Note: When independent, $\bar{\chi} = 0$. But the reverse is not necessarily true.

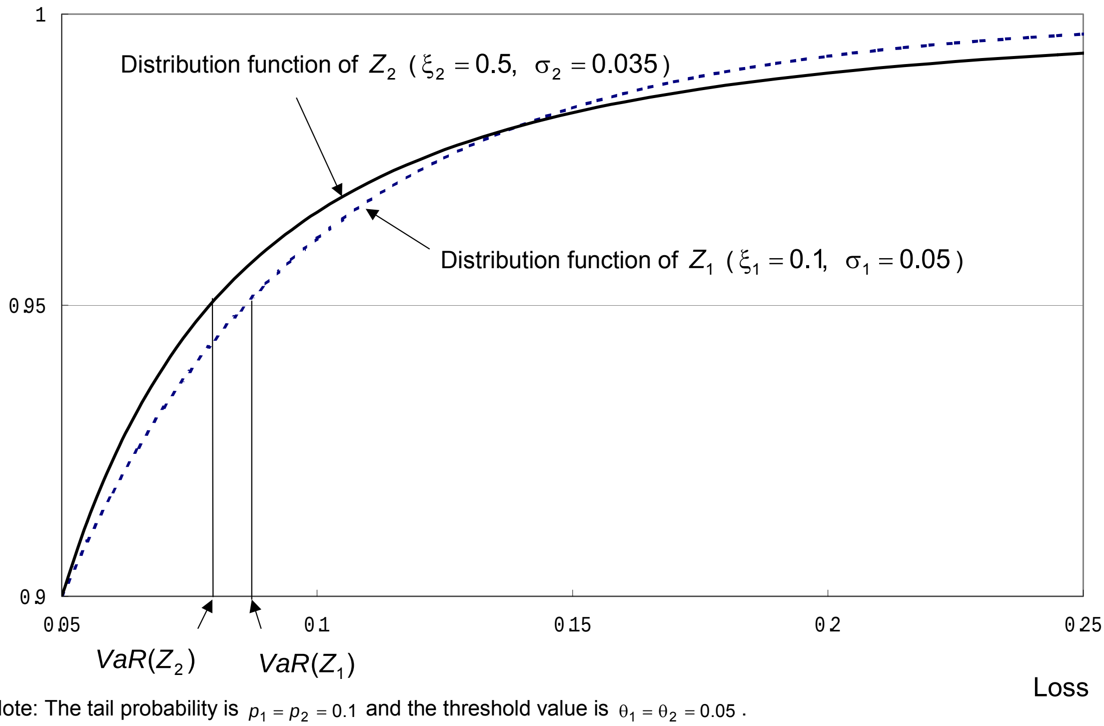
Table 4

Properties of the copulas used in this paper

	Equation	Dependence structure	χ	$\overline{\chi}$
Gumbel	$C(u,v)=\exp\{-[(-\log u)^{\alpha}+(-\log v)^{\alpha}]^{1/\alpha}\}$	Independent when $\alpha=1$ Fully dependent when $\alpha=\infty$	$\chi=2-2^{1/\alpha}$ ($\alpha\geq 1$)	$\overline{\chi}=1$
Gaussian	$C(u,v)=\Phi_{\rho}(\Phi^{-1}(u),\Phi^{-1}(v))$	Independent when $\rho=0$ Fully dependent when $\rho=\pm 1$	$\chi=0$ $(-1<\rho<1)$	$\overline{\chi}=\rho$
Frank	$C(u,v)=-\frac{1}{\delta}\ln\left(\frac{1-e^{-\delta}-(1-e^{-\delta u})(1-e^{-\delta v})}{1-e^{-\delta}}\right)$	Independent when $\delta=0$ Fully dependent when $\delta=\pm\infty$	$\chi=0$	$\overline{\chi}=0$

Figure 6

Example plot of the distribution of exceedances



Note: The tail probability is $p_1=p_2=0.1$ and the threshold value is $\theta_1=\theta_2=0.05$.

Table 5

Threshold value $\overline{\kappa}_{VaR}$ for the tail risk of VaR

(Tail probability: p = 0.1, confidence level: 95%)

		ξ_1									
		0.10	**0.20**	**0.30**	**0.40**	**0.50**	**0.60**	**0.70**	**0.80**	**0.90**	**1.00**
ξ_2	0.10	–	–	–	–	–	–	–	–	–	–
	0.20	1.036	–	–	–	–	–	–	–	–	–
	0.30	1.073	1.036	–	–	–	–	–	–	–	–
	0.40	1.113	1.074	1.037	–	–	–	–	–	–	–
	0.50	1.154	1.114	1.075	1.037	–	–	–	–	–	–
	0.60	1.198	1.156	1.116	1.076	1.038	–	–	–	–	–
	0.70	1.243	1.200	1.158	1.117	1.077	1.038	–	–	–	–
	0.80	1.291	1.246	1.202	1.160	1.118	1.078	1.038	–	–	–
	0.90	1.341	1.294	1.249	1.205	1.162	1.120	1.079	1.039	–	–
	1.00	1.393	1.345	1.298	1.252	1.207	1.163	1.121	1.079	1.039	–

(Tail probability: p = 0.1, confidence level: 99%)

		ξ_1									
		0.10	**0.20**	**0.30**	**0.40**	**0.50**	**0.60**	**0.70**	**0.80**	**0.90**	**1.00**
ξ_2	0.10	–	–	–	–	–	–	–	–	–	–
	0.20	1.129	–	–	–	–	–	–	–	–	–
	0.30	1.281	1.134	–	–	–	–	–	–	–	–
	0.40	1.460	1.292	1.139	–	–	–	–	–	–	–
	0.50	1.670	1.479	1.304	1.144	–	–	–	–	–	–
	0.60	1.919	1.699	1.498	1.315	1.149	–	–	–	–	–
	0.70	2.213	1.960	1.728	1.516	1.325	1.154	–	–	–	–
	0.80	2.563	2.269	2.001	1.756	1.535	1.336	1.158	–	–	–
	0.90	2.980	2.638	2.325	2.041	1.784	1.553	1.346	1.162	–	–
	1.00	3.476	3.077	2.713	2.381	2.081	1.811	1.570	1.356	1.167	–

(Tail probability: p = 0.05, confidence level: 99%)

		ξ_1									
		0.10	**0.20**	**0.30**	**0.40**	**0.50**	**0.60**	**0.70**	**0.80**	**0.90**	**1.00**
ξ_2	0.10	–	–	–	–	–	–	–	–	–	–
	0.20	1.087	–	–	–	–	–	–	–	–	–
	0.30	1.185	1.090	–	–	–	–	–	–	–	–
	0.40	1.294	1.190	1.092	–	–	–	–	–	–	–
	0.50	1.416	1.302	1.195	1.094	–	–	–	–	–	–
	0.60	1.552	1.428	1.310	1.200	1.097	–	–	–	–	–
	0.70	1.706	1.569	1.440	1.319	1.205	1.099	–	–	–	–
	0.80	1.878	1.727	1.585	1.452	1.327	1.210	1.101	–	–	–
	0.90	2.072	1.906	1.749	1.602	1.464	1.335	1.215	1.103	–	–
	1.00	2.291	2.107	1.933	1.771	1.618	1.476	1.343	1.220	1.105	–

Note: VaR has tail risk when σ_1/σ_2 is more than $\overline{\kappa}_{VaR}$.

Figure 7

Varied scale parameters and the tail risk of VaR

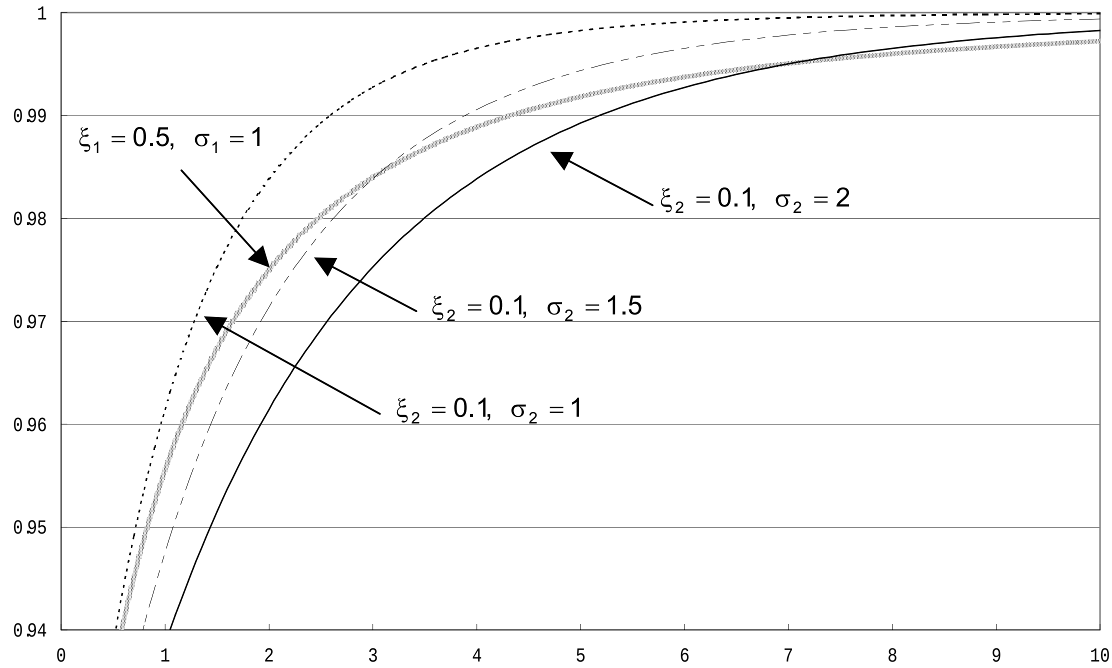


Figure 8

Varied tail indices and the tail risk of VaR

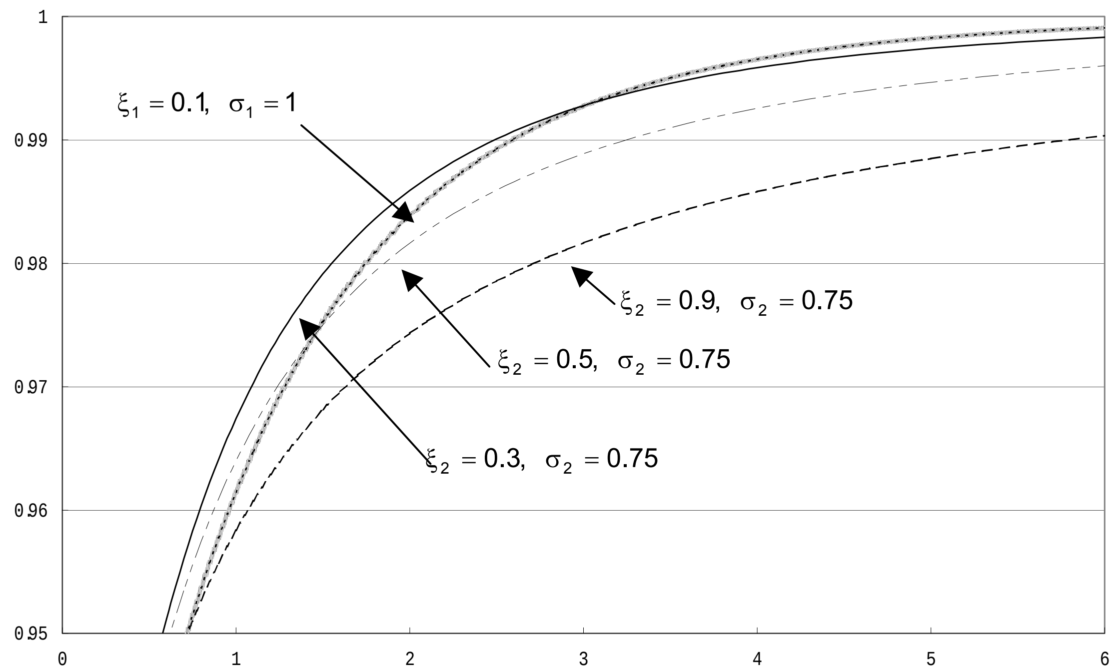


Table 6

Threshold value $\overline{\kappa}_{ES}$ for the tail risk of expected shortfall

(Tail probability: $p = 0.1$, confidence level: 95%)

		ξ_1									
		0.10	**0.20**	**0.30**	**0.40**	**0.50**	**0.60**	**0.70**	**0.80**	**0.90**	**1.00**
ξ_2	0.10	–	–	–	–	–	–	–	–	–	–
	0.20	1.142	–	–	–	–	–	–	–	–	–
	0.30	1.325	1.161	–	–	–	–	–	–	–	–
	0.40	1.571	1.376	1.185	–	–	–	–	–	–	–
	0.50	1.916	1.678	1.446	1.220	–	–	–	–	–	–
	0.60	2.436	2.133	1.838	1.551	1.271	–	–	–	–	–
	0.70	3.305	2.894	2.494	2.104	1.725	1.357	–	–	–	–
	0.80	5.047	4.420	3.808	3.213	2.634	2.072	1.527	–	–	–
	0.90	10.281	9.004	7.758	6.545	5.366	4.221	3.111	2.037	–	–
	1.00	–	–	–	–	–	–	–	–	–	–

(Tail probability: p = 0.1, confidence level: 99%)

		ξ_1									
		0.10	**0.20**	**0.30**	**0.40**	**0.50**	**0.60**	**0.70**	**0.80**	**0.90**	**1.00**
ξ_2	0.10	–	–	–	–	–	–	–	–	–	–
	0.20	1.230	–	–	–	–	–	–	–	–	–
	0.30	1.547	1.257	–	–	–	–	–	–	–	–
	0.40	1.998	1.624	1.292	–	–	–	–	–	–	–
	0.50	2.670	2.171	1.727	1.337	–	–	–	–	–	–
	0.60	3.741	3.042	2.419	1.873	1.401	–	–	–	–	–
	0.70	5.626	4.574	3.638	2.817	2.107	1.504	–	–	–	–
	0.80	9.575	7.784	6.191	4.793	3.586	2.559	1.702	–	–	–
	0.90	21.852	17.765	14.129	10.940	8.184	5.841	3.884	2.282	–	–
	1.00	–	–	–	–	–	–	–	–	–	–

(Tail probability: p = 0.05, confidence level: 99%)

		ξ_1									
		0.10	**0.20**	**0.30**	**0.40**	**0.50**	**0.60**	**0.70**	**0.80**	**0.90**	**1.00**
ξ_2	0.10	–	–	–	–	–	–	–	–	–	–
	0.20	1.187	–	–	–	–	–	–	–	–	–
	0.30	1.437	1.210	–	–	–	–	–	–	–	–
	0.40	1.780	1.499	1.239	–	–	–	–	–	–	–
	0.50	2.276	1.917	1.584	1.278	–	–	–	–	–	–
	0.60	3.040	2.560	2.116	1.708	1.336	–	–	–	–	–
	0.70	4.347	3.660	3.025	2.442	1.910	1.430	–	–	–	–
	0.80	7.013	5.906	4.881	3.940	3.082	2.307	1.613	–	–	–
	0.90	15.136	12.747	10.535	8.503	6.651	4.978	3.482	2.158	–	–
	1.00	–	–	–	–	–	–	–	–	–	–

Note: Expected shortfall has tail risk when σ_1/σ_2 is more than $\overline{\kappa}_{ES}$. When $\xi = 1$, we are unable to calculate expected shortfall as the first moment diverges.

Figure 9

Upward bias when using exceedances for risk measurement

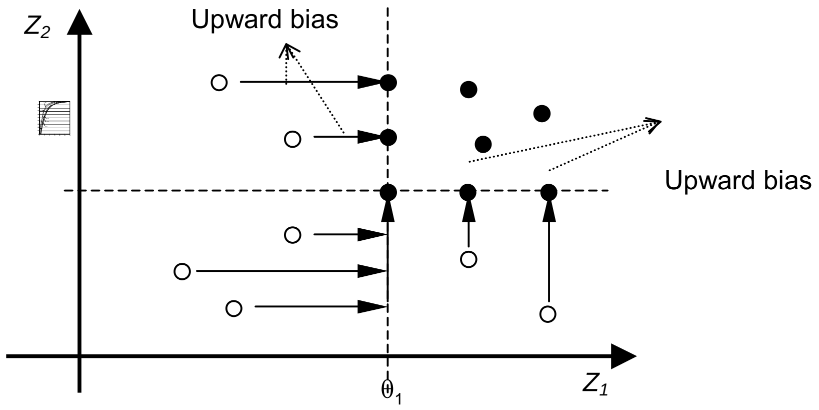


Figure 10

The marginal distribution function assumed in this paper

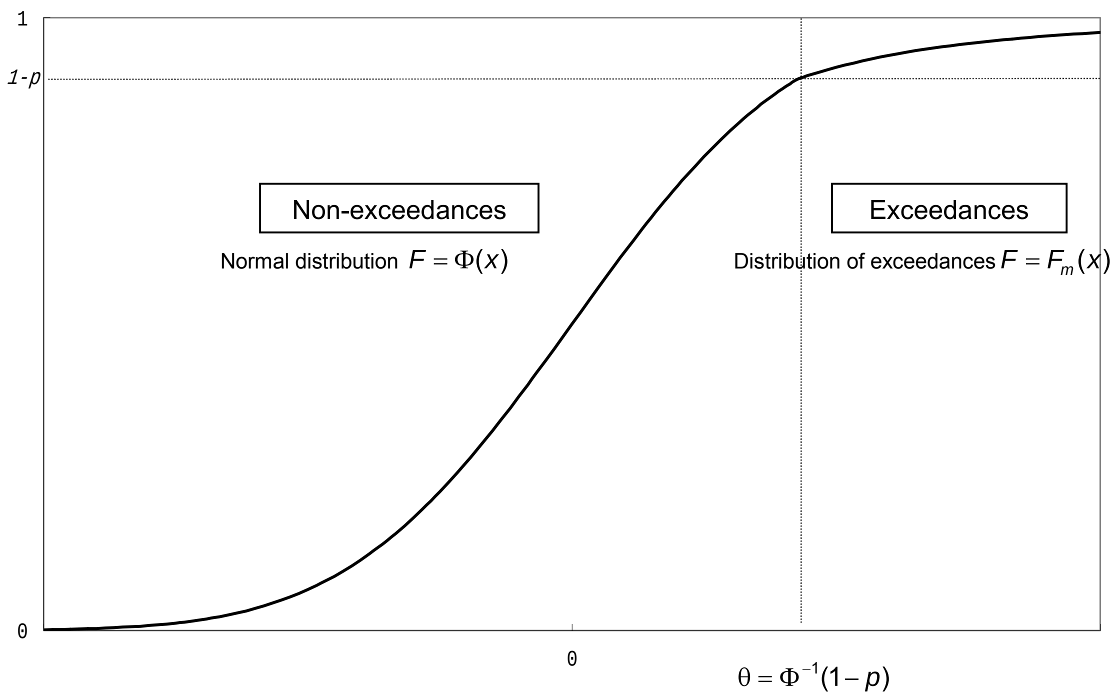
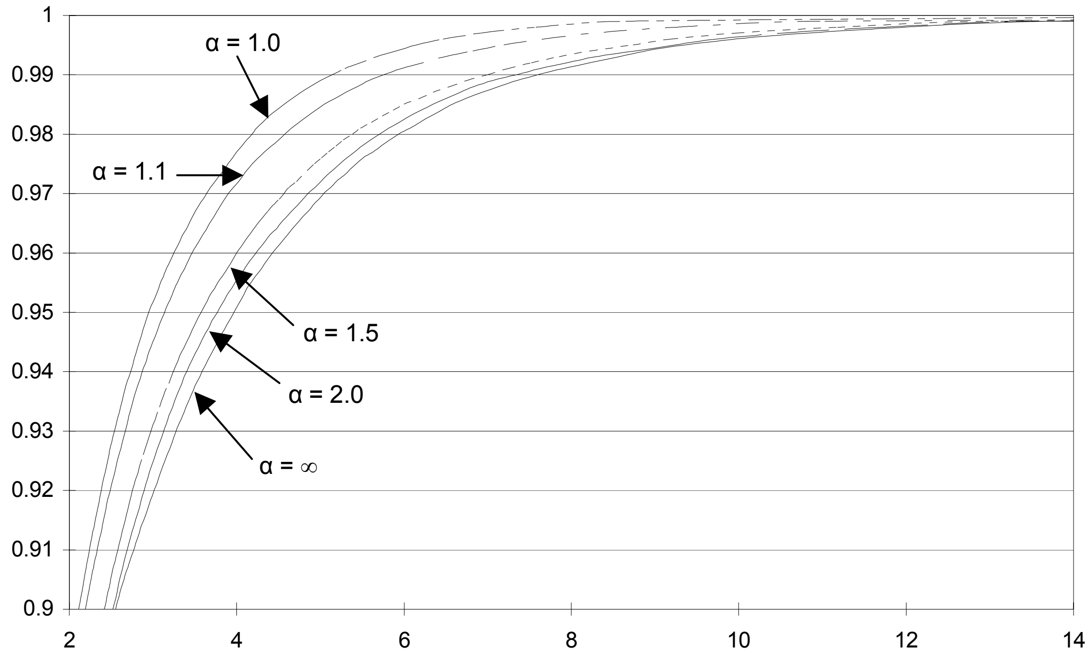


Figure 11

Empirical distribution functions of the sums under the Gumbel copula



Note: Empirical distributions are plotted from one million simulations with the marginal distribution parameters set at $\xi = 0.5$, $\sigma = 1$, $p = 0.1$.

Table 7

VaR and expected shortfall under changes in the dependence parameter using a specific copula

Gumbel	$\xi = 0.1$					
α	**VaR (95%)**	**VaR (99%)**	**VaR (99.9%)**	**ES (95%)**	**ES (99%)**	**ES (99.9%)**
1.0	2.971	5.165	8.748	4.357	6.715	10.670
1.1	3.150	5.777	10.724	4.852	7.915	13.702
1.2	3.299	6.252	11.822	5.189	8.623	14.974
1.3	3.412	6.563	12.429	5.425	9.071	15.676
1.4	3.505	6.798	12.861	5.597	9.374	16.117
1.5	3.577	6.980	13.111	5.725	9.586	16.410
1.6	3.634	7.087	13.295	5.822	9.740	16.615
1.7	3.682	7.178	13.417	5.898	9.857	16.767
1.8	3.718	7.247	13.485	5.958	9.948	16.886
1.9	3.748	7.307	13.547	6.007	10.020	16.983
2.0	3.772	7.357	13.602	6.048	10.078	17.060
5.0	3.957	7.672	13.966	6.311	10.417	17.561
10.0	3.981	7.694	14.033	6.342	10.456	17.595
∞	3.993	7.703	14.219	6.352	10.502	17.613
Gaussian	$\xi = 0.1$					
ρ	**VaR (95%)**	**VaR (99%)**	**VaR (99.9%)**	**ES (95%)**	**ES (99%)**	**ES (99.9%)**
0	2.971	5.165	8.748	4.357	6.715	10.670
0.1	3.124	5.435	9.275	4.585	7.086	11.257
0.2	3.250	5.687	9.747	4.786	7.423	11.842
0.3	3.366	5.932	10.262	4.986	7.770	12.473
0.4	3.476	6.180	10.798	5.183	8.129	13.159
0.5	3.576	6.424	11.324	5.380	8.505	13.891
0.6	3.671	6.671	11.939	5.577	8.898	14.663
0.7	3.761	6.923	12.507	5.775	9.309	15.464
0.8	3.842	7.198	13.132	5.978	9.736	16.288
0.9	3.921	7.501	13.727	6.189	10.172	17.149
1	3.993	7.703	14.219	6.352	10.502	17.613
Frank	$\xi = 0.1$					
δ	**VaR (95%)**	**VaR (99%)**	**VaR (99.9%)**	**ES (95%)**	**ES (99%)**	**ES (99.9%)**
0	2.971	5.165	8.748	4.357	6.715	10.670
1	3.171	5.438	9.071	4.600	7.017	11.025
2	3.348	5.687	9.392	4.817	7.290	11.344
3	3.492	5.901	9.656	5.000	7.524	11.618
4	3.607	6.074	9.875	5.153	7.720	11.852
5	3.699	6.226	10.056	5.278	7.884	12.049
6	3.770	6.349	10.217	5.380	8.022	12.218
7	3.828	6.451	10.362	5.466	8.141	12.363
8	3.874	6.539	10.484	5.538	8.245	12.489
9	3.914	6.614	10.599	5.600	8.337	12.601
∞	3.993	7.703	14.219	6.352	10.502	17.613

Note: VaR and expected shortfall are calculated from one million simulations for each copula with the marginal distribution parameters set at $\sigma = 1$, $p = 0.1$. The tail index values are shown in the upper left of each table.

Table 7 (cont)

Gumbel	$\xi = 0.25$					
α	**VaR (95%)**	**VaR (99%)**	**VaR (99.9%)**	**ES (95%)**	**ES (99%)**	**ES (99.9%)**
1.0	3.125	6.065	12.465	5.083	8.858	17.463
1.1	3.302	6.694	14.824	5.595	10.170	21.106
1.2	3.437	7.162	16.085	5.949	10.994	23.018
1.3	3.543	7.501	16.986	6.200	11.538	24.174
1.4	3.628	7.745	17.557	6.384	11.920	24.944
1.5	3.696	7.920	18.004	6.521	12.195	25.479
1.6	3.750	8.049	18.214	6.626	12.398	25.863
1.7	3.792	8.152	18.429	6.708	12.554	26.154
1.8	3.827	8.231	18.594	6.773	12.675	26.383
1.9	3.852	8.284	18.652	6.827	12.773	26.568
2.0	3.874	8.339	18.732	6.871	12.852	26.718
5.0	4.036	8.699	19.286	7.159	13.330	27.802
10.0	4.059	8.726	19.414	7.194	13.388	27.911
∞	4.071	8.735	19.778	7.206	13.454	27.837
Gaussian	$\xi = 0.25$					
ρ	**VaR (95%)**	**VaR (99%)**	**VaR (99.9%)**	**ES (95%)**	**ES (99%)**	**ES (99.9%)**
0	3.125	6.065	12.465	5.083	8.858	17.463
0.1	3.288	6.354	13.068	5.330	9.284	18.200
0.2	3.412	6.618	13.669	5.542	9.657	18.876
0.3	3.529	6.886	14.259	5.753	10.051	19.682
0.4	3.635	7.152	14.947	5.964	10.468	20.593
0.5	3.730	7.412	15.689	6.176	10.914	21.629
0.6	3.819	7.667	16.531	6.388	11.395	22.804
0.7	3.900	7.938	17.371	6.602	11.913	24.111
0.8	3.967	8.218	18.229	6.822	12.469	25.539
0.9	4.027	8.541	19.083	7.052	13.058	27.123
1	4.071	8.735	19.778	7.206	13.454	27.837
Frank	$\xi = 0.25$					
δ	**VaR (95%)**	**VaR (99%)**	**VaR (99.9%)**	**ES (95%)**	**ES (99%)**	**ES (99.9%)**
0	3.125	6.065	12.465	5.083	8.858	17.463
1	3.328	6.345	12.869	5.335	9.180	17.847
2	3.506	6.608	13.170	5.561	9.478	18.210
3	3.654	6.847	13.453	5.755	9.739	18.531
4	3.770	7.034	13.740	5.916	9.960	18.803
5	3.863	7.202	14.000	6.050	10.145	19.037
6	3.935	7.340	14.168	6.159	10.302	19.237
7	3.991	7.451	14.308	6.250	10.437	19.409
8	4.035	7.554	14.468	6.328	10.556	19.566
9	4.071	7.641	14.598	6.394	10.662	19.705
∞	4.071	8.735	19.778	7.206	13.454	27.837

Note: VaR and expected shortfall are calculated from one million simulations for each copula with the marginal distribution parameters set at $\sigma = 1$, $p = 0.1$. The tail index values are shown in the upper left of each table.

Table 7 (cont)

Gumbel	$\xi = 0.5$					
α	**VaR (95%)**	**VaR (99%)**	**VaR (99.9%)**	**ES (95%)**	**ES (99%)**	**ES (99.9%)**
1.0	3.442	8.441	27.131	7.419	17.092	53.729
1.1	3.595	9.024	30.316	7.929	18.507	57.999
1.2	3.715	9.501	31.524	8.310	19.550	61.639
1.3	3.800	9.850	32.876	8.585	20.293	64.136
1.4	3.873	10.078	34.013	8.789	20.839	65.995
1.5	3.927	10.268	34.691	8.942	21.249	67.384
1.6	3.972	10.398	35.156	9.060	21.563	68.453
1.7	4.005	10.501	35.501	9.153	21.811	69.273
1.8	4.033	10.576	35.800	9.229	22.007	69.936
1.9	4.051	10.632	35.911	9.290	22.168	70.512
2.0	4.068	10.682	36.003	9.341	22.301	70.991
5.0	4.186	11.084	36.846	9.701	23.260	75.714
10.0	4.203	11.106	37.187	9.759	23.447	76.842
∞	4.213	11.115	38.301	9.755	23.448	75.100
Gaussian	$\xi = 0.5$					
ρ	**VaR (95%)**	**VaR (99%)**	**VaR (99.9%)**	**ES (95%)**	**ES (99%)**	**ES (99.9%)**
0	3.442	8.441	27.131	7.419	17.092	53.729
0.1	3.615	8.803	28.052	7.728	17.747	55.729
0.2	3.739	9.077	28.675	7.968	18.231	57.090
0.3	3.851	9.379	29.438	8.209	18.763	58.693
0.4	3.949	9.679	30.552	8.451	19.337	60.525
0.5	4.037	9.943	31.695	8.693	19.947	62.477
0.6	4.106	10.216	32.864	8.934	20.614	64.683
0.7	4.167	10.481	34.683	9.176	21.358	67.279
0.8	4.207	10.753	36.224	9.425	22.204	70.588
0.9	4.230	11.062	37.467	9.691	23.159	74.816
1	4.213	11.115	38.301	9.755	23.448	75.100
Frank	$\xi = 0.5$					
δ	**VaR (95%)**	**VaR (99%)**	**VaR (99.9%)**	**ES (95%)**	**ES (99%)**	**ES (99.9%)**
0	3.442	8.441	27.131	7.419	17.092	53.729
1	3.643	8.751	27.474	7.686	17.449	54.247
2	3.821	9.042	27.927	7.930	17.793	54.692
3	3.973	9.299	28.258	8.141	18.105	55.133
4	4.093	9.521	28.649	8.318	18.375	55.491
5	4.185	9.691	29.054	8.465	18.601	55.791
6	4.255	9.861	29.387	8.587	18.792	56.074
7	4.308	10.004	29.730	8.688	18.955	56.312
8	4.351	10.110	29.853	8.774	19.101	56.522
9	4.382	10.212	29.870	8.847	19.233	56.723
∞	4.213	11.115	38.301	9.755	23.448	75.100

Note: VaR and expected shortfall are calculated from one million simulations for each copula with the marginal distribution parameters set at σ = 1, *p* = 0.1. The tail index values are shown in the upper left of each table.

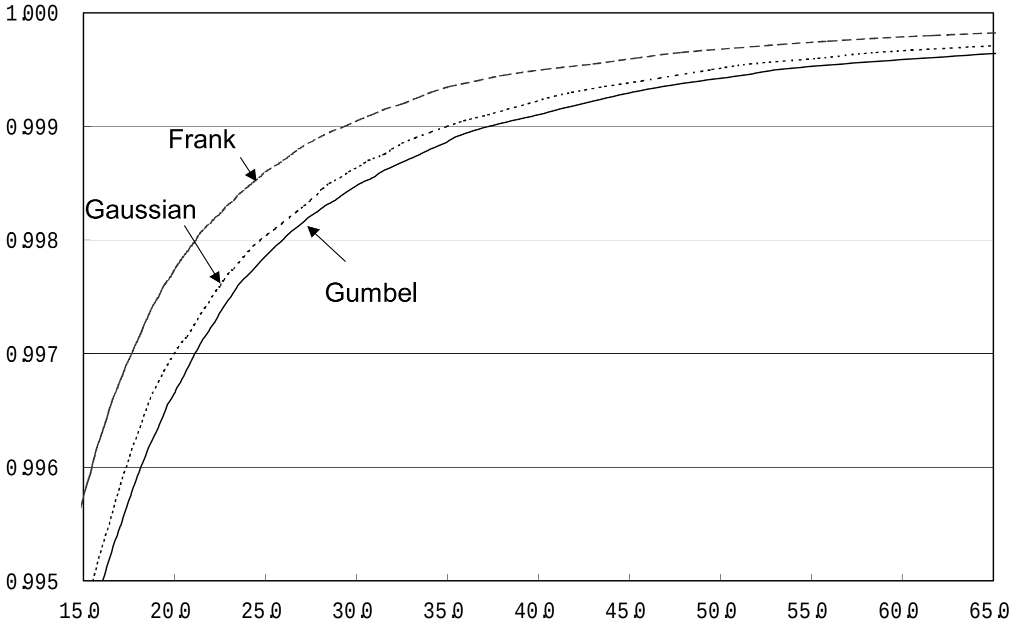
Table 7 (cont)						
Gumbel	$\xi = 0.75$					
α	**VaR (95%)**	**VaR (99%)**	**VaR (99.9%)**	**ES (95%)**	**ES (99%)**	**ES (99.9%)**
1.0	3.847	12.654	68.724	14.106	45.232	232.931
1.1	3.961	13.076	73.107	14.193	44.817	220.165
1.2	4.054	13.468	74.485	14.574	46.065	226.977
1.3	4.117	13.752	74.705	14.878	47.107	232.902
1.4	4.167	13.980	75.957	15.110	47.934	237.781
1.5	4.209	14.130	78.154	15.288	48.578	241.740
1.6	4.243	14.277	77.773	15.427	49.087	244.924
1.7	4.263	14.314	78.758	15.540	49.504	247.554
1.8	4.278	14.362	79.165	15.633	49.861	249.744
1.9	4.291	14.373	79.241	15.713	50.164	251.761
2.0	4.302	14.380	78.839	15.781	50.431	253.590
5.0	4.355	14.716	78.988	16.542	53.710	282.245
10.0	4.364	14.714	80.040	16.844	55.155	295.725
∞	4.373	14.720	83.395	16.517	53.579	275.707
Gaussian	$\xi = 0.75$					
ρ	**VaR (95%)**	**VaR (99%)**	**VaR (99.9%)**	**ES (95%)**	**ES (99%)**	**ES (99.9%)**
0	3.847	12.654	68.724	14.106	45.232	232.931
0.1	4.026	13.186	70.836	14.668	47.050	243.941
0.2	4.145	13.434	71.028	15.092	48.451	254.063
0.3	4.254	13.751	72.412	15.531	49.982	265.049
0.4	4.344	14.094	74.657	15.921	51.324	273.791
0.5	4.411	14.387	77.344	16.217	52.268	278.229
0.6	4.468	14.556	78.944	16.429	52.907	279.463
0.7	4.493	14.736	81.197	16.610	53.526	280.122
0.8	4.500	14.931	83.456	16.802	54.359	283.397
0.9	4.468	15.092	84.647	17.040	55.548	291.229
1	4.373	14.720	83.395	16.517	53.579	275.707
Frank	$\xi = 0.75$					
δ	**VaR (95%)**	**VaR (99%)**	**VaR (99.9%)**	**ES (95%)**	**ES (99%)**	**ES (99.9%)**
0	3.847	12.654	68.724	14.106	45.232	232.931
1	4.051	12.988	68.816	14.397	45.680	234.046
2	4.229	13.318	69.598	14.665	46.116	234.846
3	4.376	13.620	70.071	14.897	46.507	235.493
4	4.494	13.879	70.484	15.091	46.843	235.963
5	4.580	14.069	70.999	15.251	47.117	236.298
6	4.650	14.258	71.637	15.383	47.344	236.603
7	4.703	14.398	73.037	15.493	47.537	236.907
8	4.739	14.515	72.559	15.587	47.708	237.163
9	4.767	14.634	72.669	15.669	47.873	237.456
∞	4.373	14.720	83.395	16.517	53.579	275.707

Note: VaR and expected shortfall are calculated from one million simulations for each copula with the marginal distribution parameters set at $\sigma = 1$, $p = 0.1$. The tail index values are shown in the upper left of each table.

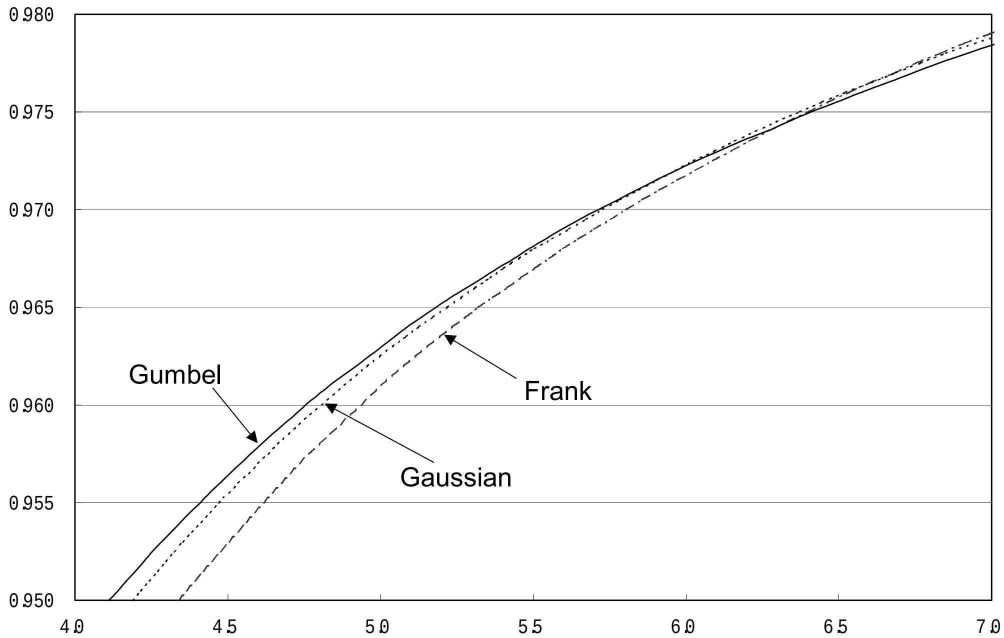
Figure 12

Empirical distributions under Gumbel, Gaussian and Frank copulas

(Portion with a cumulative probability of at least 99.5%)



(Portion with a cumulative probability of 95–98%)



Note: The marginal distribution parameters are set at $\xi = 0.5$, $\sigma = 1$, $p = 0.1$. The empirical distributions are generated by conducting one million simulations for each copula. For all of the copula parameters, the Spearman's rho is set at $\rho_S = 0.5$.

Table 8

VaR and expected shortfall under different copulas

$\xi = 0.1$

	VaR (95%)	VaR (99%)	VaR (99.9%)	ES (95%)	ES (99%)	ES (99.9%)
Independent	2.971	5.165	8.748	4.357	6.715	10.670
Fully dependent	3.993	7.703	14.219	6.352	10.502	17.613

Spearman's rho=0.2

	VaR (95%)	VaR (99%)	VaR (99.9%)	ES (95%)	ES (99%)	ES (99.9%)
Frank	3.212	5.493	9.152	4.651	7.080	11.098
Gaussian	3.261	5.709	9.784	4.804	7.454	11.897
Gumbel	3.245	6.080	11.426	5.069	8.381	14.566

Spearman's rho=0.5

	VaR (95%)	VaR (99%)	VaR (99.9%)	ES (95%)	ES (99%)	ES (99.9%)
Frank	3.547	5.982	9.770	5.073	7.617	11.728
Gaussian	3.594	6.463	11.425	5.416	8.575	14.027
Gumbel	3.601	7.024	13.184	5.766	9.653	16.500

Spearman's rho=0.8

	VaR (95%)	VaR (99%)	VaR (99.9%)	ES (95%)	ES (99%)	ES (99.9%)
Frank	3.869	6.529	10.478	5.531	8.235	12.477
Gaussian	3.851	7.236	13.207	6.005	9.792	16.399
Gumbel	3.858	7.526	13.836	6.185	10.261	17.312

$\xi = 0.25$

	VaR (95%)	VaR (99%)	VaR (99.9%)	ES (95%)	ES (99%)	ES (99.9%)
Independent	3.125	6.065	12.465	5.083	8.858	17.463
Fully dependent	4.071	8.735	19.778	7.206	13.454	27.837

Spearman's rho=0.2

	VaR (95%)	VaR (99%)	VaR (99.9%)	ES (95%)	ES (99%)	ES (99.9%)
Frank	3.369	6.403	12.914	5.387	9.248	17.927
Gaussian	3.422	6.643	13.728	5.561	9.691	18.944
Gumbel	3.389	6.988	15.598	5.822	10.707	22.383

Spearman's rho=0.5

	VaR (95%)	VaR (99%)	VaR (99.9%)	ES (95%)	ES (99%)	ES (99.9%)
Frank	3.711	6.934	13.600	5.831	9.843	18.659
Gaussian	3.747	7.455	15.861	6.214	10.998	21.830
Gumbel	3.720	7.979	18.086	6.566	12.284	25.647

Spearman's rho=0.8

	VaR (95%)	VaR (99%)	VaR (99.9%)	ES (95%)	ES (99%)	ES (99.9%)
Frank	4.031	7.544	14.456	6.320	10.544	19.551
Gaussian	3.974	8.263	18.334	6.851	12.544	25.735
Gumbel	3.949	8.526	19.090	7.020	13.106	27.229

Note: VaR and expected shortfall are calculated by conducting one million simulations for each copula. The marginal distribution parameters are set at $\sigma = 1$, $p = 0.1$.

Table 8 (cont)

$\xi = 0.5$

	VaR (95%)	VaR (99%)	VaR (99.9%)	ES (95%)	ES (99%)	ES (99.9%)
Independent	3.442	8.441	27.131	7.419	17.092	53.729
Fully dependent	4.213	11.115	38.301	9.755	23.448	75.100

Spearman's rho=0.2

	VaR (95%)	VaR (99%)	VaR (99.9%)	ES (95%)	ES (99%)	ES (99.9%)
Frank	3.684	8.812	27.657	7.742	17.527	54.353
Gaussian	3.748	9.105	28.750	7.989	18.277	57.226
Gumbel	3.672	9.325	31.444	8.172	19.177	60.376

Spearman's rho=0.5

	VaR (95%)	VaR (99%)	VaR (99.9%)	ES (95%)	ES (99%)	ES (99.9%)
Frank	4.031	9.407	28.422	8.224	18.232	55.300
Gaussian	4.052	9.988	31.896	8.736	20.062	62.854
Gumbel	3.947	10.332	34.770	8.993	21.384	67.848

Spearman's rho=0.8

	VaR (95%)	VaR (99%)	VaR (99.9%)	ES (95%)	ES (99%)	ES (99.9%)
Frank	4.347	10.100	29.790	8.766	19.087	56.502
Gaussian	4.211	10.798	36.249	9.459	22.322	71.084
Gumbel	4.119	10.888	36.572	9.518	22.757	72.817

$\xi = 0.75$

	VaR (95%)	VaR (99%)	VaR (99.9%)	ES (95%)	ES (99%)	ES (99.9%)
Independent	3.847	12.654	68.724	14.106	45.232	232.931
Fully dependent	4.373	14.720	83.395	16.517	53.579	275.707

Spearman's rho=0.2

	VaR (95%)	VaR (99%)	VaR (99.9%)	ES (95%)	ES (99%)	ES (99.9%)
Frank	4.092	13.022	69.291	14.459	45.778	234.268
Gaussian	4.157	13.465	71.011	15.131	48.589	255.050
Gumbel	4.028	13.288	73.602	14.429	45.581	224.180

Spearman's rho=0.5

	VaR (95%)	VaR (99%)	VaR (99.9%)	ES (95%)	ES (99%)	ES (99.9%)
Frank	4.433	13.722	70.411	14.989	46.666	235.724
Gaussian	4.424	14.411	77.312	16.260	52.397	278.633
Gumbel	4.222	14.188	79.041	15.348	48.795	243.099

Spearman's rho=0.8

	VaR (95%)	VaR (99%)	VaR (99.9%)	ES (95%)	ES (99%)	ES (99.9%)
Frank	4.737	14.512	72.461	15.578	47.691	237.134
Gaussian	4.496	14.932	83.944	16.830	54.489	284.102
Gumbel	4.326	14.549	80.106	16.057	51.537	261.953

Note: VaR and expected shortfall are calculated by conducting one million simulations for each copula. The marginal distribution parameters are set at $\sigma = 1$, $p = 0.1$.

Table 9

VaR and expected shortfall under different copulas for different marginal distributions (example)

($\xi_1 = 0.1, \quad \xi_2 = 0.1, \quad \sigma_1 = 1, \quad \sigma_2 = 2, \quad \rho_S = 0.2, \quad p = 0.1$ **)**

	VaR (95%)	**VaR (99%)**	**VaR (99.9%)**	**ES (95%)**	**ES (99%)**	**ES (99.9%)**
Frank	3.8542	7.4521	13.7438	6.1475	10.1535	17.1047
Gaussian	3.8806	7.7226	14.3812	6.3062	10.5660	17.7964
Gumbel	3.8569	8.1702	16.3774	6.6234	11.7285	21.3039

Table 10

Estimation of the parameters of the distribution of exceedances of daily log changes of the foreign exchange rates (per one US dollar)

Developed countries

Japan (yen)

% of excess	ξ	σ	Threshold	VaR (95%)	Expected shortfall (95%)	VaR (99%)	Expected shortfall (99%)
Right tail							
1%	– 0.3988	0.0085	0.0169	– 0.0024	0.0092	0.0169	0.0230
2%	0.0485	0.0048	0.0141	0.0097	0.0146	0.0174	0.0226
3%	– 0.0169	0.0054	0.0117	0.0090	0.0143	0.0176	0.0228
4%	0.1482	0.0040	0.0110	0.0101	0.0146	0.0171	0.0228
5%	0.1126	0.0039	0.0102	0.0102	0.0145	0.0170	0.0223
6%	0.0767	0.0042	0.0092	0.0100	0.0146	0.0173	0.0225
7%	0.0767	0.0042	0.0086	0.0100	0.0146	0.0173	0.0225
8%	0.0950	0.0039	0.0081	0.0100	0.0146	0.0172	0.0224
9%	0.0761	0.0041	0.0076	0.0100	0.0146	0.0173	0.0225
10%	0.0796	0.0040	0.0072	0.0100	0.0146	0.0173	0.0225
Left tail							
1%	– 0.0162	0.0094	0.0199	0.0045	0.0140	0.0198	0.0290
2%	0.0875	0.0071	0.0157	0.0093	0.0166	0.0207	0.0290
3%	0.0996	0.0067	0.0128	0.0095	0.0166	0.0206	0.0289
4%	0.1083	0.0064	0.0111	0.0097	0.0166	0.0206	0.0289
5%	0.1880	0.0054	0.0101	0.0101	0.0167	0.0202	0.0292
6%	0.1647	0.0054	0.0091	0.0101	0.0168	0.0204	0.0291
7%	0.1484	0.0053	0.0083	0.0101	0.0166	0.0202	0.0285
8%	0.1603	0.0052	0.0075	0.0101	0.0167	0.0204	0.0290
9%	0.2138	0.0046	0.0072	0.0101	0.0167	0.0201	0.0295
10%	0.1848	0.0048	0.0066	0.0102	0.0168	0.0203	0.0293

Germany (mark)

% of excess	ξ	σ	Threshold	VaR (95%)	Expected shortfall (95%)	VaR (99%)	Expected shortfall (99%)
Right tail							
1%	0.0484	0.0039	0.0146	0.0085	0.0123	0.0146	0.0187
2%	0.0642	0.0036	0.0122	0.0089	0.0126	0.0147	0.0187
3%	0.0633	0.0035	0.0107	0.0090	0.0126	0.0147	0.0187
4%	0.0596	0.0034	0.0097	0.0090	0.0126	0.0147	0.0187
5%	0.0741	0.0033	0.0091	0.0091	0.0126	0.0147	0.0187
6%	0.0443	0.0035	0.0083	0.0090	0.0126	0.0148	0.0187
7%	– 0.0326	0.0040	0.0075	0.0088	0.0127	0.0151	0.0187
8%	– 0.0876	0.0045	0.0067	0.0088	0.0128	0.0153	0.0188
9%	– 0.0482	0.0042	0.0064	0.0088	0.0127	0.0151	0.0188
10%	– 0.0496	0.0042	0.0059	0.0088	0.0127	0.0151	0.0188
Left tail							
1%	– 0.0958	0.0045	0.0153	0.0073	0.0122	0.0152	0.0194
2%	0.0365	0.0038	0.0127	0.0093	0.0130	0.0153	0.0193
3%	– 0.0024	0.0040	0.0109	0.0088	0.0129	0.0153	0.0194
4%	– 0.0721	0.0046	0.0094	0.0084	0.0128	0.0156	0.0195
5%	– 0.0334	0.0044	0.0086	0.0086	0.0128	0.0154	0.0194
6%	– 0.0045	0.0041	0.0080	0.0087	0.0128	0.0153	0.0194
7%	0.0137	0.0040	0.0074	0.0088	0.0128	0.0153	0.0194
8%	0.0029	0.0040	0.0069	0.0088	0.0128	0.0153	0.0194
9%	– 0.0275	0.0043	0.0063	0.0088	0.0129	0.0154	0.0193
10%	– 0.0226	0.0043	0.0058	0.0088	0.0129	0.0154	0.0194

Table 10 (cont)

UK (pound)

% of excess	ξ	σ	Threshold	VaR (95%)	Expected shortfall (95%)	VaR (99%)	Expected shortfall (99%)
Right tail							
1%	– 0.0461	0.0029	0.0114	0.0066	0.0096	0.0114	0.0142
2%	0.0777	0.0022	0.0102	0.0082	0.0104	0.0117	0.0142
3%	– 0.0782	0.0030	0.0087	0.0071	0.0100	0.0118	0.0144
4%	– 0.1037	0.0032	0.0077	0.0070	0.0100	0.0119	0.0144
5%	– 0.1188	0.0035	0.0068	0.0068	0.0100	0.0120	0.0146
6%	– 0.1120	0.0035	0.0062	0.0068	0.0100	0.0119	0.0145
7%	– 0.1160	0.0036	0.0056	0.0068	0.0100	0.0120	0.0146
8%	– 0.1120	0.0036	0.0052	0.0069	0.0099	0.0119	0.0145
9%	– 0.0967	0.0035	0.0048	0.0069	0.0099	0.0119	0.0145
10%	– 0.0770	0.0034	0.0045	0.0069	0.0099	0.0118	0.0145
Left tail							
1%	0.3167	0.0023	0.0119	0.0091	0.0110	0.0119	0.0152
2%	0.1368	0.0026	0.0099	0.0076	0.0103	0.0118	0.0151
3%	0.0112	0.0033	0.0082	0.0065	0.0099	0.0119	0.0153
4%	0.0688	0.0029	0.0075	0.0069	0.0100	0.0118	0.0152
5%	0.0654	0.0029	0.0069	0.0069	0.0100	0.0118	0.0152
6%	0.0505	0.0029	0.0063	0.0068	0.0100	0.0118	0.0152
7%	0.0521	0.0029	0.0059	0.0068	0.0100	0.0118	0.0152
8%	0.0284	0.0030	0.0054	0.0068	0.0100	0.0119	0.0152
9%	0.0127	0.0031	0.0050	0.0068	0.0100	0.0120	0.0152
10%	– 0.0314	0.0034	0.0045	0.0068	0.0101	0.0121	0.0152

Asia

Hong Kong (dollar)

% of excess	ξ	σ	Threshold	VaR (95%)	Expected shortfall (95%)	VaR (99%)	Expected shortfall (99%)
Right tail							
1%	0.7191	0.0002	0.0006	0.0004	0.0006	0.0006	0.0012
2%	0.5968	0.0001	0.0005	0.0004	0.0006	0.0006	0.0011
3%	0.2707	0.0002	0.0003	0.0002	0.0005	0.0006	0.0010
4%	0.2644	0.0002	0.0003	0.0002	0.0005	0.0006	0.0010
5%	0.2691	0.0002	0.0002	0.0002	0.0005	0.0006	0.0010
6%	0.2847	0.0002	0.0002	0.0002	0.0005	0.0006	0.0010
7%	0.3002	0.0001	0.0002	0.0002	0.0005	0.0006	0.0009
8%	0.2942	0.0001	0.0002	0.0002	0.0005	0.0006	0.0010
9%	0.2776	0.0002	0.0001	0.0002	0.0005	0.0006	0.0010
10%	0.3097	0.0001	0.0001	0.0002	0.0005	0.0006	0.0009
Left tail							
1%	0.0653	0.0004	0.0006	0.0000	0.0004	0.0006	0.0011
2%	0.1254	0.0003	0.0004	0.0001	0.0004	0.0006	0.0011
3%	0.2045	0.0003	0.0003	0.0002	0.0005	0.0006	0.0010
4%	0.2474	0.0002	0.0003	0.0002	0.0005	0.0006	0.0010
5%	0.2558	0.0002	0.0002	0.0002	0.0005	0.0006	0.0010
6%	0.2757	0.0002	0.0002	0.0002	0.0005	0.0006	0.0009
7%	0.2966	0.0001	0.0002	0.0002	0.0004	0.0005	0.0009
8%	0.2875	0.0001	0.0002	0.0002	0.0005	0.0006	0.0009
9%	0.2767	0.0001	0.0001	0.0002	0.0005	0.0006	0.0009
10%	0.3071	0.0001	0.0001	0.0002	0.0004	0.0005	0.0009

Table 10 (cont)

Indonesia (rupiah)

% of excess	ξ	σ	Threshold	VaR (95%)	Expected shortfall (95%)	VaR (99%)	Expected shortfall (99%)
Right tail							
1%	– 0.2216	0.0419	0.0673	– 0.0142	0.0349	0.0670	0.1014
2%	– 0.1642	0.0436	0.0380	– 0.0055	0.0381	0.0663	0.0998
3%	0.1050	0.0291	0.0279	0.0135	0.0444	0.0620	0.0986
4%	0.3070	0.0208	0.0228	0.0183	0.0463	0.0587	0.1046
5%	0.3138	0.0193	0.0183	0.0183	0.0463	0.0586	0.1051
6%	0.3457	0.0173	0.0153	0.0186	0.0466	0.0581	0.1071
7%	0.4031	0.0149	0.0134	0.0188	0.0473	0.0574	0.1121
8%	0.4179	0.0138	0.0116	0.0187	0.0476	0.0573	0.1137
9%	0.3819	0.0139	0.0097	0.0188	0.0470	0.0576	0.1097
10%	0.4088	0.0128	0.0084	0.0187	0.0475	0.0574	0.1130
Left tail							
1%	0.4270	0.0187	0.0569	0.0350	0.0514	0.0567	0.0894
2%	0.1896	0.0250	0.0354	0.0142	0.0401	0.0537	0.0889
3%	0.2215	0.0213	0.0269	0.0167	0.0412	0.0536	0.0886
4%	0.2307	0.0195	0.0216	0.0174	0.0415	0.0535	0.0884
5%	0.2198	0.0189	0.0172	0.0172	0.0414	0.0537	0.0883
6%	0.2280	0.0178	0.0141	0.0174	0.0415	0.0536	0.0883
7%	0.2114	0.0178	0.0111	0.0173	0.0415	0.0539	0.0879
8%	0.2377	0.0164	0.0092	0.0174	0.0415	0.0534	0.0886
9%	0.2277	0.0163	0.0071	0.0174	0.0416	0.0537	0.0886
10%	0.2743	0.0146	0.0062	0.0173	0.0416	0.0529	0.0907

Malaysia (ringgit)

% of excess	ξ	σ	Threshold	VaR (95%)	Expected shortfall (95%)	VaR (99%)	Expected shortfall (99%)
Right tail							
1%	0.0026	0.0121	0.0205	0.0010	0.0130	0.0204	0.0325
2%	0.0473	0.0106	0.0139	0.0044	0.0150	0.0213	0.0328
3%	– 0.0210	0.0119	0.0089	0.0028	0.0146	0.0218	0.0332
4%	0.0130	0.0111	0.0060	0.0035	0.0147	0.0215	0.0330
5%	0.0581	0.0100	0.0041	0.0041	0.0148	0.0211	0.0328
6%	0.1713	0.0082	0.0031	0.0046	0.0148	0.0203	0.0337
7%	0.3473	0.0061	0.0025	0.0047	0.0152	0.0195	0.0378
8%	0.5202	0.0046	0.0022	0.0046	0.0167	0.0193	0.0473
9%	0.6630	0.0035	0.0019	0.0044	0.0199	0.0194	0.0644
10%	0.7371	0.0030	0.0016	0.0043	0.0232	0.0196	0.0815
Left tail							
1%	– 0.1915	0.0152	0.0200	– 0.0087	0.0086	0.0199	0.0326
2%	0.0294	0.0117	0.0113	0.0006	0.0124	0.0194	0.0318
3%	0.0059	0.0121	0.0063	0.0002	0.0123	0.0197	0.0320
4%	0.1299	0.0096	0.0043	0.0022	0.0129	0.0188	0.0320
5%	0.3236	0.0068	0.0033	0.0033	0.0133	0.0176	0.0344
6%	0.5093	0.0048	0.0027	0.0037	0.0144	0.0169	0.0414
7%	0.6266	0.0038	0.0023	0.0037	0.0161	0.0166	0.0507
8%	0.7174	0.0030	0.0019	0.0036	0.0187	0.0165	0.0643
9%	0.7713	0.0026	0.0017	0.0036	0.0213	0.0165	0.0780
10%	0.7626	0.0024	0.0014	0.0036	0.0208	0.0165	0.0754

Table 10 (cont)

Philippines (peso)

% of excess	ξ	σ	Threshold	VaR (95%)	Expected shortfall (95%)	VaR (99%)	Expected shortfall (99%)
Right tail							
1%	0.1785	0.0126	0.0224	0.0048	0.0163	0.0224	0.0376
2%	0.2242	0.0102	0.0146	0.0060	0.0168	0.0222	0.0376
3%	0.2923	0.0082	0.0112	0.0074	0.0173	0.0219	0.0378
4%	0.2938	0.0076	0.0088	0.0072	0.0173	0.0218	0.0379
5%	0.2805	0.0073	0.0071	0.0071	0.0172	0.0219	0.0378
6%	0.3479	0.0062	0.0062	0.0073	0.0174	0.0215	0.0391
7%	0.3059	0.0063	0.0050	0.0073	0.0173	0.0217	0.0381
8%	0.3465	0.0056	0.0044	0.0073	0.0174	0.0215	0.0391
9%	0.3839	0.0051	0.0039	0.0072	0.0175	0.0214	0.0404
10%	0.4156	0.0046	0.0035	0.0072	0.0177	0.0213	0.0418
Left tail							
1%	0.4173	0.0083	0.0185	0.0088	0.0160	0.0185	0.0326
2%	0.3344	0.0075	0.0125	0.0065	0.0148	0.0182	0.0323
3%	0.2872	0.0072	0.0091	0.0057	0.0144	0.0184	0.0322
4%	0.3775	0.0056	0.0077	0.0065	0.0148	0.0179	0.0331
5%	0.5049	0.0042	0.0069	0.0069	0.0154	0.0173	0.0363
6%	0.4317	0.0043	0.0060	0.0068	0.0150	0.0176	0.0340
7%	0.3333	0.0048	0.0050	0.0067	0.0147	0.0181	0.0319
8%	0.3170	0.0047	0.0043	0.0067	0.0147	0.0182	0.0316
9%	0.2980	0.0047	0.0036	0.0067	0.0147	0.0183	0.0313
10%	0.2915	0.0046	0.0031	0.0067	0.0147	0.0183	0.0311

Singapore (Singapore dollar)

% of excess	ξ	σ	Threshold	VaR (95%)	Expected shortfall (95%)	VaR (99%)	Expected shortfall (99%)
Right tail							
1%	– 0.1834	0.0059	0.0103	– 0.0008	0.0059	0.0103	0.0153
2%	– 0.0190	0.0048	0.0070	0.0025	0.0074	0.0104	0.0151
3%	0.1158	0.0038	0.0057	0.0038	0.0078	0.0101	0.0150
4%	0.2528	0.0029	0.0049	0.0043	0.0080	0.0098	0.0153
5%	0.2665	0.0027	0.0043	0.0043	0.0080	0.0098	0.0154
6%	0.2867	0.0025	0.0039	0.0044	0.0080	0.0097	0.0155
7%	0.2710	0.0024	0.0035	0.0044	0.0080	0.0098	0.0154
8%	0.3463	0.0021	0.0033	0.0044	0.0081	0.0096	0.0161
9%	0.3118	0.0021	0.0030	0.0044	0.0081	0.0097	0.0157
10%	0.3256	0.0020	0.0028	0.0044	0.0081	0.0096	0.0159
Left tail							
1%	0.1481	0.0057	0.0103	0.0021	0.0074	0.0103	0.0169
2%	0.2112	0.0045	0.0070	0.0033	0.0079	0.0103	0.0169
3%	0.2843	0.0036	0.0056	0.0039	0.0082	0.0102	0.0170
4%	0.3276	0.0030	0.0048	0.0041	0.0083	0.0101	0.0172
5%	0.3626	0.0026	0.0043	0.0043	0.0084	0.0100	0.0174
6%	0.2964	0.0028	0.0036	0.0042	0.0083	0.0102	0.0169
7%	0.3076	0.0026	0.0033	0.0042	0.0083	0.0102	0.0170
8%	0.3191	0.0024	0.0030	0.0042	0.0083	0.0101	0.0170
9%	0.3253	0.0023	0.0027	0.0042	0.0083	0.0101	0.0171
10%	0.3368	0.0022	0.0025	0.0042	0.0083	0.0101	0.0173

Table 10 (cont)

South Korea (won)

% of excess	ξ	σ	Threshold	VaR (95%)	Expected shortfall (95%)	VaR (99%)	Expected shortfall (99%)
Right tail							
1%	– 0.2322	0.0491	0.0222	– 0.0742	– 0.0162	0.0218	0.0618
2%	0.6963	0.0118	0.0136	0.0056	0.0260	0.0240	0.0867
3%	0.8231	0.0071	0.0104	0.0075	0.0341	0.0233	0.1235
4%	0.8915	0.0050	0.0090	0.0080	0.0461	0.0228	0.1830
5%	0.7355	0.0053	0.0074	0.0074	0.0276	0.0239	0.0899
6%	0.6621	0.0053	0.0063	0.0073	0.0249	0.0244	0.0756
7%	0.6627	0.0048	0.0055	0.0073	0.0249	0.0244	0.0757
8%	0.7024	0.0041	0.0050	0.0073	0.0264	0.0242	0.0833
9%	0.6871	0.0039	0.0045	0.0073	0.0257	0.0243	0.0800
10%	0.6852	0.0036	0.0041	0.0073	0.0257	0.0243	0.0797
Left tail							
1%	0.0755	0.0280	0.0224	– 0.0203	0.0066	0.0222	0.0525
2%	0.3220	0.0156	0.0122	– 0.0003	0.0168	0.0242	0.0529
3%	0.5929	0.0087	0.0089	0.0051	0.0208	0.0224	0.0633
4%	0.7563	0.0056	0.0075	0.0063	0.0258	0.0213	0.0871
5%	0.8596	0.0041	0.0066	0.0066	0.0355	0.0207	0.1361
6%	0.8818	0.0034	0.0060	0.0066	0.0399	0.0206	0.1583
7%	0.7022	0.0039	0.0051	0.0065	0.0232	0.0214	0.0730
8%	0.6696	0.0038	0.0045	0.0066	0.0222	0.0215	0.0674
9%	0.6921	0.0034	0.0041	0.0065	0.0229	0.0214	0.0712
10%	0.7417	0.0029	0.0039	0.0065	0.0251	0.0214	0.0828

Taiwan (New Taiwan dollar)

% of excess	ξ	σ	Threshold	VaR (95%)	Expected shortfall (95%)	VaR (99%)	Expected shortfall (99%)
Right tail							
1%	0.0930	0.0059	0.0086	– 0.0003	0.0053	0.0086	0.0151
2%	0.2709	0.0041	0.0053	0.0020	0.0063	0.0084	0.0152
3%	0.3819	0.0030	0.0042	0.0028	0.0068	0.0083	0.0155
4%	0.3914	0.0026	0.0034	0.0029	0.0068	0.0082	0.0156
5%	0.4001	0.0024	0.0029	0.0029	0.0068	0.0082	0.0157
6%	0.3876	0.0022	0.0025	0.0029	0.0068	0.0082	0.0155
7%	0.4118	0.0020	0.0022	0.0029	0.0068	0.0082	0.0158
8%	0.4509	0.0018	0.0020	0.0029	0.0069	0.0081	0.0164
9%	0.4265	0.0018	0.0017	0.0029	0.0069	0.0082	0.0160
10%	0.4155	0.0017	0.0015	0.0029	0.0068	0.0082	0.0158
Left tail							
1%	– 0.2632	0.0069	0.0071	– 0.0069	0.0015	0.0070	0.0125
2%	– 0.0737	0.0051	0.0044	– 0.0004	0.0047	0.0079	0.0124
3%	0.0507	0.0040	0.0032	0.0012	0.0053	0.0077	0.0122
4%	0.4018	0.0022	0.0028	0.0024	0.0058	0.0070	0.0135
5%	0.5538	0.0016	0.0025	0.0025	0.0062	0.0068	0.0157
6%	0.5043	0.0016	0.0022	0.0025	0.0060	0.0068	0.0148
7%	0.4385	0.0016	0.0019	0.0025	0.0059	0.0069	0.0137
8%	0.4565	0.0015	0.0017	0.0025	0.0059	0.0069	0.0140
9%	0.4426	0.0014	0.0015	0.0025	0.0059	0.0069	0.0138
10%	0.3959	0.0015	0.0013	0.0025	0.0058	0.0070	0.0131

Table 10 (cont)

Thailand (baht)

% of excess	ξ	σ	Threshold	VaR (95%)	Expected shortfall (95%)	VaR (99%)	Expected shortfall (99%)
Right tail							
1%	0.3821	0.0131	0.0236	0.0078	0.0192	0.0235	0.0447
2%	0.3663	0.0102	0.0158	0.0078	0.0193	0.0238	0.0446
3%	0.4287	0.0078	0.0124	0.0088	0.0199	0.0235	0.0455
4%	0.3511	0.0080	0.0097	0.0080	0.0194	0.0240	0.0441
5%	0.2481	0.0090	0.0071	0.0071	0.0192	0.0250	0.0430
6%	0.3139	0.0076	0.0062	0.0076	0.0193	0.0244	0.0437
7%	0.3705	0.0065	0.0053	0.0077	0.0194	0.0239	0.0453
8%	0.4283	0.0056	0.0047	0.0077	0.0197	0.0237	0.0477
9%	0.4112	0.0055	0.0040	0.0077	0.0196	0.0237	0.0468
10%	0.4298	0.0051	0.0035	0.0077	0.0198	0.0237	0.0478
Left tail							
1%	0.2904	0.0132	0.0229	0.0060	0.0176	0.0229	0.0414
2%	0.2742	0.0111	0.0143	0.0053	0.0172	0.0227	0.0413
3%	0.2957	0.0097	0.0099	0.0053	0.0172	0.0225	0.0415
4%	0.3043	0.0087	0.0074	0.0056	0.0173	0.0225	0.0416
5%	0.4270	0.0066	0.0063	0.0063	0.0178	0.0216	0.0445
6%	0.4958	0.0055	0.0054	0.0064	0.0183	0.0212	0.0477
7%	0.5661	0.0046	0.0048	0.0065	0.0192	0.0209	0.0526
8%	0.5245	0.0045	0.0041	0.0065	0.0186	0.0211	0.0493
9%	0.4422	0.0049	0.0033	0.0066	0.0179	0.0213	0.0444
10%	0.4430	0.0046	0.0028	0.0066	0.0179	0.0214	0.0444

Eastern Europe

Czech (Czech koruna)

% of excess	ξ	σ	Threshold	VaR (95%)	Expected shortfall (95%)	VaR (99%)	Expected shortfall (99%)
Right tail							
1%	0.3606	0.0045	0.0169	0.0114	0.0153	0.0169	0.0239
2%	0.2503	0.0045	0.0135	0.0098	0.0145	0.0168	0.0240
3%	0.2221	0.0043	0.0116	0.0095	0.0145	0.0169	0.0240
4%	0.1999	0.0042	0.0103	0.0094	0.0144	0.0171	0.0240
5%	0.1833	0.0042	0.0094	0.0094	0.0145	0.0171	0.0240
6%	0.1495	0.0044	0.0084	0.0092	0.0145	0.0174	0.0241
7%	0.1319	0.0045	0.0076	0.0091	0.0145	0.0175	0.0241
8%	0.1260	0.0044	0.0070	0.0091	0.0145	0.0175	0.0241
9%	0.1209	0.0044	0.0064	0.0091	0.0145	0.0176	0.0241
10%	0.2034	0.0040	0.0061	0.0090	0.0148	0.0177	0.0257
Left tail							
1%	0.0091	0.0053	0.0160	0.0075	0.0127	0.0159	0.0213
2%	0.0485	0.0049	0.0125	0.0081	0.0130	0.0159	0.0212
3%	0.0786	0.0044	0.0108	0.0086	0.0132	0.0159	0.0211
4%	0.0950	0.0041	0.0098	0.0089	0.0133	0.0158	0.0210
5%	0.1036	0.0039	0.0089	0.0089	0.0133	0.0158	0.0210
6%	0.1910	0.0033	0.0084	0.0091	0.0133	0.0156	0.0214
7%	0.0911	0.0039	0.0076	0.0089	0.0134	0.0159	0.0210
8%	0.0920	0.0038	0.0071	0.0089	0.0133	0.0159	0.0210
9%	0.0832	0.0039	0.0066	0.0089	0.0134	0.0159	0.0210
10%	0.0669	0.0040	0.0061	0.0089	0.0134	0.0160	0.0210

Table 10 (cont)

Hungary (forint)

% of excess	ξ	σ	Threshold	VaR (95%)	Expected shortfall (95%)	VaR (99%)	Expected shortfall (99%)
Right tail							
1%	0.6080	0.0038	0.0151	0.0112	0.0148	0.0151	0.0248
2%	0.4775	0.0035	0.0118	0.0091	0.0135	0.0146	0.0240
3%	0.3707	0.0037	0.0098	0.0081	0.0129	0.0148	0.0237
4%	0.3213	0.0037	0.0085	0.0077	0.0128	0.0150	0.0236
5%	0.3517	0.0035	0.0076	0.0076	0.0130	0.0151	0.0246
6%	0.3412	0.0035	0.0068	0.0074	0.0131	0.0155	0.0253
7%	0.3412	0.0033	0.0063	0.0074	0.0131	0.0154	0.0252
8%	0.3397	0.0032	0.0058	0.0074	0.0131	0.0155	0.0253
9%	0.3416	0.0030	0.0055	0.0074	0.0130	0.0153	0.0251
10%	0.3373	0.0030	0.0051	0.0074	0.0131	0.0155	0.0253
Left tail							
1%	0.0688	0.0056	0.0129	0.0044	0.0097	0.0129	0.0189
2%	0.3084	0.0032	0.0109	0.0083	0.0118	0.0133	0.0191
3%	0.1789	0.0037	0.0090	0.0072	0.0113	0.0135	0.0190
4%	0.1508	0.0036	0.0079	0.0071	0.0113	0.0135	0.0188
5%	0.1605	0.0034	0.0072	0.0072	0.0113	0.0135	0.0188
6%	0.1296	0.0036	0.0064	0.0070	0.0113	0.0137	0.0190
7%	0.1317	0.0035	0.0058	0.0070	0.0113	0.0137	0.0190
8%	0.0990	0.0038	0.0051	0.0070	0.0114	0.0139	0.0191
9%	0.0781	0.0039	0.0046	0.0070	0.0114	0.0140	0.0191
10%	0.0497	0.0042	0.0040	0.0070	0.0115	0.0142	0.0191

Poland (zloty)

% of excess	ξ	σ	Threshold	VaR (95%)	Expected shortfall (95%)	VaR (99%)	Expected shortfall (99%)
Right tail							
1%	– 0.2344	0.0107	0.0163	– 0.0047	0.0080	0.0163	0.0249
2%	– 0.0221	0.0078	0.0119	0.0046	0.0125	0.0173	0.0248
3%	0.0484	0.0067	0.0095	0.0061	0.0130	0.0171	0.0246
4%	0.2413	0.0048	0.0084	0.0073	0.0134	0.0163	0.0253
5%	0.2813	0.0043	0.0074	0.0074	0.0134	0.0162	0.0256
6%	0.1926	0.0047	0.0064	0.0073	0.0134	0.0165	0.0248
7%	0.2054	0.0045	0.0057	0.0073	0.0134	0.0165	0.0249
8%	0.2809	0.0038	0.0054	0.0073	0.0134	0.0162	0.0258
9%	0.2698	0.0038	0.0049	0.0073	0.0134	0.0163	0.0256
10%	0.2999	0.0035	0.0046	0.0073	0.0134	0.0162	0.0262
Left tail							
1%	– 0.0731	0.0088	0.0137	– 0.0014	0.0078	0.0136	0.0218
2%	0.2848	0.0046	0.0106	0.0068	0.0118	0.0141	0.0220
3%	0.4115	0.0033	0.0094	0.0079	0.0124	0.0139	0.0226
4%	0.3912	0.0030	0.0084	0.0077	0.0123	0.0139	0.0225
5%	0.2105	0.0039	0.0071	0.0071	0.0120	0.0146	0.0216
6%	0.1813	0.0039	0.0063	0.0071	0.0120	0.0146	0.0212
7%	0.1560	0.0041	0.0055	0.0069	0.0121	0.0149	0.0214
8%	0.1513	0.0041	0.0049	0.0069	0.0121	0.0149	0.0215
9%	0.1540	0.0040	0.0045	0.0069	0.0121	0.0149	0.0215
10%	0.2049	0.0036	0.0042	0.0069	0.0120	0.0147	0.0219

Table 10 (cont)

Slovakia (Slovakian koruna)

% of excess	ξ	σ	Threshold	VaR (95%)	Expected shortfall (95%)	VaR (99%)	Expected shortfall (99%)
Right tail							
1%	0.6071	0.0034	0.0150	0.0116	0.0148	0.0150	0.0235
2%	0.3394	0.0038	0.0121	0.0092	0.0133	0.0151	0.0222
3%	0.2923	0.0036	0.0104	0.0087	0.0131	0.0151	0.0222
4%	0.2068	0.0040	0.0090	0.0082	0.0130	0.0155	0.0222
5%	0.1710	0.0042	0.0079	0.0079	0.0130	0.0157	0.0223
6%	0.1572	0.0042	0.0071	0.0079	0.0130	0.0158	0.0224
7%	0.1134	0.0045	0.0062	0.0078	0.0131	0.0161	0.0225
8%	0.0972	0.0047	0.0055	0.0077	0.0131	0.0162	0.0226
9%	0.0885	0.0047	0.0049	0.0077	0.0132	0.0163	0.0226
10%	0.1554	0.0044	0.0044	0.0076	0.0135	0.0166	0.0241
Left tail							
1%	0.1208	0.0050	0.0154	0.0081	0.0128	0.0154	0.0211
2%	0.0945	0.0050	0.0118	0.0074	0.0125	0.0154	0.0213
3%	0.0954	0.0048	0.0098	0.0074	0.0124	0.0154	0.0213
4%	0.1204	0.0044	0.0086	0.0077	0.0125	0.0152	0.0211
5%	0.0840	0.0047	0.0074	0.0074	0.0125	0.0155	0.0213
6%	0.0456	0.0050	0.0063	0.0072	0.0125	0.0157	0.0214
7%	0.0605	0.0048	0.0056	0.0073	0.0125	0.0156	0.0214
8%	0.0457	0.0049	0.0049	0.0073	0.0125	0.0157	0.0214
9%	0.0604	0.0048	0.0044	0.0073	0.0125	0.0156	0.0214
10%	0.0568	0.0048	0.0039	0.0073	0.0125	0.0156	0.0214

Central and South America

Brazil (real)

% of excess	ξ	σ	Threshold	VaR (95%)	Expected shortfall (95%)	VaR (99%)	Expected shortfall (99%)
Right tail							
1%	0.3537	0.0122	0.0237	0.0087	0.0193	0.0236	0.0424
2%	1.3430	0.0025	0.0189	0.0176	0.0155	0.0217	0.0035
3%	0.9208	0.0026	0.0177	0.0166	0.0373	0.0227	0.1137
4%	0.7524	0.0027	0.0167	0.0162	0.0253	0.0232	0.0538
5%	0.6026	0.0030	0.0158	0.0158	0.0233	0.0239	0.0435
6%	0.5014	0.0032	0.0150	0.0156	0.0227	0.0244	0.0402
7%	0.3285	0.0044	0.0137	0.0152	0.0225	0.0256	0.0381
8%	0.1712	0.0065	0.0115	0.0147	0.0232	0.0277	0.0388
9%	0.0834	0.0084	0.0092	0.0143	0.0240	0.0296	0.0406
10%	0.0609	0.0090	0.0078	0.0142	0.0243	0.0302	0.0413
Left tail							
1%	0.4644	0.0078	0.0174	0.0085	0.0154	0.0173	0.0318
2%	0.4784	0.0058	0.0121	0.0078	0.0149	0.0168	0.0321
3%	0.4397	0.0050	0.0100	0.0077	0.0148	0.0170	0.0315
4%	0.3022	0.0058	0.0077	0.0065	0.0143	0.0178	0.0304
5%	0.3351	0.0051	0.0067	0.0067	0.0143	0.0176	0.0307
6%	0.2791	0.0054	0.0055	0.0065	0.0143	0.0179	0.0302
7%	0.2658	0.0053	0.0046	0.0064	0.0143	0.0180	0.0301
8%	0.2595	0.0054	0.0036	0.0064	0.0146	0.0185	0.0310
9%	0.2758	0.0048	0.0033	0.0064	0.0143	0.0179	0.0302
10%	0.2698	0.0049	0.0027	0.0064	0.0144	0.0182	0.0306

Table 10 (cont)

Chile (peso)

% of excess	ξ	σ	Threshold	VaR (95%)	Expected shortfall (95%)	VaR (99%)	Expected shortfall (99%)
Right tail							
1%	0.1755	0.0038	0.0101	0.0047	0.0082	0.0101	0.0147
2%	0.1296	0.0035	0.0078	0.0047	0.0083	0.0103	0.0147
3%	0.1988	0.0029	0.0066	0.0052	0.0085	0.0102	0.0148
4%	0.2410	0.0026	0.0059	0.0053	0.0085	0.0101	0.0148
5%	0.1555	0.0028	0.0051	0.0051	0.0085	0.0103	0.0146
6%	0.1719	0.0026	0.0047	0.0052	0.0084	0.0102	0.0145
7%	0.2308	0.0023	0.0044	0.0052	0.0085	0.0101	0.0149
8%	0.2172	0.0023	0.0040	0.0052	0.0085	0.0102	0.0148
9%	0.1676	0.0025	0.0037	0.0052	0.0085	0.0102	0.0145
10%	0.1768	0.0024	0.0034	0.0052	0.0084	0.0102	0.0145
Left tail							
1%	0.1465	0.0030	0.0090	0.0047	0.0075	0.0090	0.0125
2%	0.1423	0.0028	0.0068	0.0044	0.0073	0.0089	0.0124
3%	0.1466	0.0026	0.0058	0.0046	0.0074	0.0089	0.0124
4%	0.1605	0.0024	0.0051	0.0046	0.0074	0.0088	0.0124
5%	0.1223	0.0026	0.0044	0.0044	0.0073	0.0090	0.0125
6%	0.1050	0.0026	0.0039	0.0044	0.0074	0.0090	0.0126
7%	0.1225	0.0025	0.0036	0.0044	0.0073	0.0090	0.0125
8%	0.1314	0.0024	0.0033	0.0044	0.0073	0.0089	0.0125
9%	0.1109	0.0025	0.0029	0.0044	0.0074	0.0090	0.0126
10%	0.1078	0.0024	0.0027	0.0044	0.0074	0.0090	0.0125

Columbia (peso)

% of excess	ξ	σ	Threshold	VaR (95%)	Expected shortfall (95%)	VaR (99%)	Expected shortfall (99%)
Right tail							
1%	0.4404	0.0035	0.0136	0.0095	0.0126	0.0136	0.0199
2%	0.1930	0.0047	0.0097	0.0057	0.0106	0.0132	0.0198
3%	0.1832	0.0045	0.0078	0.0056	0.0106	0.0132	0.0199
4%	0.1950	0.0041	0.0067	0.0058	0.0106	0.0132	0.0198
5%	0.2351	0.0036	0.0060	0.0060	0.0107	0.0130	0.0199
6%	0.2376	0.0034	0.0054	0.0060	0.0107	0.0130	0.0199
7%	0.2091	0.0035	0.0048	0.0060	0.0107	0.0131	0.0196
8%	0.2156	0.0033	0.0044	0.0060	0.0106	0.0130	0.0195
9%	0.2120	0.0032	0.0040	0.0060	0.0107	0.0131	0.0196
10%	0.2197	0.0031	0.0037	0.0060	0.0106	0.0129	0.0195
Left tail							
1%	0.0402	0.0044	0.0102	0.0033	0.0076	0.0102	0.0147
2%	0.0379	0.0042	0.0075	0.0038	0.0079	0.0104	0.0149
3%	0.0812	0.0037	0.0061	0.0042	0.0081	0.0103	0.0147
4%	0.1062	0.0034	0.0051	0.0044	0.0081	0.0102	0.0146
5%	0.1764	0.0030	0.0046	0.0046	0.0082	0.0101	0.0149
6%	0.1322	0.0030	0.0040	0.0046	0.0081	0.0101	0.0145
7%	0.1290	0.0030	0.0036	0.0046	0.0081	0.0101	0.0145
8%	0.1815	0.0027	0.0032	0.0046	0.0082	0.0101	0.0149
9%	0.1482	0.0027	0.0029	0.0046	0.0081	0.0100	0.0144
10%	0.1513	0.0027	0.0027	0.0046	0.0081	0.0100	0.0144

Table 10 (cont)

Mexico (new peso)

% of excess	ξ	σ	Threshold	VaR (95%)	Expected shortfall (95%)	VaR (99%)	Expected shortfall (99%)
Right tail							
1%	0.3016	0.0230	0.0299	0.0004	0.0207	0.0297	0.0626
2%	0.3313	0.0181	0.0149	0.0005	0.0204	0.0288	0.0628
3%	0.6560	0.0091	0.0115	0.0076	0.0266	0.0263	0.0809
4%	0.6746	0.0073	0.0093	0.0078	0.0272	0.0261	0.0836
5%	0.7538	0.0055	0.0082	0.0082	0.0307	0.0256	0.1014
6%	0.8536	0.0041	0.0075	0.0083	0.0414	0.0250	0.1557
7%	0.8022	0.0039	0.0068	0.0083	0.0343	0.0252	0.1200
8%	0.6683	0.0044	0.0059	0.0083	0.0265	0.0258	0.0790
9%	0.6236	0.0044	0.0053	0.0084	0.0253	0.0260	0.0720
10%	0.6459	0.0040	0.0049	0.0084	0.0259	0.0259	0.0754
Left tail							
1%	0.2564	0.0168	0.0194	– 0.0029	0.0121	0.0193	0.0418
2%	0.4498	0.0094	0.0130	0.0059	0.0172	0.0206	0.0438
3%	0.5355	0.0066	0.0102	0.0073	0.0182	0.0202	0.0459
4%	0.4714	0.0064	0.0081	0.0068	0.0176	0.0206	0.0438
5%	0.5003	0.0055	0.0069	0.0069	0.0178	0.0204	0.0449
6%	0.5702	0.0044	0.0062	0.0071	0.0185	0.0200	0.0487
7%	0.6416	0.0036	0.0058	0.0071	0.0196	0.0198	0.0549
8%	0.6085	0.0035	0.0052	0.0071	0.0190	0.0199	0.0516
9%	0.5535	0.0036	0.0047	0.0072	0.0182	0.0200	0.0470
10%	0.5714	0.0033	0.0044	0.0071	0.0185	0.0200	0.0484

Peru (new sol)

% of excess	ξ	σ	Threshold	VaR (95%)	Expected shortfall (95%)	VaR (99%)	Expected shortfall (99%)
Right tail							
1%	0.6771	0.0031	0.0071	0.0041	0.0073	0.0071	0.0168
2%	0.8039	0.0015	0.0058	0.0048	0.0085	0.0072	0.0208
3%	0.5744	0.0017	0.0048	0.0040	0.0071	0.0075	0.0151
4%	0.6083	0.0014	0.0044	0.0041	0.0072	0.0074	0.0157
5%	0.4925	0.0015	0.0039	0.0039	0.0069	0.0076	0.0142
6%	0.3714	0.0017	0.0034	0.0038	0.0067	0.0079	0.0132
7%	0.3776	0.0016	0.0032	0.0038	0.0068	0.0079	0.0134
8%	0.3624	0.0017	0.0028	0.0037	0.0069	0.0082	0.0139
9%	0.3597	0.0017	0.0026	0.0037	0.0070	0.0083	0.0140
10%	0.2650	0.0019	0.0023	0.0037	0.0068	0.0083	0.0130
Left tail							
1%	0.4988	0.0036	0.0071	0.0031	0.0063	0.0070	0.0141
2%	0.4935	0.0026	0.0049	0.0030	0.0062	0.0070	0.0141
3%	0.4013	0.0024	0.0039	0.0027	0.0060	0.0072	0.0135
4%	0.4080	0.0021	0.0032	0.0028	0.0061	0.0072	0.0135
5%	0.4251	0.0019	0.0028	0.0028	0.0061	0.0072	0.0136
6%	0.4658	0.0016	0.0026	0.0029	0.0062	0.0071	0.0140
7%	0.4128	0.0017	0.0022	0.0028	0.0061	0.0072	0.0135
8%	0.4700	0.0014	0.0021	0.0029	0.0062	0.0071	0.0142
9%	0.4722	0.0013	0.0019	0.0029	0.0062	0.0071	0.0142
10%	0.4566	0.0013	0.0018	0.0029	0.0061	0.0071	0.0140

Table 10 (cont)

Venezuela (bolívar)

% of excess	ξ	σ	Threshold	VaR (95%)	Expected shortfall (95%)	VaR (99%)	Expected shortfall (99%)
Right tail							
1%	0.8298	0.0156	0.0139	0.0000	0.0240	0.0138	0.1049
2%	1.0274	0.0061	0.0080	0.0044	–	0.0140	–
3%	0.9737	0.0042	0.0060	0.0043	0.1022	0.0143	0.4837
4%	0.7617	0.0047	0.0042	0.0033	0.0198	0.0157	0.0719
5%	0.7413	0.0041	0.0032	0.0032	0.0190	0.0158	0.0678
6%	0.8267	0.0031	0.0027	0.0033	0.0240	0.0154	0.0933
7%	0.8490	0.0026	0.0023	0.0034	0.0263	0.0153	0.1055
8%	0.8781	0.0022	0.0020	0.0033	0.0310	0.0153	0.1288
9%	0.9145	0.0019	0.0018	0.0033	0.0414	0.0153	0.1811
10%	0.8621	0.0019	0.0016	0.0034	0.0280	0.0152	0.1139
Left tail							
1%	0.5490	0.0087	0.0119	0.0025	0.0105	0.0118	0.0311
2%	0.4895	0.0066	0.0066	0.0017	0.0099	0.0120	0.0301
3%	0.4854	0.0053	0.0044	0.0020	0.0101	0.0121	0.0297
4%	0.5257	0.0043	0.0031	0.0022	0.0103	0.0119	0.0307
5%	0.5957	0.0034	0.0024	0.0024	0.0108	0.0116	0.0335
6%	0.6275	0.0029	0.0019	0.0025	0.0112	0.0115	0.0354
7%	0.6329	0.0026	0.0015	0.0025	0.0113	0.0115	0.0359
8%	0.5854	0.0026	0.0011	0.0025	0.0107	0.0116	0.0327
9%	0.6640	0.0021	0.0009	0.0024	0.0118	0.0115	0.0388
10%	0.6883	0.0019	0.0007	0.0024	0.0123	0.0115	0.0414

Note: The exchange rate data are sourced from Bloomberg. The estimation period is 1 November 1993-29 October 2001.

The values of ξ and σ under the generalised Pareto distribution are estimated with the maximum likelihood estimation on the exceedances of daily logarithm changes in the exchange rates. VaR and expected shortfall are calculated using each of the estimated parameters.

Table 11

Estimation of the bivariate extreme value distribution of daily log changes of the Southeast Asian exchange rates

Currencies		α	ξ_1	σ_1	θ_1	ξ_2	σ_2	θ_2
Indonesia (rupiah)	Malaysia (ringgit)	1.2658	0.4088	0.0128	0.0084	0.7371	0.0030	0.0016
Indonesia (rupiah)	Philippines (peso)	1.3056	0.4088	0.0128	0.0084	0.4156	0.0046	0.0035
Indonesia (rupiah)	Singapore (dollar)	1.3316	0.4088	0.0128	0.0084	0.3256	0.0020	0.0028
Indonesia (rupiah)	Thailand (baht)	1.3855	0.4088	0.0128	0.0084	0.4298	0.0051	0.0035
Malaysia (ringgit)	Philippines (peso)	1.2578	0.7371	0.0030	0.0016	0.4156	0.0046	0.0035
Malaysia (ringgit)	Singapore (dollar)	1.5288	0.7371	0.0030	0.0016	0.3256	0.0020	0.0028
Malaysia (ringgit)	Thailand (baht)	1.3186	0.7371	0.0030	0.0016	0.4298	0.0051	0.0035
Philippines (peso)	Singapore (dollar)	1.3120	0.4156	0.0046	0.0035	0.3256	0.0020	0.0028
Philippines (peso)	Thailand (baht)	1.4267	0.4156	0.0046	0.0035	0.4298	0.0051	0.0035
Singapore (dollar)	Thailand (baht)	1.4364	0.3256	0.0020	0.0028	0.4298	0.0051	0.0035

Note: The exchange rate data are sourced from Bloomberg. The estimation period is 1 November 1993-29 October 2001. The estimation is for the right tails of the logarithm changes. The tail probabilities are set at $p_1 = p_2 = 0.1$.

Table 12

VaR and expected shortfall of the simulated sums of the exchange rates

Currencies: Indonesia (rupiah) and Malaysia (ringgit)

α=1.266 (Spearman's rho=0.340)

	VaR (95%)	VaR (99%)	ES (95%)	ES (99%)
Frank	0.02337	0.06852	0.06079	0.15357
Gaussian	0.02331	0.06958	0.06186	0.15783
Gumbel	0.02257	0.07041	0.06412	0.17071

Currencies: Indonesia (rupiah) and Philippines (peso)

α=1.306 (Spearman's rho=0.195)

	VaR (95%)	VaR (99%)	ES (95%)	ES (99%)
Frank	0.02464	0.06573	0.05481	0.12279
Gaussian	0.02464	0.06746	0.05611	0.12702
Gumbel	0.02408	0.07002	0.05855	0.13811

Currencies: Indonesia (rupiah) and Singapore (dollar)

α=1.332 (Spearman's rho=0.360)

	VaR (95%)	VaR (99%)	ES (95%)	ES (99%)
Frank	0.02118	0.05993	0.04980	0.11490
Gaussian	0.02133	0.06094	0.05061	0.11699
Gumbel	0.02132	0.06270	0.05203	0.12180

Currencies: Indonesia (rupiah) and Thailand (baht)

α=1.386 (Spearman's rho=0.203)

	VaR (95%)	VaR (99%)	ES (95%)	ES (99%)
Frank	0.02562	0.06788	0.05664	0.12629
Gaussian	0.02551	0.07015	0.05830	0.13219
Gumbel	0.02482	0.07298	0.06106	0.14513

Currencies: Malaysia (ringgit) and Philippines (peso)

α=1.258 (Spearman's rho=0.151)

	VaR (95%)	VaR (99%)	ES (95%)	ES (99%)
Frank	0.01161	0.03427	0.03382	0.09490
Gaussian	0.01154	0.03504	0.03550	0.10266
Gumbel	0.01111	0.03570	0.03648	0.10855

Currencies: Malaysia (ringgit) and Singapore (dollar)

α=1.529 (Spearman's rho=0.154)

	VaR (95%)	VaR (99%)	ES (95%)	ES (99%)
Frank	0.00844	0.02442	0.02660	0.08047
Gaussian	0.00834	0.02558	0.02844	0.08865
Gumbel	0.00811	0.02677	0.02919	0.09196

Currencies: Malaysia (ringgit) and Thailand (baht)

α=1.319 (Spearman's rho=0.448)

	VaR (95%)	VaR (99%)	ES (95%)	ES (99%)
Frank	0.01232	0.03692	0.03583	0.09971
Gaussian	0.01220	0.03778	0.03766	0.10826
Gumbel	0.01166	0.03850	0.03884	0.11547

Currencies: Philippines (peso) and Singapore (dollar)

α=1.312 (Spearman's rho=0.473)

	VaR (95%)	VaR (99%)	ES (95%)	ES (99%)
Frank	0.01043	0.02497	0.02116	0.04533
Gaussian	0.01047	0.02588	0.02179	0.04721
Gumbel	0.01035	0.02720	0.02288	0.05150

Currencies: Philippines (peso) and Thailand (baht)

α=1.427 (Spearman's rho=0.252)

	VaR (95%)	VaR (99%)	ES (95%)	ES (99%)
Frank	0.01455	0.03650	0.03066	0.06663
Gaussian	0.01440	0.03802	0.03185	0.07121
Gumbel	0.01395	0.03992	0.03366	0.07996

Currencies: Singapore (dollar) and Thailand (baht)

α=1.436 (Spearman's rho=0.411)

	VaR (95%)	VaR (99%)	ES (95%)	ES (99%)
Frank	0.01114	0.02754	0.02344	0.05152
Gaussian	0.01114	0.02882	0.02427	0.05418
Gumbel	0.01102	0.03037	0.02549	0.05885

Appendix A: Copula of multivariate exceedances

This appendix explains the finding of Ledford and Tawn (1996) that the copula of multivariate exceedances converges to the extreme value copula.

We consider the copula of multivariate maxima before considering the copula of multivariate exceedances. The following theorem gives the foundations for describing the asymptotic joint distribution of multivariate maxima (see Resnick (1987), Proposition 5.11 for the proof).

Theorem

Suppose that $\{(Z_{1j}, Z_{2j}); j = 1,\ldots,n\}$ are independent and identically distributed two-dimensional random vectors with the joint distribution function *F*. Also suppose that the marginal distribution of these two-dimensional random vectors is a Fréchet distribution. In other words, for each *i*, *j*, $\Pr[Z_{ij} \le z_{ij}] = \exp(-1/z_{ij})$. Define the vector of component-wise maxima as $M_{Z_i,n} = \max(Z_{i1}, Z_{i2},\ldots,Z_{in})$. Then, the following holds:

$$\Pr[\frac{M_{Z_{1,n}}}{n} \le z_1, \frac{M_{Z_{2,n}}}{n} \le z_2] = F^n(nz_1, nz_2) \to G(z_1, z_2)\text{, as } n \to \infty\text{,}$$

where $G(z_1, z_2) = \exp\{-V(z_1, z_2)\}$,

$$V(z_1, z_2) = \int_0^1 \max\{sz_1^{-1}, (1-s)z_2^{-1}\} dH(s).$$

H is a non-negative measure on [0,1] that satisfies the following condition:

$$\int_0^1 s dH(s) = \int_0^1 (1-s) dH(s) = 1.$$

Using this theorem, Ledford and Tawn (1996) show that the copula of multivariate exceedances converges to the bivariate extreme value copula as follows.

Suppose that $\{(Z_{1j}, Z_{2j}); j = 1,\ldots,n\}$ are independent and identically distributed two-dimensional random vectors with the joint distribution function F_*. Also assume that the marginal distribution of (Z_1, Z_2) is a Fréchet distribution. In other words, for each *i*, $\Pr[Z_i \le z_i] = \exp(-1/z_i)$. Based on Proposition 5.15 in Resnick (1987), F_* is within the domain of attraction of G_* if and only if the following holds:

$$\lim_{t\to\infty} \frac{-\log F_*(tz_1, tz_2)}{-\log F_*(t,t)} = \frac{-\log G_*(z_1, z_2)}{-\log G_*(1,1)}. \tag{A-1}$$

As this is an asymptotic result, Ledford and Tawn (1996) assume that this also holds with a sufficiently large value of $t = t_c$. That is, the following holds for a large value of $t = t_c$:

$$\frac{-\log F_*(t_c z_1, t_c z_2)}{-\log F_*(t_c, t_c)} = \frac{-\log G_*(z_1, z_2)}{-\log G_*(1,1)}. \tag{A-2}$$

Define z'_j as $z'_j = t_c z_j$. With (A-2), the following holds when z'_j is above some high threshold θ_j:

$$\log F_*(z'_1, z'_2) = \log F_*(t_c, t_c) \frac{\log G_*(z'_1/t_c, z'_2/t_c)}{\log G_*(1,1)}. \tag{A-3}$$

G_* satisfies the following condition since G_* is the extreme value distribution,

$$G(z'_1, z'_2) = \exp\{-V(z'_1, z'_2)\}\text{,} \tag{A-4}$$

where $V(z_1,z_2) = \int_0^1 \max\{sz_1^{-1},(1-s)z_2^{-1}\}dH(s)$.

Here, H is a non-negative measure on [0,1] that satisfies $\int_0^1 sdH(s) = \int_0^1 (1-s)dH(s) = 1$.

As V is a homogeneous function of order –1, this leads to the following relation (where z' is now expressed by z).

$$F_*(z_1,z_2) = \exp\left\{V(z_1,z_2)\frac{t_c \log F_*(t_c,t_c)}{V(1,1)}\right\} = \exp\{V(z_1,z_2)K\}, \tag{A-5}$$

where K is a constant.

To determine the value of K we consider the value of F_* at the threshold θ_j. If we suppose that this threshold value is the $1-\lambda_j$ quantile, θ_j is derived as $\theta_j = -1/\log(1-\lambda_j)$. Setting $z_1 = \theta_1 = -1/\log(1-\lambda_1)$ and $z_2 = \infty$ in (A-5), we obtain the following:

$$F_*(-1/\log(1-\lambda_1),\infty) = \exp\{V(-1/\log(1-\lambda_1),\infty)K\}. \tag{A-6}$$

The left-hand side of equation (A-6) is equal to $1-\lambda_1$ because it is the distribution function at the $1-\lambda_1$ quantile. On the other hand, the right-hand side of equation (A-6) is equal to $\exp\{-K\log(1-\lambda_1)\}$, as shown below:

$$\begin{aligned} V(-1/\log(1-\lambda_1),\infty) &= \int_0^1 \max\{s(-\log(1-\lambda_1)),(1-s)/\infty\}dH(s) \\ &= -\log(1-\lambda_1)\int_0^1 sdH(s) = -\log(1-\lambda_1) \end{aligned}. \tag{A-7}$$

As $1-\lambda_1 = \exp\{-K\log(1-\lambda_1)\}$, we find that K = –1. Setting this into (A-5), F_* is obtained as follows:

$$F_*(z_1,z_2) = \exp\{-V(z_1,z_2)\}, \tag{A-8}$$

where $V(z_1,z_2)$ is the same as in (A-4).

This shows that the asymptotic joint distribution of the multivariate exceedances whose marginal distribution is a Fréchet distribution is given by (A-8).

We use this result to obtain the copula of multivariate exceedances whose marginals are not Fréchet distributions. Define u_i as $u_i \equiv \Pr[Z_i \le z_i] = \exp(-1/z_i)$. Set $z_i = -1/\log u_i$ into $G(z_1,z_2) = \exp\{-V(z_1,z_2)\}$ to obtain the following copula:

$$C(u_1,u_2) = \exp\{-V(-\frac{1}{\log u_1},-\frac{1}{\log u_2})\}, \tag{A-9}$$

where $V(z_1,z_2)$ is the same as (A-4).

With copula invariance, this is the copula of exceedances for all marginals since the copula is invariant under increasing continuous transformations.[49]

[49] Proposition 5.10 in Resnick (1987) shows that this approach is appropriate.

Appendix B:
Tail risk of VaR under the generalised Pareto distribution

This appendix analyses the tail risk of VaR under the generalised Pareto distribution employing Feller's convolution theorem.[50] We assume that the marginal distributions of asset losses are the generalised Pareto and have the same tail index.

This assumption is different from the assumption in Sections 3 and 4 in two aspects. First, in Sections 3 and 4, we assume that only the exceedances follow the generalised Pareto distribution. In this appendix, we assume that the both exceedances and non-exceedances follow the same generalised Pareto distribution. Second, in Sections 3 and 4, we assume that the tail index is different among assets. In this appendix, we assume that the tail index is equal across assets. Thus, under the assumption in Sections 3 and 4, we are unable to employ the convolution theorem used in this appendix.

Feller (1971, p 278) and Embrechts et al (1997, Lemma 1.3.1) utilise the convolution theorem for regularly varying distribution functions to examine the properties of the sum of the independent random variables with the same tail index. We explain their conclusions, incorporating our concept of tail risk.

Suppose that two independent random variables Z_1 and Z_2 have the same distribution functions as follows:

$$G_{\xi,\sigma}(x) = 1-(1+\xi\cdot\frac{x}{\sigma})_+^{-1/\xi}. \tag{B-1}$$

The distribution function of the sum of the two random variables Z_1 and Z_2 is derived from the convolution of equation (B-1), as follows:

$$H(x) \equiv \Pr\{Z_1+Z_2 \le x\} \equiv \int_0^x G_{\xi,\sigma}(x-y)dG_{\xi,\sigma}(y). \tag{B-2}$$

The function $\overline{G}_{\xi,\sigma}(x) \equiv 1-G_{\xi,\sigma}(x)$ is transformed as follows:

$$\overline{G}_{\xi,\sigma}(x) = \left(1+\xi\cdot\frac{x}{\sigma}\right)_+^{-1/\xi} = x^{-1/\xi}\left(\frac{1}{x}+\frac{\xi}{\sigma}\right)_+^{-1/\xi}. \tag{B-3}$$

Since the term $(1/x+\xi(x-\theta)/\sigma)_+^{-1/\xi}$ on the right-hand side of equation (B-3) is *slowly varying*,[51] using Feller's convolution theorem (see Feller (1971, p 278), or Embrechts et al (1997, Lemma 1.3.1), the following relation holds when the value of x is sufficiently large:

$$\overline{H}(x) \approx x^{-1/\xi}\left\{\left(\frac{1}{x}+\frac{\xi}{\sigma}\right)_+^{-1/\xi}+\left(\frac{1}{x}+\frac{\xi}{\sigma}\right)_+^{-1/\xi}\right\} = 2\left(1+\xi\cdot\frac{x}{\sigma}\right)_+^{-1/\xi}, \tag{B-4}$$

where $\overline{H}(x) \equiv 1-H(x)$. Therefore, the distribution function of the sum of two independent random variables Z_1 and Z_2 is as follows, when the value of x is sufficiently large:

$$H(x) \approx 1-2\left(1+\xi\cdot\frac{x}{\sigma}\right)_+^{-1/\xi}. \tag{B-5}$$

---

50 Geluk et al (2000) adopt Feller's convolution theorem for analysing the portfolio diversification effect under fat-tailed distributions.

51 Slowly varying functions are those functions $L(x)$ that satisfy the following condition (see Feller (1971) for details):

$$\lim_{x\to\infty}\frac{L(tx)}{L(x)} = 1$$

Meanwhile, the distribution function of the sum of two fully dependent random variables whose distribution function is given by (B-1) follows the same distribution as $2Z_1$. Thus, the distribution function $I(x)$ of the sum of two fully dependent variables is given below:

$$I(x) \equiv \Pr\{2Z_1 \le x\} = \Pr\{Z_1 \le x/2\} = G_{\xi,\sigma}(x/2) = 1 - \left(1 + \xi \cdot \frac{x}{2\sigma}\right)_+^{-1/\xi}. \tag{B-6}$$

VaR has tail risk when the two distribution functions $H(x)$ and $I(x)$ intersect (that is, when there is a solution to $H(x) = I(x)$), and when the VaR confidence level is lower than the cumulative probability of this intersection. In the case of $\xi < 1$, there is a solution to $H(x) = I(x)$, and the cumulative probability $p(\xi)$ at the intersection is as follows:[52]

$$p(\xi) = 1 - 2\left(1 + \frac{2^{\xi} - 1}{1 - 2^{\xi-1}}\right)^{-1/\xi} \quad (\xi < 1). \tag{B-7}$$

With some calculations (B-7), we find that the tail index must be 0.9 or higher for VaR to have tail risk at the confidence levels of 95% and 99%.

The tail index is 0.9 or higher only when the distribution is so fat that the 1.2-th moment is infinite. Such a fat-tailed distribution is rarely found in financial data. Thus, under the assumptions of this appendix, we find that VaR does not have tail risk as long as the confidence level is sufficiently high.

References

Acerbi, C, C Nordio and C Sirtori (2001): "Expected shortfall as a tool for financial risk management", Working Paper, Italian Association for Financial Risk Management.

Acerbi, C and D Tasche (2001): "On the coherence of expected shortfall", Working Paper, Center for Mathematical Sciences, Munich University of Technology.

Artzner, P, F Delbaen, J M Eber and D Heath (1997): "Thinking coherently", *Risk*, Vol 10, No 11, pp 68-71.

——— (1999): "Coherent measures of risk", *Mathematical Finance*, Vol 9, No 3, pp 203-28.

Basak, S and A Shapiro (2001): "Value-at-risk based risk management: optimal policies and asset prices", *The Review of Financial Studies*, Vol 14, No 2, pp 371-405.

Bouyé, E (2001): "Multivariate extremes at work for portfolio risk management", Working Paper, Financial Econometrics Research Centre, CUBS, London.

Capéraà, P, A L Fougères and C Genest (1997): "A nonparametric estimation procedure for bivariate extreme value copulas", *Biometrika*, Vol 84, No 3, pp 567-77.

Coles, S (2001): *An introduction to statistical modeling of extreme values*, Springer-Verlag.

Coles, S, J E Hefferman and J A Tawn (1999): "Dependence measures for extreme value analysis", *Extremes*, Vol 2, No 4, pp 339-65.

Committee on the Global Financial System (1999): *A review of financial market events in Autumn 1998*, October.

——— (2000): *Stress testing by large financial institutions: current practice and aggregation issues*, April.

[52] When $\xi \ge 1$, VaR is not sub-additive and has no tail risk. As $H(x) < I(x)$ for all x, the VaR for independence is larger than the VaR for full dependence. This shows that VaR is not sub-additive. On the other hand, full dependence dominates independence in the first order stochastic dominance. Thus, the ordering by VaR is consistent with the ordering by the first order stochastic dominance.

Danielsson, J (2001): "The emperor has no clothes: limits to risk modelling", Working Paper, Financial Markets Group, London School of Economics.

Dunbar, N (2001): *Inventing money: the story of Long-Term Capital Management and the legends behind it*, John Wiley and Sons.

Embrechts, P, C Klüppelberg and T Mikosch (1997): *Modelling extremal events for insurance and finance*, Springer-Verlag.

Embrechts, P, L De Haan and X Huang (2000): "Modelling multivariate extremes", *Extremes and Integrated Risk Management*, P Embrechts, editor, Risk Waters Group, pp 59-67.

Embrechts, P, A McNeil, and D Straumann (2002): "Correlation and dependency in risk management: properties and pitfalls", *Risk Management: Value at Risk and Beyond*, M A H Dempster, editor, Cambridge University Press, pp 176-223.

Feller, W (1971): *An introduction to probability theory and its applications, Volume 2*, John Wiley and Sons.

Frees, E W and E A Valdez (1998): "Understanding relationships using copulas", *North American Actuarial Journal*, Vol 2, No 1, pp 1-25.

Geluk, J, L Peng and C de Vries (2000): "Convolutions of heavy tailed random variables and applications to portfolio diversification and MA(1) time series", *Advances in Applied Probability*, Vol 32, No 4, pp 1011-26.

Gumbel, E J (1960): "Bivariate exponential distributions", *Journal of the American Statistical Association*, Vol 55, No 292, pp 698-707.

Hartmann, E, S Straetmans and C G de Vries (2000): "Asset market linkages in crisis periods", Working Paper.

Hauksson, H A, M Dacorogna, T Domenig, U Müller and G Samorodnitsky (2000): "Multivariate extremes, aggregation and risk estimation", Working Paper, Olsen & Associates.

Hefferman, J E (2000): "A directory of coefficients of tail dependence", *Extremes*, Vol 3, No 3, pp 279-90.

Joe, H (1997): *Multivariate models and dependence concepts*, Chapman & Hall.

Jorion, P (2000): "Risk management lessons from long-term capital management", *European Financial Management* Vol 6, pp 277-300.

Kotz, S and S Nadarajah (2000): *Extreme value distributions theory and applications*, Imperial College Press.

Ledford, A W and J A Tawn (1996): "Statistics for near dependence in multivariate extreme values", *Biometrika*, Vol 83, pp 169-87.

——— (1997): "Modelling dependence within joint tail regions", *Journal of Royal Statistical Society Series B*, Vol 59, No 2, pp 475-99.

Longin, F and B Solnik (2001): "Extreme correlation of international equity markets", *Journal of Finance*, Vol 56, No 2, pp 649-76.

Lucas, A, P Klaassen, P Spreij and S Straetmans (2001): "An analytic approach to credit risk of large corporate bond and loan portfolios", *Journal of Banking and Finance*, Vol 25, No 9, pp 1635-64.

McNeil, A J (2000): "Extreme value theory for risk managers", *Extremes and Integrated Risk Management*, Risk Waters Group, pp 3-18.

Nelsen, R B (1999): *An introduction to Copulas*, Springer-Verlag New York.

Reiss, R D and M Thomas (2000): *Statistical analysis of extreme values: from insurance, finance, hydrology, and other fields*, Birkhäuser Verlag.

Resnick, R I (1987): *Extreme values, regular variations, and point process*, Springer-Verlag, New York, Berlin, Heidelberg.

Rockafeller, R T and S Uryasev (2000): "Optimization of conditional value-at-risk", *Journal of Risk*, Vol 2, No 3, pp 21-41.

Tawn, J A (1988): “Bivariate extreme value theory: models, and estimation”, *Biometrika*, Vol 75, No 3, pp 397-415.

Yamai, Y and T Yoshiba (2002a): “On the validity of value-at-risk: comparative analysis with expected shortfall”, *Monetary and Economic Studies*, Vol 20, No 1, Bank of Japan, pp 57-86.

——— (2002b): “Comparative analyses of expected shortfall and VaR: their estimation error, decomposition, and optimization”, *Monetary and Economic Studies*, Vol 20, No 1, Bank of Japan, pp 87-122.

——— (2002c) “Comparative analyses of expected shortfall and VaR (2): expected utility maximization and tail risk”, *Monetary and Economic Studies*, Vol 20, No 2, Bank of Japan, pp 95-115.

Extreme tails for linear portfolio credit risk models[1]

André Lucas,[2,3] Pieter Klaassen,[2,4]
Peter Spreij[5] and Stefan Straetmans[6]

Abstract

We consider the extreme tail behaviour of the CreditMetrics model for portfolio credit losses. We generalise the model to allow for alternative distributions of the risk factors. We consider two special cases and provide alternative tail approximations. The results reveal that one has to be careful in applying extreme value theory for computing extreme quantiles efficiently. The applicability of extreme value theory in characterising the tail shape very much depends on the exact distributional assumptions for the systematic and idiosyncratic credit risk factors.

1. Introduction

The management of *market* risks by banks and other lending institutions - especially investment banks - has gained in importance in recent years due to growing proprietary trading portfolios on the banks' balance sheets; see, for example, the popularity of the value-at-risk (VaR) concept. However, *credit* risk management is perhaps even more important within the financial sector because it directly relates to a bank's core function of financial intermediation.

Until recently, the bulk of the credit risk literature mainly concentrated on assessing the credit risk of individual exposures in isolation, ie without taking into account the potential for credit quality comovements and defaults; see, for example, Altman (1983), Caouette et al (1998) or the Journal of Banking and Finance (2001, vol 25 (1)) as starting references. More recently a portfolio view on credit losses has emerged by recognising that changes in credit quality tend to comove over the business cycle and that one can diversify part of the credit risk by a clever composition of the loan portfolio across regions, industries and countries. Thus in order to assess the credit risk of a loan portfolio, a bank must not only investigate the creditworthiness of its customers, but also identify the concentration risks and possible comovements of risk factors in the portfolio.

Several approaches have been developed in order to determine the credit loss distribution at the portfolio level; see, for example, *CreditMetrics* by Gupton et al (1997), *CreditRisk+* by Credit Suisse (1997), *PortfolioManager* by KMV (Kealhofer (1995)) or *CreditPortfolio View* by McKinsey (Wilson (1997a,b)). Despite the apparent differences between these approaches, they exhibit a common underlying framework; see Koyluoglu and Hickman (1998) and Gordy (2000). In a recent paper we extended the one-factor *CreditMetrics* approach to allow for general dependencies on and distributions of credit risk factors; see, for example, Lucas et al (2001a). We also introduced a limit law to efficiently approximate loss quantiles for portfolios with a finite number of exposures; see Lucas et

---

[1] We thank seminar participants at EURANDOM (Eindhoven) and risk conferences in Basel and Montreal for useful comments. Correspondence to: alucas@feweb.vu.nl, s.straetmans@berfin.unimaas.nl, pieter.klaassennl@abnamro.com or spreijwins@uva.nl.

[2] Department of Finance and Financial Sector Management, Vrije Universiteit, De Boelelaan 1105, NL-1081HV Amsterdam, Netherlands.

[3] Tinbergen Institute Amsterdam, Keizersgracht 482, NL-1017EG Amsterdam, Netherlands.

[4] ABN AMRO Bank NV, Financial Markets Risk Management (HQ 2035), PO Box 283, NL-1000EA Amsterdam, Netherlands.

[5] Korteweg-de Vries Institute, University of Amsterdam, Plantage Muidergracht 24, NL-1018TV Amsterdam, Netherlands.

[6] Limburg Institute of Financial Economics (LIFE), Maastricht University, Tongersestraat 53, NL-6200MD Maastricht, Netherlands.

al (2001b) and Finger (1999). This limit law can be used in order to perform analyses of the sensitivity of the credit loss quantiles to changes in the exposure characteristics, such as credit quality, the degree of systematic risk, and the maturity profile.

Suppose, however, that a credit risk manager is also interested in calculating credit loss quantiles for very high confidence levels or, stated differently, for very low tail probabilities q. Such tail probabilities may be much smaller than the usual 5% or 1%. These extreme credit loss quantiles may be of interest for the sake of testing certain stress scenarios. An initial way of calculating these quantiles consists in using the closed-form expression of the credit loss limit law from Lucas et al (2001b). However, in order to be able to derive this expression, one has to choose a probability distribution for the latent variable triggering credit migrations and defaults in the *CreditMetrics* setup. It can be shown that the quantile calculations may be quite sensitive to varying this distributional choice for the latent variable.[7] Moreover, in case more than one systematic risk factor is present, analytical techniques may be unavailable. In such cases, the manager has to resort to simulations. If the desired tail probability is extremely small, an unduly large number of simulations might be called for.

In order to circumvent this risk of misspecification, one can also estimate credit loss quantiles by directly focusing upon the distributional tail of portfolio credit losses. It is now generally accepted as a stylised fact that the tail of credit loss distributions behaves fairly different from the tail of a normal distribution. In particular, the portfolio credit losses exhibit more probability mass in the tails than a normal distribution with identical mean and variance. In fact, using the toolkit of *extreme value theory (EVT)*, we have shown in our previous paper that the tail probabilities of portfolio credit losses are polynomially declining to zero whereas a normal distribution has a tail that declines at an exponential rate. Stated differently, extreme portfolio credit losses happen relatively more frequently than one would expect on the basis of a normally distributed random variable. As a result, common rules of thumb for calculating loss quantiles based on the normal paradigm no longer apply. For example, the 99.9% quantile may lie much more than three standard deviations above the distributional mean, which is the number one would expect for the normal distribution.

Distributions with a polynomial tail decay are also called heavy-tailed or fat-tailed distributions. The statistical theory of EVT shows that a wide class of distributional models all display polynomially declining tails. Stated otherwise, if one is only interested in the tail behaviour of an empirical process, one does not need to know the whole distribution. For statistical inference on the extreme quantiles, it is sufficient to know that the stochastic process exhibits heavy tails. Apart from providing statistical derivations of limit laws for sample maxima, EVT also provides various estimators for the rate of tail decay in the case of fat tails, the so-called tail index. Quantile estimators that use these tail index estimates as an input can then easily be formulated.

EVT has become increasingly popular in financial research as a tool for modelling the tail of return distributions with an eye towards calculating risk measures such as value-at-risk (VaR). Exploiting the empirical stylised fact of heavy-tailed financial returns (Mandelbrot (1952)), EVT provides extreme quantile estimates for confidence levels q typically beyond the tail of the empirical distribution function ($q < n^{-1}$ with n the sample size). Good starting references on applications of EVT in market risk management include Daníelsson and de Vries (2000), Longin (2000), Longin and Solnik (2001), Embrechts et al (1997) and Embrechts (2000). Diebold et al (1998) provide a discussion of pitfalls and opportunities in the use of extreme value analysis in financial risk management.

EVT techniques have reportedly been employed in empirical work to limit the number of simulations needed to reliably estimate far-out quantiles. This is especially relevant if multiple risk factors are present. To our knowledge, however, there are no theoretical papers on the applicability of EVT to estimating credit loss quantiles far out in the distributional tail. In this paper we investigate the accuracy (estimation error) of EVT techniques for credit loss distributions. More specifically we investigate how far one should go into the distributional tail in order to obtain extreme value quantile estimates that are reasonably close to their exact underlying values. The latter quantile values are calculated for two different parametric distributions of the factor model components triggering default: the factors are assumed to be either normally distributed or Student-t distributed. For both cases, we

[7] Lucas et al (2001a) consider Gaussian (as in the *CreditMetrics* setup) and non-Gaussian parameterisations for the latent variable and find that minor changes in these distributional assumptions can have large effects on extreme credit loss quantiles.

find that the confidence levels q should be chosen extremely low in order to obtain an acceptable level of estimation risk. This evidence raises doubts over the practical use of extreme value analysis in the field of credit risk management.

The remainder of this article is structured as follows. In Section 2 we briefly review the *CreditMetrics* setup towards deriving the analytic distribution of portfolio credit losses. In Sections 3 and 4 we apply extreme value analysis to the tails of the portfolio credit loss distribution and compare the EVT quantile values with their true underlying counterparts in order to assess estimation risk. Concluding remarks are in Section 5.

2. Theoretical framework

Consider a credit portfolio consisting of n bonds. As we eventually want to focus upon the accuracy of extreme value analysis for estimating credit loss quantiles far into the tail, we keep the model setup relatively stylised to highlight the main issues. In particular, we consider bonds with identical characteristics (equal initial ratings, unit face values (1), equal default probabilities, etc). Moreover, we allow for only two end-of-period states for the bond: defaulted and not defaulted.[8]

In our benchmark setting, each bond j, where $j = 1, ..., n$, is characterised by a latent variable S_j triggering a bond's default. A logical, though not the only, candidate for S_j is the company's "surplus", ie the difference between the market value of assets and that of liabilities. Default occurs when the surplus falls below a threshold s^*. Given our assumption of a uniform default probability for the entire portfolio, s^* does not depend on j. The credit loss on individual exposures j is now given by the indicator variable

$1_{\{S_j < s^*\}}$.

We assume that the company surplus variable S_j obeys the linear factor model

$$S_j = \rho f + \sqrt{1-\rho^2}\varepsilon_j \tag{1}$$

with f and ε_j representing systematic influences (business cycle conditions, stock market fluctuations) and firm-specific shocks, respectively. Non-linear extensions of this model can be found in Lucas et al (2001a). The systematic and idiosyncratic shocks are assumed to follow stationary distributions $f \sim G(\cdot)$ and $\varepsilon_j \overset{iid}{\sim} F(\cdot)$. These distributional assumptions imply that our model encompasses the Gaussian *CreditMetrics* setup.

It can now easily be shown that

$$\begin{aligned} C &= \lim_{n\to\infty} n^{-1}\sum_{j=1}^{n} 1(S_j < s^*) \overset{a.s.}{=} P\left[S_j < s^* \mid f\right] \qquad (2) \\ &= P\left[\varepsilon_j < \frac{s^* - \rho f}{\sqrt{1-\rho^2}} \middle| f\right] \\ &= F\left(\frac{s^* - \rho f}{\sqrt{1-\rho^2}}\right). \end{aligned}$$

Notice that equation (2) is equivalent to Theorem 1 in Lucas et al (2001b). By conditioning on the common factor f in (2), one effectively averages out all idiosyncratic risk ε_j just as in the case of linear portfolio theory. Indeed, within the CAPM model only systematic risk persists when the number of assets increases. The limit law in (2) generalises this feature to the non-linear context of credit risk

8 The effects of portfolio heterogeneity on the credit loss distribution and its tail are discussed in Lucas et al (2001a,b).

management. Moreover, it is important to realise that the above limit law holds irrespective of the precise distributional assumptions on f and ε_j.

Knowledge of the limit law's analytic expression enables risk managers to calculate the loss distribution's quantiles for given confidence levels q without the need to resort to simulations. This follows from the following chain of equalities:[9]

$$\begin{aligned} P[C > c] &= P\left[F\left(\frac{s^* - \rho f}{\sqrt{1-\rho^2}}\right) > c\right] \\ &= P\left[f < \frac{s^* - \sqrt{1-\rho^2}\,F^{-1}(c)}{\rho}\right] \\ &= G\left(\frac{s^* - \sqrt{1-\rho^2}\,F^{-1}(c)}{\rho}\right) \\ &= q, \end{aligned} \tag{3}$$

which can be rewritten as

$$c = F\left[\frac{s^* - \rho G^{-1}(q)}{\sqrt{1-\rho^2}}\right]. \tag{4}$$

Clearly the use of the analytic quantile formula (4) requires knowledge of $G(\cdot)$ and $F(\cdot)$. However, a credit risk manager might only be interested in knowing the credit-at-risk for very low values of q for the sake of, for example, stress testing. In the next section we investigate to what extent extreme value analysis might be of use to the credit risk manager in order to calculate these extreme credit risk quantiles.

3. Analysing extreme tails

It has been established previously that portfolio credit losses in (2) exhibit a heavy tail; see Lucas et al (2001a,b) for a formal proof. This property can be expressed analytically as:

$$P(C > c) = (1-c)^{\alpha} L(1/(1-c)), \tag{5}$$

where α is the tail index of C governing the tail decay towards zero and $L(\cdot)$ stands for a slowly varying function, ie $\lim_{x\to\infty} L(tx)/L(t) = 1$, for $t > 0$. Examples are $L(x) = \ln(x)$ and $L(x) = K$ for some constant K. Clearly, the lower the tail index, the more likely extreme credit losses become.

It can be easily shown that there is a direct relation between the tail properties of the factors f and ε_j in (1) and the value of α. For example, if (f, ε_j) are standard normally distributed, then $\alpha = (1 - \rho^2) / \rho^2$. For Student-t distributed risk factors with corresponding degrees of freedom μ and ν for f and ε_j, respectively, we have $\alpha = \mu/\nu$; see Lucas et al (2001a). The tail result for the Gaussian case might appear somewhat counterintuitive at first sight, as normally distributed (thin-tailed) risk factors lead to a portfolio credit loss distribution with a polynomially (ie "fat") tail. However, the result simply reflects that a higher degree of systematic risk ρ implies a stronger domino effect of individual loans defaulting simultaneously in a credit portfolio. This effect makes the tail of the portfolio losses relatively fatter (lower α). The Student-t result leads to the observation that the tails of the credit loss distribution may be very fat if the idiosyncratic risk factor has thinner tails than the systematic risk factor ($\nu > \mu$). This makes economic sense. If f has fatter tails than ε_j, extreme realisations of S_j occur relatively more

---

[9] Analytic quantile calculations for linear *multifactor* models are more complicated, but there are still advantages over pure simulation in that the number of stochastic variables is reduced significantly by n.

often due to bad realisations of f than bad realisations of ε_j. Consequently, it is much more likely that large portions of the portfolio default simultaneously (due to systematic risk). Because of this clustering effect, extreme realisations of portfolio credit losses also become more likely, resulting in a lower rate of tail decay.

In order to evaluate the accuracy of extreme value analysis for the sake of extreme quantile estimation, we compare exact tail quantiles for specific choices of $F(\cdot)$ and $G(\cdot)$ with tail quantiles calculated by means of (4). The exact analytic quantiles are calculated by means of the quantile formula (4). We consider two specific choices of $F(\cdot)$ and $G(\cdot)$. Our first choice is the standard CreditMetrics model with $F(\cdot)$ and $G(\cdot)$ both standard normal. Second, we also consider a fat-tailed alternative where $F(\cdot)$ and $G(\cdot)$ are Student-t distributions with 3 and 5 degrees of freedom, respectively. The Student-t distributions are rescaled to have unit variance. These numbers for the degrees of freedom parameters are not unreasonable given empirical work on the tail behaviour of stock returns. Moreover, this choice of parameters ensures that the portfolio credit loss density does not diverge towards the edges of its support; see Lucas et al (2001a). We set the value of the asset correlation parameter ρ^2 to 20%, which is the value prescribed for corporate loans in the Basel proposals for the New Capital Accord; see Basel Committee on Banking Supervision (2001).

Table1

ML estimate of α for different tail probabilities *q*

q	**$\mu = \infty, \nu = \infty$**		**$\mu = 5, \nu = 3$**	
	ML	**EVT**	**ML**	**EVT**
10^{-1}	45.1	4	44.2	1.67
10^{-2}	29.8	4	9.3	1.67
10^{-3}	22.7	4	1.9	1.67
10^{-4}	18.7	4	0.9	1.67
10^{-5}	16.1	4	1.1	1.67
10^{-6}	14.2	4	1.3	1.67
10^{-7}	13.0	4	1.4	1.67
10^{-8}	12.0	4	1.5	1.67

Note: The table contains the ML estimate of the tail index α in the Weibull approximation of the tail obtained by minimising Kullback-Leibler distance in the tail, ie conditional on $c > c^*$ with c^* the $(1 - q)$-quantile of the exact credit loss distribution. The model is the CreditMetrics model with a 1% unconditional default probability, Student-t(5) distributed systematic risk factor f, and Student-t(3) distributed idiosyncratic risk factor ε_j. The correlation parameter is $\rho^2 = 20\%$. The EVT column contains the exact (limiting) EVT tail index.

Taking the tail expression in (5) as a point of departure, EVT analysis of the credit loss tail naturally starts by considering a linear (1st) Taylor approximation of the credit loss tail around the upper bound of the distributional support $c = 1$:

$$P(C > c) \approx K \cdot (1 - c)^{\alpha} \tag{6}$$

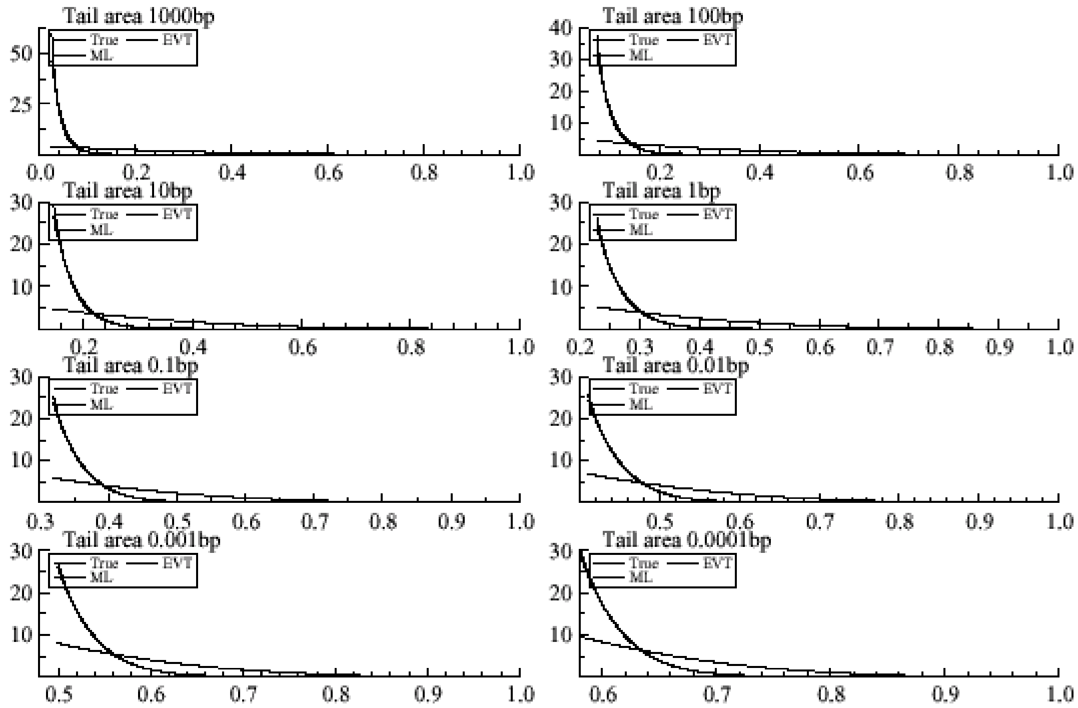
for some constants K and α, and with c close to 1. Thus we assume that the slowly varying function $L(1/(1 - c))$ is approximately constant for large c. First, we calculate extreme tail probabilities using (6) using the exact values of the tail index, ie $\alpha = (1 - \rho^2) / \rho^2 = 4$ for the Gaussian model, and $\alpha = \mu / \nu = 5/3$ for the Student-t model. Second, we estimate α by a Maximum Likelihood (ML) procedure which consists in minimising the Kullback-Leibler distance between (6) and (5) over the range $[c^*,1]$, where c^* is the $(1 - q)$-quantile of the credit loss tail for small values of q.[10]

[10] The Maximum Likelihood (ML) procedure based upon the Kullback-Leibler distance is asymptotically equivalent to applying the Hill (1975) estimator to a set of historical credit losses. This is because the Hill estimator is the ML estimator for a Pareto distribution. Note that $(1 - C)^{-1}$ has a regularly varying, ie a Pareto-type tail. Conditional upon knowledge of α (either the true value or an estimate), the scaling constant K in (6) can easily be calibrated from $\int_{c^*}^{1} K\alpha(1-c)^{\alpha-1}dc = q$ in order to ensure that the probability mass under the approximating pdf equals that under the exact pdf.

Let us now turn to the results for the linear tail approximation case. Table 1 gives the ML estimates of α for decreasing tail probabilities q, ie the lower q, the larger the corresponding tail quantile.

Figure 1

Tail approximations for the Gaussian model



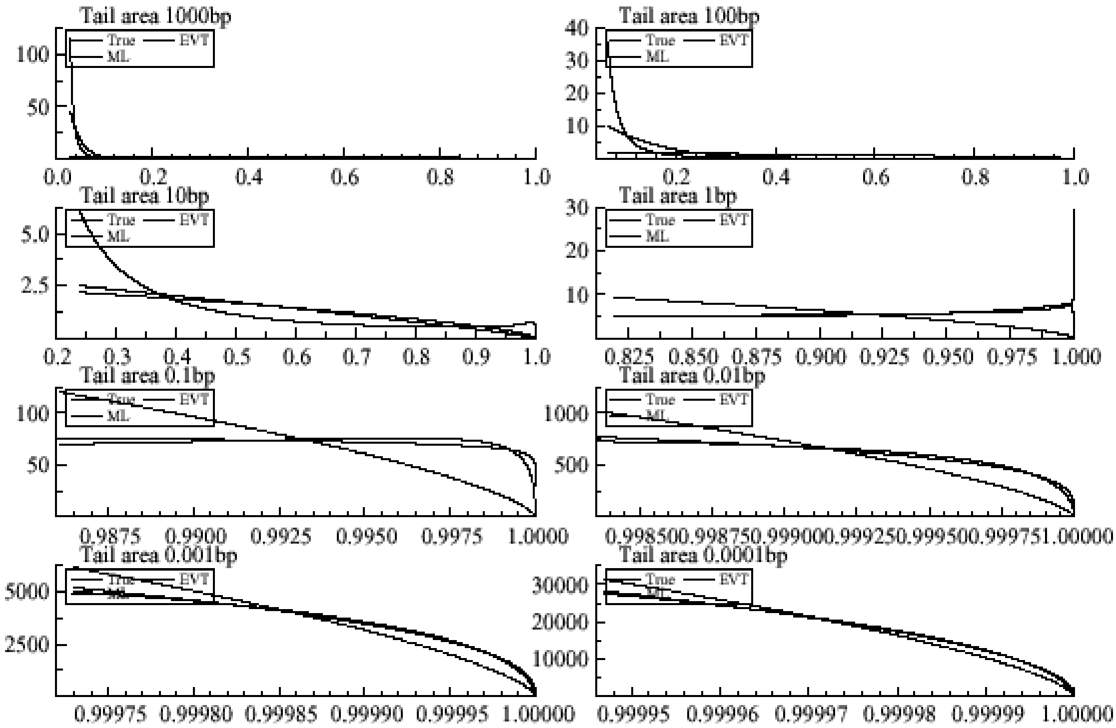
The model is the Gaussian CreditMetrics model with a 1% unconditional default probability. The title gives the tail area over which the credit loss distribution is plotted. EVT is the EVT tail approximation, while ML is the Weibull fit obtained after maximum likelihood or minimum Kullback-Leibler. The correlation parameter is $\rho^2 = 20\%$.

Clearly ML tail index estimates vary considerably with the chosen tail probability. For the Student-t model, the variation is even non-monotonic. It appears that the direction of convergence is ultimately towards its theoretical limit. But even for tail probabilities equal to a basis point of a basis point, the distance between the EVT and the ML α may be substantial, see the Gaussian model. For given values of α, K, and c (close to 1), the corresponding cumulative probabilities and densities can be derived. Conditional tail densities ($h(c|c > c^*)$) for different tail areas are shown in Figures 1 (Gaussian case) and 2 (Student-t case). Each figure contains three density curves: the exact density calculated by using the limit law for portfolio credit losses in (4), and two approximating densities $\hat{h}(c|c > c^*) = K\alpha(1-c)^{\alpha-1}$. The two approximations are the EVT, which uses the exact EVT value for α, and ML, which uses the α that minimises the Kullback-Leibler distance between the approximation $\hat{h}(\cdot)$ and the exact density $h(\cdot)$. For the ML density we impute the tail index estimates from Table 1.

At first sight, the ML fit does remarkably well in approximating the exact density. For the Gaussian models, the ML fit and exact density overlap for all practical purposes for given tail probability q. It may still be the case, however, that for varying q the approximation becomes worse. We investigate this issue in the next section in more detail. The EVT fit appears to approximate the true densities in the extreme tail, meaning that its shape resembles the extreme right-hand part of the exact density in each of the plots. For a given tail probability q, however, the EVT fit over the range $[c^*,1]$ is appallingly bad compared to the ML fit, unless one considers the Student-t case and $q = 10^{-8}$. This means that to recover the exact or limiting EVT tail shape from the exact credit loss density, one has to go *really extremely* far out into the tails. One may wonder whether credit risk managers want to know loss quantiles for $q \leq 10^{-4}$, which appears to be necessary for (exact) EVT to start to work.

Figure 2

Tail approximations for the Student-t model



The model is the CreditMetrics model with a 1% unconditional default probability, Student-t(5) distributed systematic risk factor f, and Student-t(3) distributed idiosyncratic risk factor ε_j. The title gives the tail area over which the credit loss distribution is plotted. EVT is the EVT tail approximation, while ML is the Weibull fit obtained after maximum likelihood or minimum Kullback-Leibler. The correlation parameter is ρ^2 = 20%.

Hitherto, we have compared the exact and approximate densities of credit losses on the basis of first-order tail approximation in (6). Note, however, that the linear approximation may be very imprecise because it assumes that the slowly varying function $L(\cdot)$ is approximately constant far out in the tails. As we have shown in previous work, this does not hold for the Gaussian model; see Lucas et al (2001a). This may partly explain the poor fit of the EVT approximation for moderately extreme quantiles. For the Student-t model, we can go even further. There, it can be shown analytically that the EVT fit is very poor for empirically relevant quantiles, but ultimately correct.

From Figure 2, the Student-t model produces a tail such that the conditional (tail) density starts up, goes down, then remains fairly constant over a certain range, and then slowly increases to sharply decline towards zero for c very close to the maximum loss 1. The exact EVT fit shows a conditional tail density that starts up and then decreases towards zero for $c \uparrow 1$. This is precisely the shape of the true density. To understand why the EVT tail approximation fits so badly, we consider the tail shape in more detail. In the case of a Student-t(3), the inverse cdf of $F(\cdot)$ can be approximated by

$$F^{-1}(c) \approx \sqrt[3]{\frac{1.1}{1-c}},$$

see Abramowitz and Stegun (1970), equation (26.7.7). As a result, we obtain

$$h(c) = H'(c) = \sqrt{\frac{1-\rho^2}{\rho^2}} \frac{g\left(\frac{s^* - \sqrt{1-\rho^2}F^{-1}(c)}{\rho}\right)}{f(F^{-1}(c))} \tag{7}$$

$$= \sqrt{\frac{1-\rho^2}{\rho^2}} \frac{\frac{\Gamma(3)}{\Gamma(2.5)\sqrt{3\pi}}\left[1+\frac{(s^*-\sqrt{1-\rho^2}F^{-1}(c))^2}{3\rho^2}\right]^{-3}}{\frac{\Gamma(2)}{\Gamma(1.5)\sqrt{\pi}}\left[1+(F^{-1}(c))^2\right]^{-2}}, \quad (8)$$

where the Student-t densities are parameterised to have zero mean and unit variance. Using further standard Taylor expansions, we obtain

$$h(c) = K_h \tilde{a}_0^{-3}\left(\frac{1-c}{1.1}\right)^{2/3} \sum_{k=0}^{\infty} c_k \left(\frac{1-c}{1.1}\right)^{k/3} \quad (9)$$

for c near 1, where

$$c_k = d_k + 2d_{k-2} + d_{k-4}, \quad (10)$$

$$d_k = \begin{cases} \sum_{j=\lceil k/2 \rceil}^{k} \binom{j+2}{j}\binom{j}{k-j}\left(-\tilde{a}_0\right)^{-j} \tilde{a}_1^{2j-k}\tilde{a}_2^{k-j} & \text{for } k \geq 0, \\ 0 & \text{for } k < 0, \end{cases}$$

$$\tilde{a}_0 = a_1^2, \quad (11)$$

$$\tilde{a}_1 = 2a_0 a_1, \quad (12)$$

$$\tilde{a}_2 = 1 + a_0^2, \quad (13)$$

$$a_0 = s^*/(\rho\sqrt{3}), \quad (14)$$

$$a_1 = -\sqrt{\frac{1-\rho^2}{3\rho^2}}, \quad (15)$$

$$K_h = \frac{4\sqrt{3(1-\rho^2)}}{9\rho}; \quad (16)$$

see the appendix. It is clear from (9) that near $c = 1$ the density of credit losses indeed has a Weibull expansion with $\alpha - 1 = 2/3$, or $\alpha = 5/3 = \mu/\nu$. The expression for $d_k (k \geq 0)$ is equivalent to

$$d_k = \left(-\tilde{a}_0\tilde{a}_2\right)^{k/2}\binom{2+\frac{1}{2}k}{\frac{1}{2}k} {}_2F_1\left(\frac{-k}{2}, 3+\frac{k}{2}; \frac{1}{2}; \frac{\tilde{a}_1^2}{4\tilde{a}_0\tilde{a}_2}\right),$$

where ${}_2F_1(a, b; c; z)$ is a hypergeometric function; see Abramowitz and Stegun (1970), Chapter 15. For $k \to \infty$, $|d_k|$ diverges. However,

$$\lim_{k\to\infty} \frac{\ln|d_k|}{k} = a_2$$

for some constant a_2 that does not depend on c, see equation (17) in Section 2.3.2 of Erdélyi (1953). Note, therefore, that $d_k(1-c)^k$ converges to zero for $(1-c)$ for sufficiently small values of $(1-c)$. A plot of $\ln|c_k|/k$ is given in Figure 3.

It is clear that higher-order terms in (9) than $(1-c)^{2/3}$ will be smaller in magnitude than $K > 0$ if

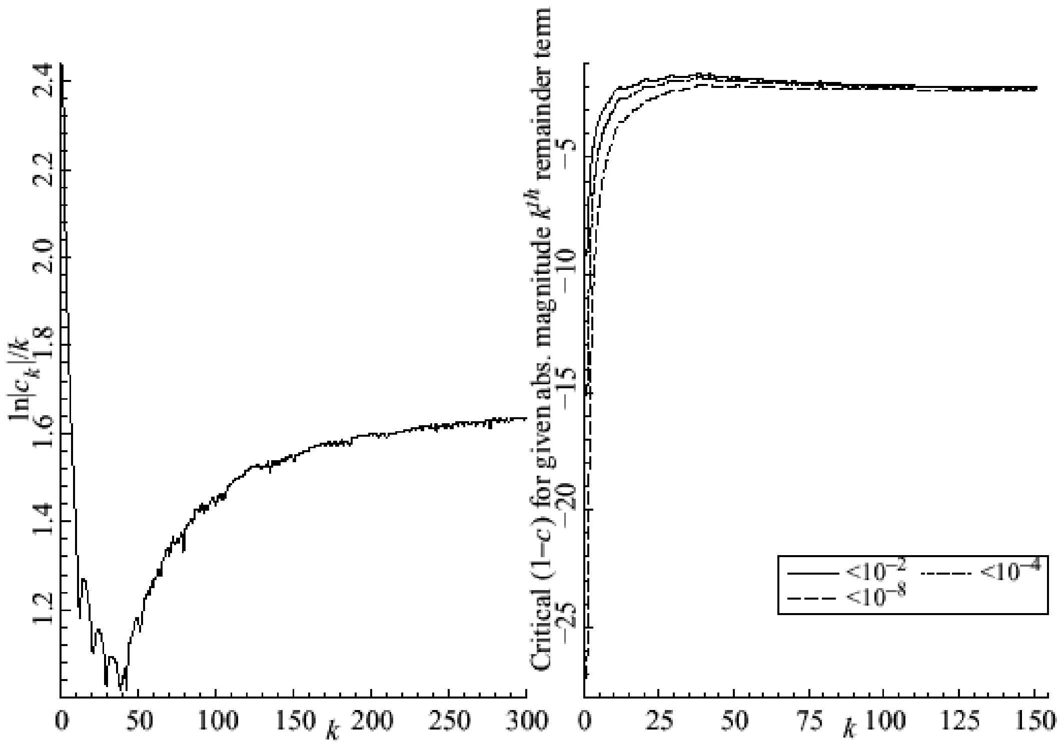
$$|c_k|\left(\frac{1-c}{1.1}\right)^{k/3} < K \Leftrightarrow 1-c < 1.1\left(\frac{K}{|c_k|}\right)^{3/k}.$$

A plot of the critical value of $1-c$ for different values of K is given in the right-hand panel of Figure 3. For example, if $K = 0.01$, we have for $k = 1$ that $(1-c)$ should be smaller than $7\cdot10^{-10}$, which is about *7% of a basis point of a basis point*. Clearly, the Weibull tail expansion only appears to set in in the *really extreme* tail, and not before. This explains the tail shapes in Figure 2.

The main conclusion we draw from our present computations is that one has to be very careful in applying tail expansions stemming from extreme value theory in the credit risk context. Higher-order terms may be important because they decline to zero very late, like the $(1-c)^{k/3}$ terms for k = 1,2 in the Student-t case. Moreover, the coefficients of the higher-order term may increase very steeply, also implying that one has to go further into the tails for the terms to become negligible. As a result, the extreme tail may start beyond quantiles of empirical interest. If this is the case, a different method of tail approximation might be called for altogether.

Figure 3

Coefficients ln|c_k|/k from (9) and following



The left-hand figure contains ln|c_k|/k, where c_k are the tail expansion coefficients from (9) and following. The right-hand plot gives the critical value of $(1-c)$ for which the kth-order term in the expansion is below K, where K is 10^{-2}, 10^{-4} or 10^{-8}. It is computed as $1.1(K/|c_k|)^{3/k}$.

4. Results

The ML fits in Figures 1 and 2 were reasonable for most tail areas. As this mimics the empirical application of EVT in practice to efficiently approximate a simulated version of $h(c)$, there is still some hope for the practical use of extreme value analysis in credit risk management. The applicability of EVT, however, hinges on the stability of the approximation over decreasing tail probabilities.

Of course, the estimate of α may differ for different tail areas (as shown in Table 1), but the real question is whether the fitted α produces estimates of credit loss quantiles or conditional expected credit losses that are adequate approximations to their true underlying values. To investigate this, we conducted the following experiment. Using the ML estimate of α for a specific tail probability q, we estimate the quantiles ($\hat{c}_1$ and $\hat{c}_2$) and conditional expected losses beyond those quantiles ($\hat{E}_1$ and $\hat{E}_2$) corresponding to tail probabilities of q/10 and q/100, respectively. We also calculated the percentage deviation (Δ) of the estimates from their true values. The results are in Table 2.

Table 2

Quantiles and expected losses beyond sample

q	c_0	c_1	$\hat{c}_1$	Δ	c_2	$\hat{c}_2$	Δ	E_1	$\hat{E}_1$	Δ	E_2	$\hat{E}_2$	Δ
	ML fit, μ = ∞, ν = ∞												
10^{-1}	0.03	0.08	0.07	–0.02	0.15	0.12	–0.18	0.11	0.09	–0.11	0.18	0.14	–0.24
10^{-2}	0.08	0.15	0.14	–0.01	0.23	0.21	–0.09	0.18	0.17	–0.05	0.27	0.23	–0.13
10^{-3}	0.15	0.23	0.23	–0.01	0.32	0.30	–0.05	0.27	0.26	–0.03	0.36	0.33	–0.08
10^{-4}	0.23	0.32	0.32	–0.00	0.41	0.40	–0.03	0.36	0.35	–0.02	0.45	0.43	–0.05
	EVT fit, μ = ∞, ν = ∞												
10^{-1}	0.03	0.08	0.45	5.00	0.15	0.69	3.75	0.11	0.56	4.34	0.18	0.75	3.15
10^{-2}	0.08	0.15	0.48	2.30	0.23	0.71	2.09	0.18	0.58	2.22	0.27	0.77	1.86
10^{-3}	0.15	0.23	0.52	1.27	0.32	0.73	1.28	0.27	0.62	1.29	0.36	0.78	1.18
10^{-4}	0.23	0.32	0.57	0.77	0.41	0.76	0.84	0.36	0.65	0.82	0.45	0.81	0.79
	ML fit, μ = 5, ν = 3												
10^{-1}	0.03	0.06	0.08	0.20	0.24	0.12	–0.48	0.14	0.10	–0.28	0.46	0.14	–0.69
10^{-2}	0.06	0.24	0.27	0.14	0.82	0.43	–0.48	0.46	0.34	–0.26	0.91	0.49	–0.47
10^{-3}	0.24	0.82	0.77	–0.06	0.99	0.93	–0.05	0.91	0.85	–0.07	0.99	0.96	–0.04
10^{-4}	0.82	0.99	0.99	–0.00	1.00	1.00	0.00	0.99	0.99	0.00	1.00	1.00	0.00
	EVT fit, μ = 5, ν = 3												
10^{-1}	0.03	0.06	0.76	10.70	0.24	0.94	2.95	0.14	0.85	5.26	0.46	0.96	1.09
10^{-2}	0.06	0.24	0.76	2.22	0.82	0.94	0.15	0.46	0.85	0.85	0.91	0.96	0.05
10^{-3}	0.24	0.82	0.81	–0.01	0.99	0.95	–0.03	0.91	0.88	–0.04	0.99	0.97	–0.02
10^{-4}	0.82	0.99	0.95	–0.03	1.00	0.99	–0.01	0.99	0.97	–0.02	1.00	0.99	–0.01

Note: Starting from the true quantile c_0 corresponding to a tail probability of q, we use the estimates of Table 1 to approximate the tail using a Weibull with the ML or EVT fit for α. Next, we compute the true $q/10$ quantile c_1 and the $q/100$ quantile c_2 and compare these with their Weibull approximations $\hat{c}_1$ and $\hat{c}_2$, respectively. We do the same for the conditional expected loss beyond c_1 and c_2 for the true distribution (E_1 and E_2, respectively), and beyond $\hat{c}_1$ and $\hat{c}_2$ for the Weibull approximation ($\hat{E}_1$ and $\hat{E}_2$, respectively). The fraction increase of the fitted/approximated value vis-à-vis the true one is given in the Δ columns.

Let us first consider the Gaussian model and the ML fit. If VaRs, or quantiles, slightly out of sample are estimated and the fit is very good (compare $\hat{c}_1$ with c_1), then the true VaR is underestimated by only 1% or 2%. Further out of sample, however, the approximation works less satisfactorily (compare $\hat{c}_2$ with c_2) and approximation errors increase within a range of 3% to as high as 18%. The approximation works better if the α parameter is estimated further out in the tails, ie for lower values of the tail probability q. A similar picture emerges if we consider expected losses rather than VaRs. The lower q, the better the out-of-sample approximation. Moreover, the approximation becomes worse the further we try to apply it out of sample. Also note that percentage mismatches of expected loss are already significant (11%) for $q = 10^{-1}$ and moderately out of sample ($q/10$). This is due to the fact that the expected loss also takes the goodness of fit of the tail approximation beyond the VaR quantile into account. From the quantiles we already noted that the $q/10$ quantile is approximately correct, but the tail approximation beyond that point becomes increasingly worse (see the $q/100$ quantile c_2 and $\hat{c}_2$). In any case, it is clear from all the Δ columns that the standard empirical application of EVT to the Gaussian model generally leads to an underestimation of the risk involved out of sample.

We now turn to the Student-t model and the ML fit. First, we note that the percentage and absolute approximation errors are much larger in general than for the Gaussian model. Moreover, the VaR moderately out of sample ($q/10$) may be under- or overestimated. The expected loss is underestimated. The same underestimation of risk is apparent if we look further out of sample ($q/100$) to either the VaR or the expected loss. Clearly, in the case of fat-tailed systematic and idiosyncratic risk factors, our results suggest that one should be more cautious in straightforwardly applying EVT approximations in the standard way to increase simulation efficiency and approximate risk measures out of (the simulated) sample.

Finally, we turn to the results for the Weibull approximation based on the EVT fit, ie on the exact rather than estimated tail index. The picture confirms the results from Figures 1 and 2, ie that the EVT fit works less well than the ML one. For the Gaussian model, percentage errors for the EVT fit are considerably higher than for the ML fit. Quantiles and expected losses are all more than 75% off mark. The use of the exact extreme value index α in the Weibull approximation leads to much too conservative (or prudent) estimates of risk. The picture is more subtle for fat-tailed risk factors. In particular, if one goes far out into the tails ($q = 10^{-3}$, 10^{-4}) to estimate the tail index α by ML, the EVT and ML fits produce very similar risk measures, which are both accurate to an error of about 5%. If one does not go far into the tails ($q = 0.1$), the ML fit is much better than the EVT fit for extrapolation purposes (at least up to $q/100$). For the intermediate case, $q = 0.01$, the EVT fit is much more useful if extrapolated far out into the tail ($q/100$; see $\hat{c}_2$ and $\hat{E}_2$). For nearer quantiles (see $\hat{c}_1$ and $\hat{E}_1$) the ML fit is considerably better. So the usefulness of Weibull approximations based on exact extreme value indices compared to ML estimates in the credit risk context very much depends on the tail area the ML estimate is based on and the extent of extrapolation beyond the sample envisaged for the EVT fit. If the tail area considered for ML estimation is large (high q) and one does not need to extrapolate further than $q/100$, then the exact EVT indices are of limited use. Note, however, that the approximations of quantiles and expected losses based on EVT fits improve broadly speaking when applied further out of sample ($q/100$ versus $q/10$). This holds for both the Gaussian and the Student-t models and corresponds to what one would expect. Though better, the approximation may, however, still be too prudent for empirical use.

5. Concluding remarks

The statistical theory of extreme values has been gaining in popularity within the financial research area for quite some time now. Researchers increasingly use tail index and quantile estimators (value-at-risk) in order to assess the tails of return distributions, both for single positions and for fully fledged portfolios. These statistical techniques can also be applied to calculate extreme credit loss quantiles. We investigated in this paper whether the application of extreme value theory (EVT) to the tails of portfolio credit losses is useful for the credit risk manager, ie are estimated EVT quantiles acceptably accurate or is the estimation error too large?

We started the analysis by calculating extreme quantile probabilities using the exact analytic expression of the portfolio credit loss distribution. We derived the loss distribution if the number of portfolio exposures grows large within the traditional *CreditMetrics* framework, ie portfolio exposures default either because of idiosyncratic shocks (ε_j) or because of systematic shocks (f). The analytic expression for the portfolio credit loss distribution for a large number of exposures exists upon knowledge of the distributional parameterisations for these factors. We therefore calculated the analytic credit loss quantiles conditional upon two different parametric choices for f and ε_j: Gaussian and Student-t distributed factors. The analytic portfolio credit loss distribution is heavy-tailed under either of the distributional choices for the underlying factors triggering defaults. As a consequence, we know from EVT that credit loss tail probabilities $P(C > c)$ can be factorised into a Pareto tail $(1 - c)^{\alpha}$ and a slowly varying function. We then considered a linear approximation for this factorisation and calculated extreme value probabilities, conditional upon both true values of the tail index and estimated values.

Upon comparing the analytic tail probabilities with their extreme value counterparts, we found that the extreme value probabilities come close to their true values provided one goes very far into the credit loss tail. Using higher-order expansions, we showed that very far out in the tail may mean, for empirical reasons, moving unrealistically far into the tails for higher-order terms to become negligible. It is doubtful whether credit risk managers would ever be interested in these remote tail areas. We conclude that standard use of EVT methods as applied in, for example, the market risk context is inappropriate in the credit risk context. More care should be taken when using EVT for credit risk management, and possibly a different method of tail approximation might be called for altogether.

Appendix
Proof of (9)

From (8), we obtain

$$h(c) = K_h \frac{\left[1 + \frac{(s^* - \sqrt{1-\rho^2}F^{-1}(c))}{3\rho^2}\right]^{-3}}{\left[1+(F^{-1}(c))^2\right]^{-2}} = \tag{A1}$$

$$K_h\left[1+t^2\right]^2\left[1+\frac{(s^*-\sqrt{1-\rho^2}t)^2}{3\rho^2}\right]^{-3} \tag{A2}$$

with $t = (1.1/(1-c))^{1/3}$. Define $y = 1/t$, and use the definitions in (10) to (16), then from (A2)

$$h(c) = K_h\left[1+t^2\right]^2\left[1+(a_0+a_1t)^2\right]^{-3} \tag{A3}$$

$$= K_h y^2\left[1+y^2\right]^2\left[\tilde{a}_2y^2+\tilde{a}_1y+\tilde{a}_0\right]^{-3}. \tag{A4}$$

Note that for $y \approx 0$ we have

$$(a+y)^{-3} = a^{-3}\sum_{k=0}^{\infty}(-a)^{-k}(k+2)_k\frac{y^k}{k!} = a^{-3}\sum_{k=0}^{\infty}\binom{k+2}{k}(-a)^{-k}y^k,$$

where $a_n = a\cdot(a-1)\cdots(a-n+1)$ is the Pochammer symbol. Using this result, rewrite

$$h(c) = K_h\frac{y^2}{\tilde{a}_0^3}(1+2y^2+y^4)\sum_{k=0}^{\infty}\binom{k+2}{k}(-\tilde{a}_0)^{-k}(\tilde{a}_2y^2+\tilde{a}_1y)^k \tag{A5}$$

$$= K_h\frac{y^2}{\tilde{a}_0^3}(1+2y^2+y^4)\sum_{k=0}^{\infty}\binom{k+2}{k}\left(\frac{\tilde{a}_1}{-\tilde{a}_0}\right)^k y^k\sum_{j=0}^{k}\binom{k}{j}\left(\frac{\tilde{a}_2}{\tilde{a}_1}\right)^j y^j,$$

or

$$h(c) = K_h\frac{y^2}{\tilde{a}_0^3}(1+2y^2+y^4)\sum_{k=0}^{\infty}d_ky^k, \tag{A6}$$

with

$$d_k = \left(\frac{\tilde{a}_2}{\tilde{a}_1}\right)^k\cdot\sum_{j=\lceil k/2\rceil}^{k}\binom{j+2}{j}\binom{j}{k-j}\left(\frac{\tilde{a}_1^2}{-\tilde{a}_0\tilde{a}_2}\right)^j.$$

Combining all this, we obtain

$$h(c) = K_h\frac{y^2}{\tilde{a}_0^3}\sum_{k=0}^{\infty}(d_k+2d_{k-2}+d_{k-4})y^k = K_h\frac{y^2}{\tilde{a}_0^3}\sum_{k=0}^{\infty}c_ky^k. \tag{A7}$$

References

Abramowitz, M and I Stegun (1970): *Handbook of Mathematical Functions*, New York: Dover.

Altman, E (1983): “Corporate financial distress. A complete guide to predicting, avoiding, and dealing with bankruptcy”, New York: Wiley.

Basel Committee on Banking Supervision (2001): *The New Basel Capital Accord*, technical report, Bank for International Settlements, January.

Caouette, J, E Altman and P Narayanan (1998): “Managing credit risk. The next great financial challenge”, New York: Wiley.

Credit Suisse (1997): *CreditRisk+*, downloadable: http://www.csfb.com/creditrisk.

Daníelsson, J and C de Vries (2000): “Value-at-Risk and extreme returns”, *Annale d’Economie et de Statistique* 60, 239.

Diebold, F, T Schuermann and J Stroughair (1998): “Pitfalls and opportunities in the use of extreme value theory in risk management”, in A-P N Refenes, A Burgess and J Moody (eds), *Decision technologies for computational finance*, 3-12, Deventer: Kluwer.

Embrechts, P (2000): “Extremes and integrated risk management”, London: Risk Publications.

Embrechts, P, C Klüppelberg and T Mikosch (1997): “Modelling extremal events”, vol 33 of *Applications of mathematics; stochastic modelling and applied probability*, Heidelberg: Springer Verlag.

Erdélyi, A (1953): “Higher transcendental functions”, vol 1-2, New York: McGraw-Hill.

Finger, C (1999): “Conditional approaches for CreditMetrics portfolio distributions”, *CreditMetrics Monitor* 1, 14-33, April.

Gordy, M (2000): “A comparative anatomy of credit risk models”, *Journal of Banking and Finance* 24 (1-2), 119-49.

Gupton, G, C Finger and M Bhatia (1997): *CreditMetrics - Technical Document* (1st edition), http://www.riskmetrics.com.

Hill, B (1975): “A simple general approach to inference about the tail of a distribution”, *Annals of Statistics* 3, 1163-73.

Journal of Banking and Finance (2001): vol 25 (1), pp 1-270, January.

Kealhofer, S (1995): “Managing default risk in derivative portfolios”, in *Derivative credit risk: advances in measurement and management*, London: Risk Publications.

Koyluoglu, H and A Hickman (1998): “Reconcilable differences”, *Risk*, 56-62, October.

Longin, F (2000): “From value at risk to stress testing: the extreme value approach”, *Journal of Banking and Finance* 24, 1097-130.

Longin, F and B Solnik (2001): “Extreme correlation of international equity markets”, *Journal of Finance* 56, 649-76.

Lucas, A, P Klaassen, P Spreij and S Straetmans (2001a): “Tail behaviour of credit loss distributions for general latent factor models”, TI Discussion Paper, 01-023/2, Tinbergen Institute.

——— (2001b): “An analytic approach to credit risk of large corporate bond and loan portfolios”, *Journal of Banking and Finance* 25, 1635-64.

Mandelbrot, B (1952): “The variation of certain speculative prices”, *Journal of Business* 36, 394-440.

Wilson, T (1997a): “Portfolio credit risk, part I”, *Risk*, 111-17, September.

——— (1997b): “Portfolio credit risk, part II”, *Risk*, 56-61, October.

Session 6

Market behaviour and monitoring

Interbank exposures and systemic risk[1]

Martin Blåvarg and Patrick Nimander

1. Background

Sweden underwent a severe banking crisis in the early 1990s. One of the lessons drawn was that the authorities were ill-prepared to deal with this type of situation, with regard to both crisis management and crisis prevention. After the crisis, in the mid-1990s, the Riksbank started to develop a new framework defining what its role as a non-supervisory central bank should be regarding financial stability.[2]

The starting point for this framework was that the central bank role, as well as that of other public interests in the financial sector, was built upon the existence of systemic risk. Without dwelling too much on the concept of systemic risk, it can be said that it exists because of the combination of two important factors. Firstly, the financial sector in general, and the payment system in particular, is very important for the functioning of the economy. A breakdown of the financial system will most likely carry substantial socio-economic costs. Secondly, the financial system, especially the banking system, is vulnerable to external shocks. Basically, depositors relate this to the fact that banks fund illiquid loans with liquid deposits, which makes them vulnerable to loss of depositor trust, which may lead to withdrawal of funds. Moreover, financial problems in one bank may spread to other banks and lead to losses and consequential failures of other banks (contagion). This combination of high probable social costs of failure and high fragility in the banking system is the main motive for regulating banks, according to the Banking Law Commission, which was set up with the purpose of reforming bank regulation in Sweden after the crisis.[3]

Risk of contagion between banks is thus an important element of systemic risk. Contagion in the banking system can typically be divided into direct and indirect contagion. Direct contagion arises because banks are financially exposed to one another, both through the payment system and through other types of positions such as outright loans, derivatives, repurchase agreements, etc. Indirect contagion can arise mainly through two channels. Firstly, markets may assume that direct contagion effects exist, even where this is not the case. Secondly, if one bank is struck by financial problems, markets may expect that other banks in the same system will be hit by the same problem, which in turn can lead to the other banks suffering a run by depositors.

Although risk of contagion is crucial as a motive for public interest in banking systems, it is striking how little this is reflected in regulatory systems. Regulation and supervision are to a very large extent directed at avoiding the failure of individual banks rather than the failure of the system as a whole.[4] Even if indirect contagion may be hard to influence by regulation or supervision, that should not be the case for direct contagion. In the area of payment systems, the main focus of the authorities is on the possible contagion effects that may arise due to the construction of the system. During the 1990s, a large majority of developed countries focused on using RTGS (real-time gross settlement) and DVP (delivery-versus-payment) mechanisms for making payment and settlement systems robust to individual bank failures and diminishing direct contagion effects through the system. However, little attention has been paid to the contagion effects arising outside the payment system. Many of the relevant interbank markets grew substantially during the 1990s. Global turnover on derivatives markets nearly doubled between 1995 and 2001, and turnover in foreign exchange markets more than

---

1 An earlier version of this article was published in Sveriges Riksbank's *Economic Review*, no 2, 2002, pp 19-45.

2 A description of the emergence of the Swedish banking crisis and how it has affected the authorities' monitoring and regulation of the banking system is given in Andersson & Viotti (1999).

3 The Commission's proposal is presently under consideration by the Government. For a brief description of the proposal, see Lind & Molin (1999).

4 See Acharya (2001) for a discussion on the scope for directing bank regulation to systemic risk rather than individual banks.

doubled between 1989 and 2001 (even though turnover in these markets has decreased over the last few years).[5] The higher turnover makes it probable that interbank exposures have grown as well.

The most obvious way for authorities to limit direct contagion effects would be to set regulatory limits for the size of the exposures banks were allowed to have towards one another. Most countries have rules regarding large exposures, but these are mainly set up in order to limit concentrations in banks' lending portfolios. In the EU regulatory framework, banks are not allowed to have individual counterparty exposures larger than 25% of their capital base. However, short-term exposures of less than one year between financial institutions are exempted from these rules.[6] It is common to regard the need for banks to take on large exposures to each other as an unavoidable part of their business, since they are intermediaries on interbank markets with very large flows, such as the foreign exchange and derivatives markets. The potential for direct contagion effects are thus often considered as natural.

In the field of research, the lack of data has been a general obstacle. Some work has been done on empirical measurement of contagion risks,[7] but to our knowledge there is nothing covering all interbank exposures, simply because data is not available. The lack of data is naturally connected to the low interest in this issue in the regulatory system. If supervisors do not demand the reporting of these exposures, no reporting data that can be used for research will be available. The banks' incentives to perform research themselves or provide data to outsiders are weak. Data on counterparties is normally not given freely, as this would disclose important information on the business of the bank. The incentives for banks to show their exposure to direct contagion effects may be weak, since this exposure may be one reason why the authorities may protect them in a crisis. Another reason for the lack of data in this area is simply that banks may not have felt any call to show this type of data, either from investors or supervisory authorities.

When developing the new financial stability framework at the Riksbank and trying to focus on systemic risk, the gap between the emphasis on contagion in theory on the one hand and the lack of regulatory initiatives or empirical research on the other hand was identified as a major area of concern. The Riksbank therefore wanted to develop an empirical base for estimating the effects of direct contagion. Even though the Riksbank is a non-supervisory central bank, it has a quite unique opportunity to collect information directly from financial institutions, since it has a legal right to demand any information from Swedish financial institutions. This article describes the kind of data that has been collected with the objective of analysing direct contagion effects, as well as presenting some quantitative results and drawing some conclusions as to how public authorities could deal with direct contagion.

2. Measurement of direct contagion

This section describes some of the issues that were important when the reporting of interbank exposures was developed at the Riksbank. In terms of procedure, the design of reporting was drawn up after a quite thorough investigation into the kinds of exposures Swedish banks had, what risks different types of exposures led to, how variable these exposures were over time, etc. This investigation was carried out in autumn 1998 and the reporting began in summer 1999.

The problem of direct contagion is normally seen as the risk that failure of one bank will lead to credit losses for other banks that are so great that their solvency is also threatened - if one bank falls, others will follow like a row of dominoes. To answer the question "How large could the losses be for other banks if one bank fails?" was the objective for the Riksbank when measuring direct contagion. There can be any number of reasons for one bank failing; it is just assumed that one bank fails for whatever reason. The approach targets the solvency effects of a bank failure on other banks. Failure of a bank

5 BIS (2002).

6 Individual countries may have stricter rules than this, but according to a brief survey of some EU countries made by the Swedish Financial Supervisory Authority, no country did. One country monitored interbank credit limits regularly.

7 See, for instance, Furfine (1999).

may also have liquidity impacts on other banks. The focus of the Riksbank's analysis and measurement of direct contagion has been on the solvency effect, which is reflected in the kinds of exposure that have been measured. However, the available data is also used for approximating effects on liquidity (see Section 3.6 Liquidity impact).

The willingness of banks to take on large exposures is quite dependent on the maturity. Banks may consider it fairly likely that they would receive at least some information in advance if an important counterparty were about to fail. If the time to maturity is only one day or a couple of days, it would be possible to withdraw credit exposures if a warning signal of potential failure were observed. An important issue here, therefore, is at what time horizon a bank is expected to fail, as an instantaneous failure would normally be expected to induce much greater losses than a gradual failure. In the payment system area, the focus is normally on the instantaneous failure of a bank. Interbank exposures are often of very short maturity. Interbank deposits, for instance, are predominantly overnight, at least in Sweden. As it may be difficult to measure intraday exposures globally[8] in large banks, the Riksbank chose to measure all overnight exposures, to investigate what would happen if one bank were to fail from one day to another. Although a failure of a large bank from one day to another is an unlikely event, it does happen, the failure of Barings probably being the most prominent example.

Sweden has a concentrated banking system - four large banks cover at least 80% of the system.[9] Because of its focus on systemic risk, the Riksbank concentrates its analysis on these four banks. Contagion could in general be expected to be a bigger problem in a concentrated system, since the large banks have fewer alternative counterparties in the interbank markets. As it is predominantly the failure of one of these four banks that could pose a systemic threat to the Swedish banking system, the measurement of direct contagion was conducted using the largest exposures of these four major banks. As reporting is costly for the banks, it was considered unnecessary to require all banks to do this special reporting. The difference in size between the fourth and fifth bank is so large that it is not possible that failure of one of the smaller banks could cause a loss big enough to become a threat for any of the larger banks. A failure of one of the larger banks could, on the other hand, be a threat for the smaller banks. The data collected cannot be used for analysing these latter effects.

The reporting requirements cover the 15 largest individual exposures. The reasoning behind this is that there should be few counterparties to whom banks are willing to take exposures large enough to threaten their solvency. This hypothesis has been confirmed by data (Figure 1). The size of exposures drops rapidly from the largest to the 15th largest counterparty. The 15th largest counterparty exposure is never of such a size that the failure of that counterparty would threaten the exposed bank.

One issue that was important when setting up the reporting requirements was what kind of exposures should be covered. As the purpose was to analyse what the effects on solvency would be if one of the largest counterparties failed from one day to another, it was decided to focus on exposures containing full principal credit risk. This means that the ranking was based upon uncollateralised exposures. To exclude collateralised exposures is reasonable since one of the most commonly used instruments on the Swedish interbank market is repurchase agreements with government bonds as the underlying assets. In most cases, there would be no losses on these repurchase agreements if a counterparty fails. If these exposures were not excluded, they would risk dominating the data. Collateralised exposures are reported as memo items for the 15 largest counterparties, but they do not comprise the basis for the ranking.[10]

---

8 "Globally" here refers to all business lines and all geographical locations in which a bank is active. Banks generally do not have information systems that record financial exposures on a real-time basis. The exposures are controlled by the setting of credit limits globally on particular counterparties, limits that then are distributed to different business units that may deal with that particular counterparty.

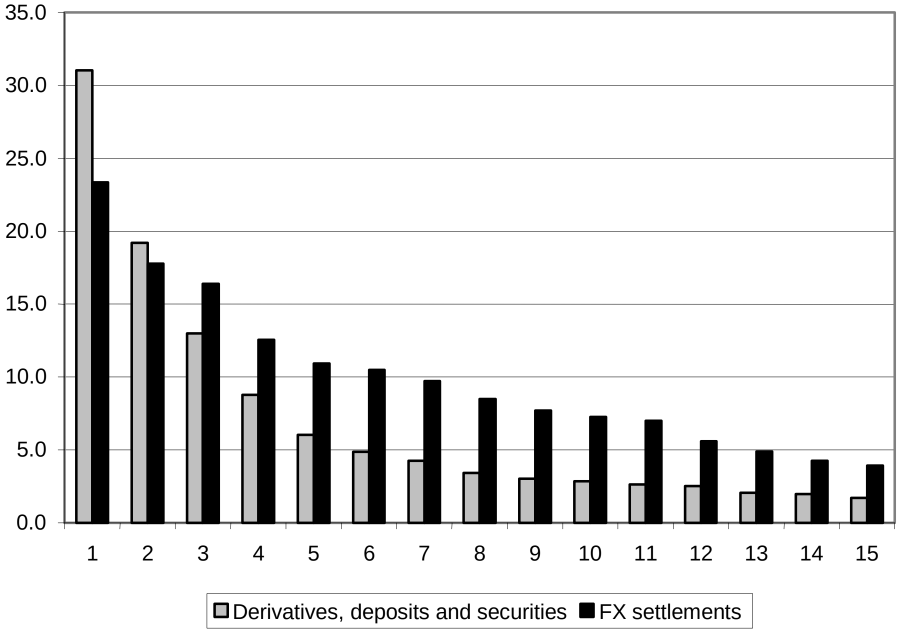
9 For a description of the structure of the Swedish banking market, see Sveriges Riksbank (2002).

10 See the reporting tables in Annex 1 for further information.

Figure 1

Swedish banks' exposure to the 15 largest counterparties. Average exposures in relation to total Tier 1 capital

In percentages



Source: Sveriges Riksbank.

The uncollateralised credit exposures that give rise to the size ranking are uncollateralised lending, holdings of securities issued by counterparties and the credit element of derivative exposures.[11] However, full principal credit risk can also arise because of settlement exposures, if payment and settlement systems are not constructed to incorporate PVP (payment versus payment) or DVP mechanisms. Swedish payment and settlement systems incorporate such mechanisms, except for foreign exchange settlement. FX settlement gives rise to a full principal credit exposure lasting on average two days. Outstanding FX settlement exposures are therefore included in the reporting. As these exposures are sometimes substantial compared to other exposures, they are not included in the size ranking of the counterparties, in order not to dominate the ranking. The 15 largest FX settlement exposures are instead ranked separately. By putting the two ranking lists together, the largest counterparties, both including and excluding FX settlement exposures, can then be established. In addition to the ranking of the largest individual exposures, the banks' total exposures within each area have been listed, in order to give a picture of the total size of interbank exposures and how concentrated these markets are.

The reporting also includes the names of each of the counterparties. This is useful for two reasons in particular. By including the names of the counterparties, the Riksbank can see if failure of one bank will affect several other Swedish banks. The names also make it possible to analyse second-round effects of contagion, that is, to construct scenarios with possible chain effects from defaults. The reporting also covers counterparties that are not financial institutions, even though it was expected that it would be mainly financial institutions to which the banks had very large exposures. This expectation has been confirmed; financial institutions dominate the ranking list, although from time to time non-financial companies are included on the lists, as well as financial companies.

[11] The positive market value of derivatives positions that a bank has against a particular counterparty. The relevant contracts are OTC derivatives rather than exchange-traded derivatives, as these exposures are normally secured. Banks often have contracts of both positive and negative value with a particular counterparty. These contracts can be netted against each other if the parties adopt netting agreements. Therefore, both gross and net exposures are reported.

The banks generally do not have information systems that collect financial exposures on a real-time or near real-time basis. Exposures are controlled by the setting of credit limits globally on particular counterparties, limits that are then distributed to different business units which may deal with that particular counterparty. To collect the actual exposures and rank them is quite burdensome and time-consuming for the banks.

As the kinds of exposures that are covered in this reporting are highly variable, it would in principle be interesting to have more frequent reporting. In order not to impose an undue burden on the banks, the Riksbank has limited the requirement to quarterly reporting. The reports are taken in for the end of the quarter, so that they coincide with the dates for financial statements, when actual exposures have to be collected globally within each institution anyway. The low frequency of reporting and the particular dates are of course a limitation for the analysis. Exposures can be expected to vary greatly from one day to another, and they are probably lower at the end of the quarter, since the banks in general do not like to show larger balance sheets than necessary. The Riksbank thus sees the reported exposures as indications of what size the exposures might be, rather than exact figures that are valid over time.

3. Reported counterparty and foreign exchange exposures

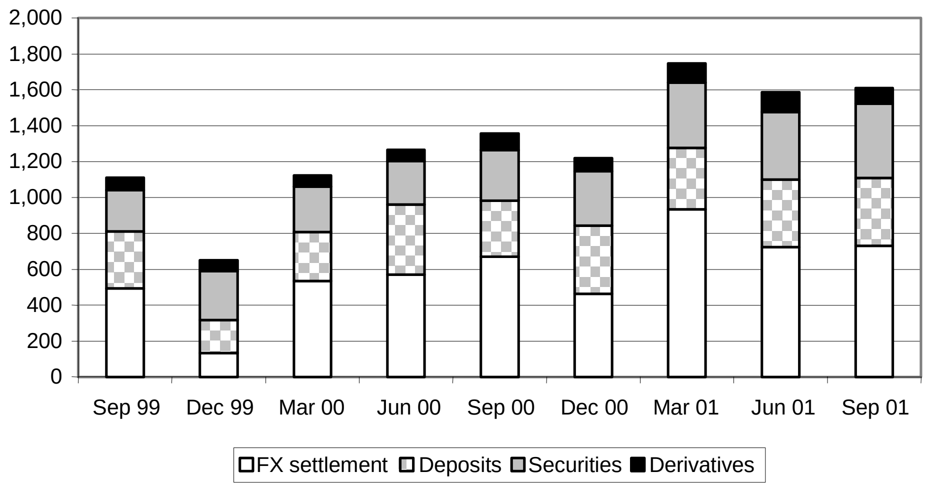
3.1 Overall size of exposures

The overall size of the reported exposures is approximately SEK 1,600 billion during 2001, for the four major Swedish banks.[12] This is a slight increase over the previous year.

Figure 2

Reported counterparty exposures by the four major Swedish banks

SEK billions



Source: Sveriges Riksbank.

The largest exposures are in the foreign exchange (FX) settlement segment, with these exposures normally making up between SEK 490 and SEK 730 billion of total exposures. Deposits have varied

[12] Reported exposures of SEK 1,600 billion can be compared to the Swedish GDP of approximately SEK 2,000 billion.

between SEK 273 and SEK 378 billion and securities between SEK 228 and SEK 414 billion. Derivatives exposure is the smallest class of exposures and has over the years increased from around SEK 60 billion to a high of SEK 110 billion and is now at SEK 87 billion. At the turn of the millennium, exposure levels were much lower, the result of very low levels of exposure to FX settlement and lower than normal exposure to deposits.

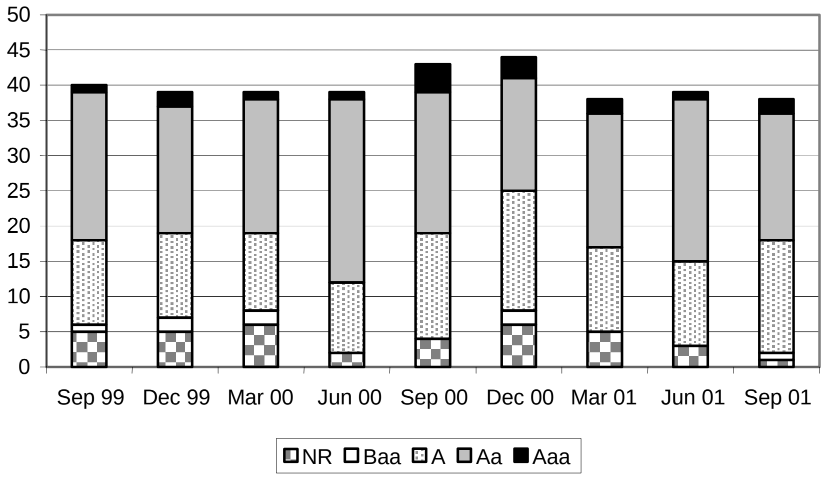
3.2 Counterparty rating

Possibly the banks' foremost means of controlling counterparty risks is to mainly expose themselves to counterparties with a high credit standing and to set limitations for exposures. One method of assessing credit standing is to study Standard & Poor's and Moody's credit ratings for the respective counterparties, as the Riksbank has no internal function for making credit assessments of banks.

The Swedish banks' counterparties have high credit ratings, according to the counterparty statistics. The average credit rating is A1/A+, which corresponds well to the ratings of the Swedish banks. The average credit rating has been at this level since the reports started in 1999.[13] The banks are largely exposed to counterparties with a credit rating of A or higher (Figure 3). There are counterparties with Baa ratings or with no rating from either S&P or Moody's. Counterparties lacking a public rating do not necessarily comprise greater credit risks than those with a rating, since the lack of credit rating could simply mean that they do not borrow directly in the market. Counterparties with no public rating from the rating agencies are normally well known by the banks that are exposed to them. The counterparties' relatively good credit standing indicates a low probability of sudden default among the counterparties.

Figure 3

Number of counterparties by rating category



Source: Sveriges Riksbank; Moody's; Standard & Poor's.

Generally, the counterparties used by the Swedish banks are internationally active foreign financial companies, Swedish and Nordic banking groups and some Swedish large and mid-sized non-financial companies.[14]

---

[13] Data was first reported for June 1999. In this article, data from September 1999 onwards is included, as the data from June does not fully correspond to the data reported later.

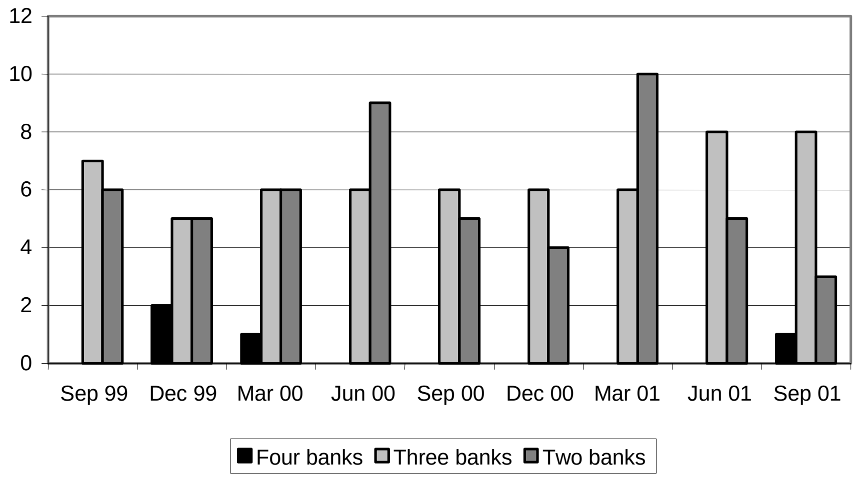
[14] Counterparties reported by a major Swedish bank can, of course, include one or more of the other major Swedish banks.

This confirms what we have seen in our work on credit risk management in the Swedish banks, ie that the Swedish banks actively manage which counterparties they do business with. Normally, limits on exposures are set through the use of ratings on the potential counterparties, either from rating agencies or internal ratings.

The four reporting banks rank their 15 largest exposures, in descending order of size. The maximum possible number of counterparties on each reporting occasion for the four major banks is thus 60. Since September 1999, the number of counterparties used by the banks has varied between 38 and 44 (Figure 3). The banks have little (or no) knowledge of which counterparties the other banks use regularly, and have no knowledge of which banks their competitors are exposed to at present. The number of counterparties reported by the banks indicates that the name concentration is not as big a problem as could have been assumed. The fact that the reported counterparties do not add up to 60 implies that there are counterparties to which more than one Swedish bank is exposed.

Figure 4

Number of Swedish banks exposed to the same counterparty



Source: Sveriges Riksbank.

The fact that more than one major Swedish bank might be exposed to the same counterparty is a possible source of risk concentration in the banking system. There are few counterparties to which all four banks are exposed at any time, but there are a number of counterparties to which two or three of the Swedish banks are exposed at any given time (Figure 4). The few counterparties shared by all four banks are not a major source of concern as they are normally highly rated counterparties to which the banks have lower levels of exposure. The counterparties shared by three of the banks deserve more attention, as this group normally includes several Swedish banks, and possibly could include some financial companies with lower credit ratings.

3.3 Direct contagion effects within the Swedish banking system

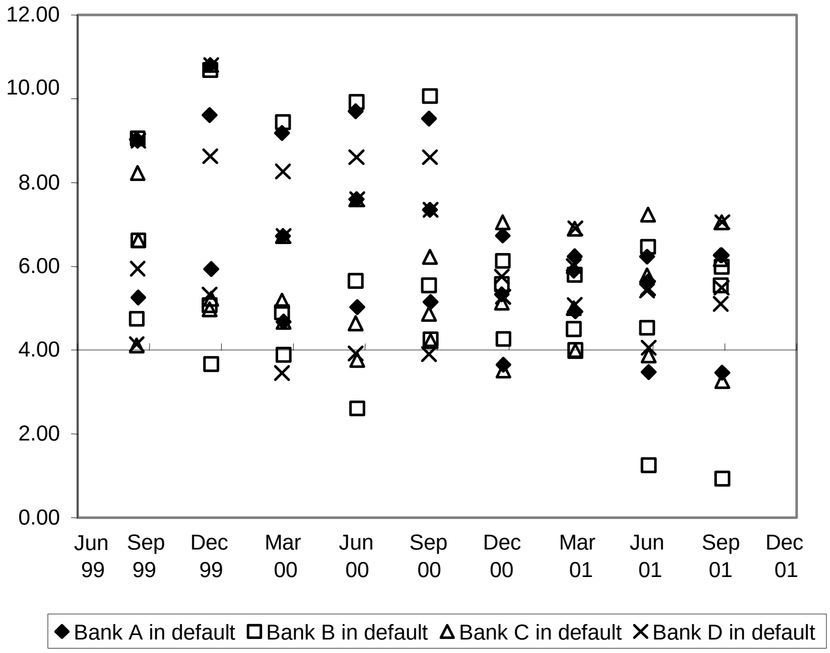
In the event of a default by one of the Swedish banks, there is a slight risk of a subsequent failure of another Swedish bank. A subsequent default could occur if one or several Swedish banks suffered such large losses that their capital was reduced below the statutory levels or to such an extent that the bank could not refinance itself in the market. In this paper, a loss big enough to lead to the Tier 1 capital of the bank falling below the required level of 4% is assumed to constitute a default. This is probably quite a conservative threshold.

Since September 1999, there have been a number of cases where a Swedish bank has had such substantial exposures towards another Swedish bank that there would have been direct risk of contagion if one of these counterparties had defaulted. In such cases, only if almost the whole of the exposed amount were lost would the exposed banks' capital actually decline sufficiently for direct contagion to occur. The Tier 1 capital ratios of the Swedish banks declined over the studied period. They were high during the first half of the period as some Swedish banks were in the process of merging or taking over other banks. Higher initial capital ratios give the banks stronger resilience to losses from counterparty exposures. The shift in Tier 1 capital ratios can clearly be seen in Figure 5. The shift occurs between September and December 2000.

Figure 5

Tier 1 capital in Swedish banks after a major Swedish bank default, assuming no recoveries

In percentages



Source: Sveriges Riksbank.

Note: Figures 5 and 6 illustrate Tier 1 capital ratios in the three surviving Swedish banks after one of the other Swedish banks has defaulted; the capital ratio is lowest after a default by Bank C.

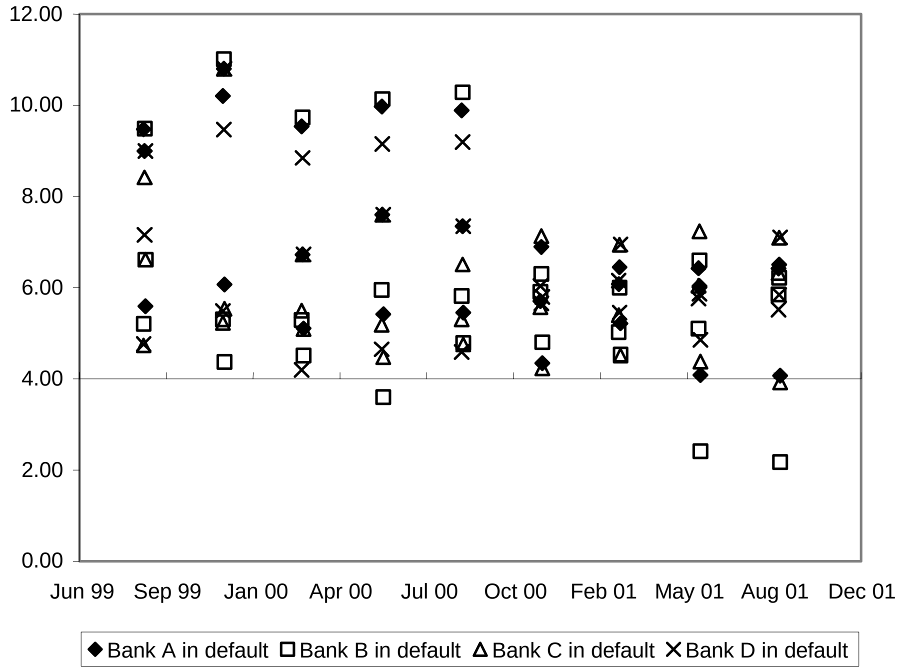
On the basis of the reported counterparty exposures and the Tier 1 capital ratios of the Swedish banks, there are 16 cases where the exposed bank's Tier 1 capital ratio would have fallen below the statutory 4% level if one of the other Swedish banks had defaulted (Figure 5). The total number of reported counterparty exposures to date is 108. These 16 cases occur assuming no recovery at all, or full loss of the total exposed amount. Assuming no recovery at all is, of course, very conservative by all standards. If we assume that the losses at default are only 75% of the exposed amounts, or 25% recovery, the number of cases where the Tier 1 capital ratio falls below 4% would be only four (Figure 6).

The severity of losses also seems to increase during the latter part of the period for which data is available. This is the effect of decreases in the Tier 1 capital ratios of all the Swedish banks, but also of higher levels of exposure between some of them. The main observation as regards direct contagion in the Swedish interbank markets is that there is a potential for large losses by some Swedish banks if other Swedish banks default. The likelihood of direct contagion in the Swedish banking system is dependent on which of the banks defaults, as there are links between the banks. Depending on which of them defaults the risk of direct contagion varies, as the exposures major banks allow themselves to other banks differ quite substantially. In the event of a counterparty default, it is only major losses with low degrees of recovery that would lead to contagion from one Swedish bank to another, almost regardless of which bank defaults. The risk of contagion effects between the banks is thus relatively slight, even though a few instances would definitely constitute very severe losses to some of the banks, even forcing the exposed bank into default.

Figure 6

Tier 1 capital in Swedish banks after a major Swedish bank default, assuming 25% recoveries

In percentages



Source: Sveriges Riksbank.

3.4 Direct contagion from abroad

We conclude that the risk of contagion within the Swedish banking system is relatively slight. There could of course be other channels through which direct contagion effects might hit the Swedish banking system. One such channel is the foreign counterparties to which the major Swedish banks are exposed.

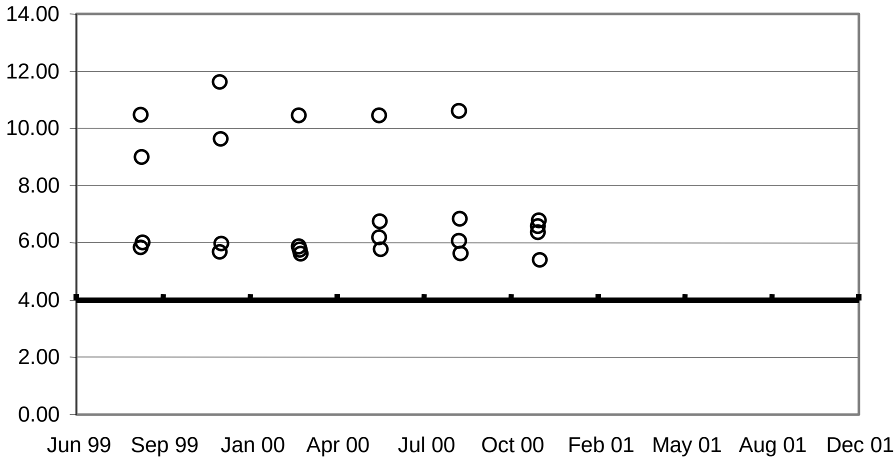
The effects on Swedish banks of a default by their largest foreign counterparty could possibly become a threat to financial stability. We have observed Tier 1 capital ratios for Swedish banks after their largest foreign counterparty has defaulted. In Figure 7, capital ratios are calculated for Swedish banks assuming full loss of the exposed amounts, and in Figure 8 we allow for 25% recovery. There are no instances when the capital ratio falls below the statutory 4% level. The effects on the system from foreign counterparties thus seem to be smaller than the effects from domestic counterparties. The foreign counterparties in these calculations have the same form of ranking as in the section on domestic exposures above.

The severity of losses on the capital ratios of Swedish banks is also lower for the foreign counterparties than for Swedish counterparties. There is a less severe effect with regard to both the number of cases where capital ratios fall below 4% and the actual capital ratios. We can only conclude that the possibility of direct contagion effects from foreign counterparties is very slight for the Swedish banking system.

Figure 7

Tier 1 capital in Swedish banks after the loss of their largest foreign counterparty, assuming no recoveries

In percentages

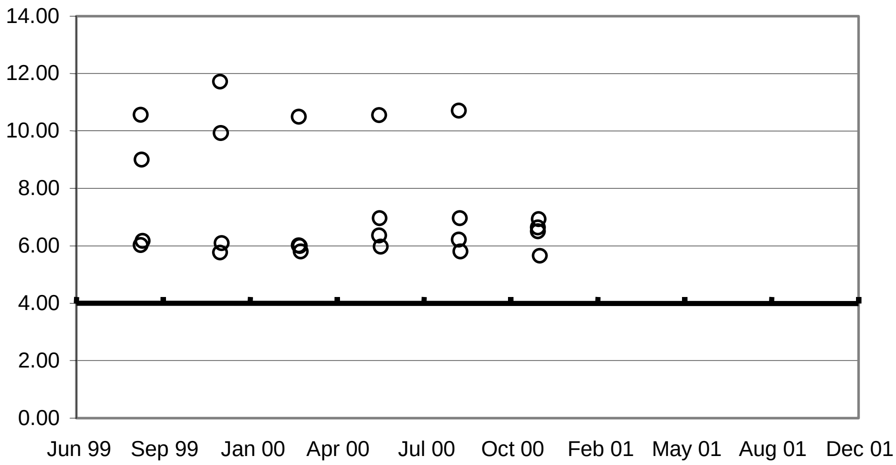


Source: Sveriges Riksbank.

Figure 8

Tier 1 capital in Swedish banks after the loss of their largest foreign counterparty, assuming 25% recoveries

In percentages



Source: Sveriges Riksbank.

3.5 Direct contagion from foreign exchange settlement

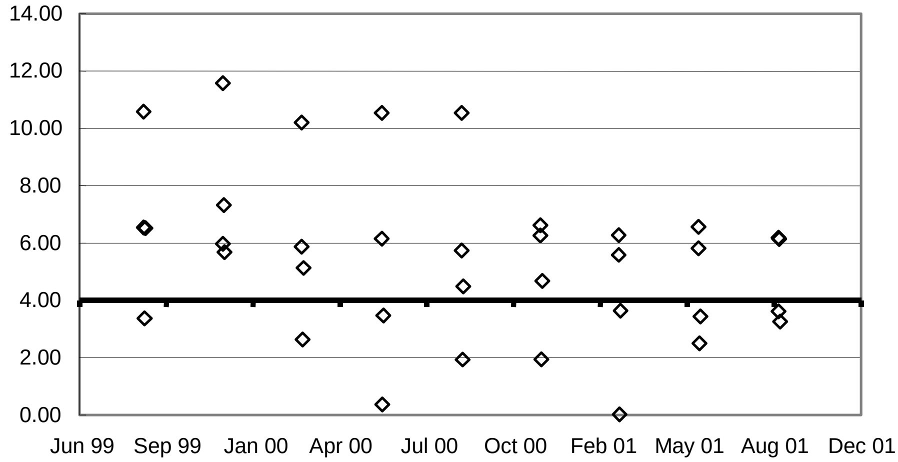
FX settlement exposure accounts for almost half of total exposures reported by the banks, which makes these exposures a likely channel for direct contagion. The effects on Swedish banks of losing their largest FX settlement exposures are calculated below. The counterparties in this case are Swedish and Nordic banks, large Swedish non-financial companies and some foreign financial companies.

The findings from the calculated Tier 1 capital ratios in Swedish banks after the loss of their largest FX exposures are that no fewer than 12 cases where capital ratios fall below the 4% threshold can be observed, assuming no recoveries. Assuming 25% recovery on the FX exposures limits the number of instances where the capital ratio falls below the statutory level to six. The number of cases where capital ratios fall below the statutory level when assuming 25% recovery decreases less than in the calculations above. This is because losses incurred by FX settlement exposures are larger than the losses above.

Figure 9

Tier 1 capital ratios in Swedish banks after the loss of their largest FX counterparty, assuming no recoveries

In percentages



Source: Sveriges Riksbank.

The size of FX settlement exposures differs markedly between the four major Swedish banks, as was the case with the size of exposures in the Swedish interbank market. The banks most at risk from FX settlement exposures are not the same banks as those most at risk from exposures to other Swedish banks. The fact that different banks have large exposures in the Swedish interbank market and the FX settlement market reduces the risk of direct contagion from one specific counterparty to several Swedish banks at the same time as the Swedish banks are vulnerable to defaults from different counterparties.

The risk of sequential direct contagion is a consequence of the possibility of one bank losing substantial amounts from the default of a foreign counterparty, the effect being that the bank itself defaults. Default by the first Swedish bank could then trigger another round of defaults among the others. This is the worst case scenario from a direct contagion perspective for the stability of the Swedish financial system.

The effects of FX settlement exposures are possibly the most severe in terms of direct contagion for the Swedish banks. The effect of defaults will diminish when foreign exchange settlement starts using PVP mechanisms within the CLS Bank.[15] The Swedish krona will not be one of the original currencies in CLS, but there are beneficial effects from trading USD/EUR on a PVP basis (Figure 11). The EUR/USD exposures reported by Swedish banks account for 19% of the total exposures, or SEK 125 billion. The effects of the krona being traded in the same way can also be assessed from Figure 11; exposures including the krona and one of the original currencies are at least 63% of total exposures and could possibly be even larger.[16] The effects of PVP in foreign exchange settlements would also

[15] For a description on CLS and how it will diminish settlement risk in foreign exchange trading, see Sveriges Riksbank (2001).

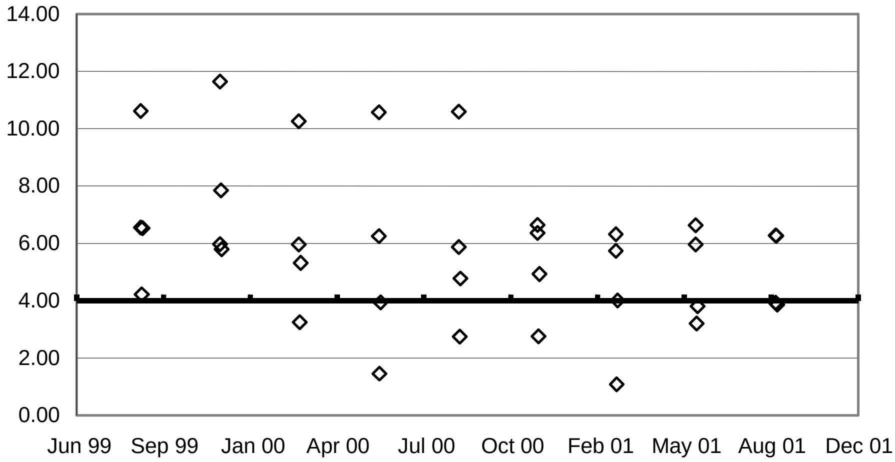
[16] Adding the exposures that are known to include SEK, USD and EUR, 11% + 33% + 19% = 63%.

reduce the level of exposures in the domestic interbank market and to foreign counterparties, as these markets also include FX settlement exposures to some extent.

Figure 10

Tier 1 capital ratios in Swedish banks after the loss of their largest FX counterparty, assuming 25% recoveries

In percentages

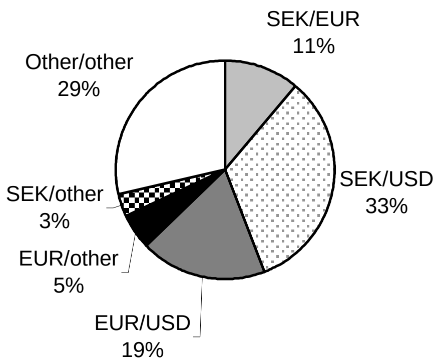


Source: Sveriges Riksbank.

Figure 11

Currency pairs

September 2001



Source: Sveriges Riksbank.

3.6 Liquidity impact

So far, the focus of the analysis of direct contagion has been on the solvency effect (ie the size of the loan loss) on Swedish banks, should one of their major counterparties default. A sudden default by a major counterparty would also comprise a liquidity effect, since repayment of the relevant claims on that counterparty would not occur. The potential liquidity impact on banks from counterparty exposures is hard to estimate, as the Riksbank's report does not cover the duration of the exposures. One can assume that the majority of exposures are of very short duration, but the duration of securities and derivatives could potentially be quite long. We therefore make the assumption that we can approximate the effects on the exposed banks' liquidity of a counterparty default by looking at the FX settlement and deposit classes of exposures. FX settlement exposures typically last for a maximum of two days. According to a survey of Swedish banks in 1998, the majority interbank deposits in Swedish banks are overnight and very few mature in more than one month. When assessing the liquidity effect on banks, it thus does not seem overwhelmingly conservative to assume that the total exposure in FX settlement and deposits to a single counterparty will be due for payment at very short notice.

Assessing the liquidity impact has so far not been part of the ongoing work at the Riksbank, but will be included in the future. Here, only a very simple calculation of the liquidity impact will be performed. The methods for doing this could probably be enhanced significantly. The effects on the liquidity of the Swedish banks have been calculated by comparing the exposure in deposits and FX settlement with data on unutilised collateral in RIX, the payment system. These calculations have been performed for the other major Swedish banks and for the largest FX settlement counterparty as reported by the banks. The full loss from a counterparty is related to the unused collateral in the payment system. If the loss is larger than the posted unused collateral, it is indicated in Table 1 below as a liquidity effect. The severity of the liquidity shortage varies considerably between the six cases.

Table 1

Liquidity effects on Swedish banks on 30 September 2001

Affected bank	**Failing bank**				
	Bank A	**Bank B**	**Bank C**	**Bank D**	**Largest FX counterparty**
Bank A	–				
Bank B		–			
Bank C		Liquidity effect	–		Liquidity effect
Bank D	Liquidity effect	Liquidity effect	Liquidity effect	–	Liquidity effect

Source: Sveriges Riksbank.

The results in Table 1 are only indicative of the possible liquidity effects, as the calculations are for one specific date. The calculations also do not take into account the fact that collateral in the Swedish payment system can be posted within minutes. The sale of other liquid assets by the bank could also mitigate liquidity effects. Another option is to borrow funds from other institutions, but in a situation where another Swedish bank has failed, this may be difficult since lenders may be reluctant to provide liquidity to a bank within the same system.

This very limited approach makes it hard to draw conclusions. However, to only take into account the collateral that is posted in the RIX system, which is readily available for immediate borrowing, is a very conservative approach. A very limited conclusion may be that it is a good sign that liquidity effects are not observed for all banks under this conservative approach.

4. Counterparty credit risk mitigation

Interbank credit exposures are often thought of as a necessary result of banking business, ie there is not much that can be done about these exposures by the banks. Especially in a concentrated banking system like the Swedish system, this is a common perception. In this section, the available methods for counterparty credit risk mitigation are briefly discussed, and it is shown that there are ways of diminishing counterparty credit exposures.

The most obvious credit risk mitigation technique is of course the setting of credit limits. There are substantial differences between the Swedish banks as regards the size of the exposures to counterparties they are willing to accept. This indicates that it is possible to set conservative credit limits, especially since these patterns are consistent over time in our data. In order to have conservative credit limits, it may be necessary to have an extensive network of counterparties, in order to diversify counterparty credit risk by using different counterparties, ie name diversification.

Swedish banks do not in general see FX settlement exposures as ordinary credit exposures. Before 1998, banks did not in general have any systems for limiting these exposures. Since then, the four largest Swedish banks have all introduced FX settlement limits. These limit systems are separate from the ordinary credit limit systems. It could be discussed whether these normal credit limits and FX settlement limits should be integrated, in order to have better control over total credit exposures within the bank.

The most important way of limiting FX settlement exposures is of course the introduction of a PVP mechanism for FX settlement. The creation of the CLS Bank is naturally a major step, which will decrease settlement exposures substantially. For the Swedish banks, however, the effect will not be that big initially, since the Swedish krona is not one of the original member currencies and a major part of Swedish banks' FX positions involve the krona (Figure 11).

As banks take on positions against each other on either side of the balance sheet, the scope for netting these exposures is important. Both positive and negative positions against the same counterparty could be netted, particularly in derivative positions. Master agreements[17] that allow for netting of derivative positions are commonly used by the Swedish banks and their most important counterparties in these markets. With respect to the positions reported to the Riksbank, netting reduces the credit positions by an average 55 to 60% for the 15 largest counterparties. It is more uncertain whether other kinds of exposures could be netted against each other in case of a failure.

Another obvious credit risk mitigation technique is the use of collateral. The most apparent area for this is financing, where banks can choose to lend to one another using uncollateralised deposits or collateralised transactions; in Sweden this is mainly done through repurchase agreements. Collateral is of course costly, and banks are not likely to always hold a sufficient amount of securities that can be used as collateral for all transactions. Another area where the use of collateral is growing is in derivatives trading. This applies especially to derivatives with long maturities, where posting collateral can be a very attractive way of hedging counterparty risk.[18]

5. Summary and policy conclusions

Sweden has a concentrated banking system, with four large banks covering at least 80% of the system, as in many other small countries. This is one reason to expect large interbank exposures within these systems, as banks may have few other alternatives than to deal with each other in the interbank markets. Data on interbank exposures shows that internal direct contagion effects are less than might have been expected in the Swedish banking system. In most cases where one of the four banks fails, the other banks will not suffer direct losses that would reduce their Tier 1 capital ratio below the regulatory level. However, this could occur on some occasions, according to the data set. Moreover, exposures are measured at the end of the quarter, so they are probably underestimated compared to exposures at peak levels, particularly intraday exposures. Therefore, a reduction of interbank exposures between the large Swedish banks is desirable in order to limit the risk of direct contagion within the Swedish system.

The risk of direct contagion from abroad mainly arises from foreign exchange settlement exposures. There are a number of cases where failure of a foreign counterparty causes one of the Swedish banks

[17] Master agreements in this context are derivatives contracts developed by industry organisations such as ISDA that allow for a standardised treatment of several derivatives deals between two counterparties, for instance regulating netting opportunities.

[18] For a discussion on the use of collateral and its implications, see CGFS (2001).

to be hit by a loss that makes its Tier 1 capital ratio decrease below the regulatory level. If FX settlement exposures are excluded, there are no cases where a Swedish bank will suffer a loss from abroad that leads to a Tier 1 capital ratio that is too low. The introduction of PVP mechanisms in foreign exchange settlement through the CLS Bank is a major advancement in risk reduction for banks active in the foreign exchange market.

Swedish banks show substantial differences with respect to the size of the individual exposures they are prepared to have to their counterparties. This indicates that it should be possible to reduce interbank exposures even in a concentrated banking system. It also leads to the conclusion that banks with large exposures in the interbank market are the ones we need to observe more closely.

The main ways to decrease the size of exposures between banks are to diversify exposures across more counterparties, to use collateralised instruments when possible, to adopt netting and to use clearing and settlement systems that provide for DVP or PVP when available. Many of the markets in which large exposures arise for the Swedish banks are international markets, where the concentrated national banking system does not pose an obstacle to diversification to a larger number of counterparties.

Swedish banks are universal banks that do not differ particularly from other large international banks. There is no reason to believe that banks in other countries differ substantially from Swedish banks with respect to exposure to direct contagion. The substantial differences with respect to the size of the largest exposures between Swedish banks suggest, however, that there may be significant differences in individual banks' exposure to direct contagion effects. One element that may lead to a larger exposure within the Swedish system compared to other countries is the substantial holdings of mortgage-backed bonds in Swedish banks. Most of the mortgage institutions are subsidiaries of the banks and are thus seen as part of the banks in the context of contagion.

The large Swedish banks have relatively high ratings and must in general be seen as rather risk conscious. The observation that banks take on exposures so large that they may not fulfil capital adequacy rules if there is a large loss on one of them suggests that the banks see sudden failure of an important counterparty as an extremely unlikely event. The reason is probably not merely the actual probability of the event occurring, but also expectations that the authorities would not allow sudden failure of an important bank. The fact that this kind of expectation exists is confirmed by the discussions that the Riksbank has had with the banks.

Moral hazard thus seems to be present with respect to exposure to direct contagion. As the fear of contagion is one of the most obvious reasons for public authorities to intervene, it is hard to see incentives for banks to decrease these exposures. To some extent, they are actually protected by the existence of risks of direct contagion, as these make government intervention more likely. Consequently, this can be seen as a market failure, which makes it reasonable to question whether there is scope for regulation in this area. In its FSSA for Sweden, the IMF stressed the importance of monitoring counterparty exposures, and suggested even more focus on these risks.[19]

In Sweden, the Riksbank has had discussions with the supervisory authority (FSA) on whether the rules on large exposures should be sharpened, in order to also take into account short-term interbank exposures. The conclusion has been not to do so at this stage. The reason is that the regulatory system is developed internationally, particularly within the European Union. The level playing field argument makes it difficult to suggest harder rules for national banks than are required by the EU system. It seems, therefore, more natural to bring up the issue in international discussions. However, the strong focus on Basel II, where these issues are not discussed, has made this quite difficult. Another reason not to introduce new rules at this stage is the creation of the CLS Bank. As quite a large portion of the contagion effects arises from FX settlement exposures, the total exposure to direct contagion might diminish substantially with the introduction of CLS. Instead of introducing stricter regulations, the Riksbank and the FSA will jointly increase the monitoring of banks' counterparty and settlement risk management, in particular the setting of credit limits. Monitoring credit limits can be an alternative to measuring actual exposures as the Riksbank currently does, especially since this may

[19] FSSA (Financial System Stability Assessment) is quite a new activity by the IMF, in which national financial systems are assessed on whether they subscribe to international standards and codes and whether the regulation and surveillance of the financial sector by the authorities are satisfactory.

be less burdensome for the banks involved and since the limits reveal the maximum exposure that the banks are willing to accept. On the other hand, individual limits reveal even more of the banks' business strategy than actual exposures, and banks may be even more reluctant to reveal this information.

Another way of improving counterparty exposure measurement would be to pick some of those counterparties that are commonly among the largest, and ask the banks to report their exposure on a day-to-day or even continuous basis. This would show whether there are high variations in exposures, and in particular whether exposures are underestimated in end-of-quarter reports, while at the same time not burdening the banks with the cumbersome work of ranking counterparties.

Another alternative to imposing stricter rules on large exposures is to consider whether it is possible to increase transparency in this area. If banks had to show their exposure to single counterparties in some form (of course without giving out the names of the counterparties), this ought to benefit the banks' investors, as it indicates the banks' ability to manage their risks. This information could be used to raise the required return on their investment or to drive down the size of exposures depending on the risk appetite of the investors.

To sum up, counterparty exposures and what they mean for systemic risk is an area where little work has been done. The Riksbank's measurement and analysis is a first step, as a means to understand the nature and the level of the problem in one particular banking system. However, more focus in the regulatory community and in the academic field would be warranted, since counterparty exposures are one of the major sources of systemic risk.

Appendix 1

	Counterparty	Derivatives	Securities	Deposits		Total	FX settlement	Stock loans	Repurchase agreements	Other collateralised loans	Exposures to companies within the same group
				Gross	Net						
1											
2											
3											
4											
5											
6											
7											
8											
9											
10											
11											
12											
13											
14											
15											
Total											

	Counterparty	SEK/EUR	SEK/USD	EUR/USD	EUR/Other	SEK/Other	Other/Other	Total
1								
2								
3								
4								
5								
6								
7								
8								
9								
10								
11								
12								
13								
14								
15								
Total 15								
Total								

References

Acharya, V (2001): “A theory of systemic risk and design of prudential bank regulation”, paper for conference on “Banks and systemic risk” held at the Bank of England, 23-25 May 2001.

Andersson, M and S Viotti (1999): “Managing and preventing financial crises”, Sveriges Riksbank, *Quarterly Review*, no 1.

Bank for International Settlements (2002): “Triennial central bank survey: foreign exchange and derivatives market activity in 2001”, March.

Committee on the Global Financial System (2001): “Collateral in wholesale financial markets: recent trends, risk management and market dynamics”, March.

Furfine, C (1999): “Interbank exposures; quantifying the risk of contagion”, *BIS Working Papers*, no 70, June, BIS.

IMF (2002): *Sweden: financial system stability assessment*.

Lind, G and J Molin (1999): “Towards new national and international banking regulations”, Sveriges Riksbank, *Quarterly Review*, no 3.

Sveriges Riksbank (2001): *Financial stability report*, no 2, November.

——— (2002): *The Swedish financial market* 2001.

Equity and bond market signals as leading indicators of bank fragility

Reint Gropp, Jukka Vesala and Giuseppe Vulpes,[1]
European Central Bank

Abstract

We analyse the ability of equity market-based distances-to-default and subordinated bond spreads to signal a material weakening in banks' financial condition. Using option pricing, we show that both indicators are complete and unbiased indicators of bank fragility. We empirically test these properties using a sample of EU banks. Two different econometric models are estimated: a series of logit-models, which are estimated for a number of different time-leads, and a proportional hazard model. We find support in favour of using both the distance-to-default and spread as leading indicators of bank fragility, regardless of our econometric specification. However, while we find robust predictive performance of the distance-to-default between six and 18 months in advance, its predictive properties are quite poor closer to the default. In contrast, subordinated debt spreads seem to have signal value close to default only. We also find that the predictive power of spreads appears to be weakened by implicit safety nets. We find no such evidence for the distances-to-default. Further, we find support for the notion that the market-based predictors of default have predictive power even controlling for balance sheet information and that both indicators may complement each other. We interpret our finding as providing some measure of support for the use of market information in supervisors' early warning models.

1. Introduction

From a supervisory perspective the securities issued by banks are interesting for two reasons: first, market prices of debt and equity may increase banks' funding cost and, therefore, induce market discipline, which may complement traditional supervisory practices (such as capital requirements and on-site inspections) in ensuring the safety and soundness of banks. The market may play a particularly useful role in disciplining the risks of large, complex and internationalised banking organisations. Second, supervisors are considering the use of market data to complement traditional balance sheet data for assessing bank fragility. Market prices may efficiently summarise information beyond and above that contained in other sources. Moreover, market information is available at a very high frequency. Supervisors could use these signals as screening devices or inputs into supervisors' early warning models geared at identifying banks which should be more closely scrutinised.[2] Recently, it has also been suggested that subordinated debt spreads might be used as triggers for supervisors' disciplining action (Evanoff and Wall (2000a), Flannery (2000)).

A number of studies have analysed whether the market prices of the securities issued by banks signal the risks incurred by them. If the prices reflect banks' risks this is taken as evidence that markets can

---

[1] Reint.Gropp@ecb.int, Jukka.Vesala@ecb.int, Giuseppe.Vulpes@ecb.int. Research assistance by Sandrine Corvoisier and Roberto Rossetti is gratefully acknowledged. The views expressed in this paper are solely those of the authors and not those of the ECB. We thank Allen Berger, Jürg Blum, Max Bruche, Vitor Gaspar, Myron Kwast, Mark Levonian, Simone Manganelli and seminar participants at the European Central Bank and the Basel Committee Research Task Force Conference on Applied Banking Research, which took place on 12-13 June 2001 in Oslo, for useful comments. All errors are the authors'.

[2] Supervisory early warning models combine a set of bank-level financial indicators (balance sheet, income statement and market indicators), as well as sometimes also other variables (eg macroeconomic conditions), to make a prediction about the future state of a bank. A growing number of supervisory agencies have been experimenting with this kind of model (see Gilbert et al (1999)).

indeed exert effective discipline on banks.[3] Studies using US data have found that banks' subordinated debenture spreads in the secondary market do reflect banks' (or bank holding companies') risks measured through balance sheet and other indicators (Flannery and Sorescu (1996), Jagtiani et al (2000), Flannery (1998 and 2000)). Morgan and Stiroh (2001) find the same to hold for the debenture spreads at issue. Sironi (2000) is the only study that we are aware of which provides evidence for European banks. He also concludes that banks' debenture spreads at issue tend to reflect cross-sectional differences in risk.

There is also some evidence that market signals could usefully complement supervisors' traditional information. Evanoff and Wall (2000b) find that subordinated debt spreads have some leading properties over supervisory CAMEL ratings. Conversely, DeYoung et al (2000) observe that on-site examinations produce information that affects the spreads. However, they find that spread changes reflect anticipated supervisory responses more often than new information. For example, bond investors in troubled banks react positively to increased supervisory oversight, hence substituting the market's own discipline. Finally, Berger et al (2000) conclude that supervisory assessments are generally less predictive of future changes in performance than equity and bond market indicators.

Finally, others have analysed the complementary role of the information contained in market prices vis-à-vis the information contained in rating agencies' assessments. Rating agencies are typically argued to be conservative and to respond mainly to risks which have already materialised (Altman and Saunders (2000)). Hand et al (1992) find that only unanticipated rating changes produce reaction in the US bond or equity markets (see also Goh and Ederington (1993)). Using European data, Gropp and Richards (2001) find that banks' bond spreads do not react to rating announcements, while equity prices do.

In general, research has focused on bond rather than equity market signals. This has been the case in part because mandatory subordinated debt issuance by banks has been prominently recommended as a new tool to discipline banks (eg Calomiris (1997)). The argument relies on the conjecture that subordinated debt holders have particularly strong incentives to monitor banks' risks, because they are uninsured and have junior status. In addition, signals based on equity prices are considered to be biased, because equity holders benefit from the upside gains that accrue from increased risk-taking (eg Hancock and Kwast (2001) and Berger et al (2000)). The relative importance of this moral hazard problem becomes the more pronounced the closer the bank is to insolvency, or the lower its charter value (eg Keeley (1990), Demsetz et al (1996), Gropp and Vesala (2001)).

However, as we will argue in this paper, there are several aspects which suggest that equity market signals may be attractive as monitoring devices. First, we show that unbiased equity-based fragility indicators can be derived. Second, there is broad consensus that the equity markets are efficient in processing available information. Empirical evidence strongly supports that equity holders respond rationally to news concerning: banks' asset quality (Docking et al (1997)), risks in LDC loans (eg Smirlock and Kaufold (1987), Musumeci and Sinkey (1990)), other banks' problems (eg Aharoney and Swary (1996)), or rating changes (ibid). Third, while bond spreads are conceptually simple, their implementation is difficult. For example, different bonds issued by the same bank may yield different estimates of the spread (Hancock and Kwast (2001)). Moreover, monitoring must concentrate on sufficiently liquid bonds in order to eliminate liquidity premia. In the European context, the construction of appropriate risk-free yield curves, which is a necessary ingredient to the calculation of spreads, may also be difficult especially for smaller countries, as further explained below.

In this paper, we first examine the properties of the market indicators in terms of their capability of capturing the major elements affecting default probability (*completeness*) and their alignment with supervisors' interests (*unbiasedness*). We show that a distance-to-default measure, derived using option pricing theory from the equity market data, is both complete and unbiased, as are uninsured bond yield spreads, provided that banks' asset value is still sufficiently high. Thus, these indicators are preferred over biased direct equity price-based measures and could represent useful leading indicators of bank fragility. The theory also suggests, however, that spreads may react only relatively late to a deterioration in the quality of a bank.

[3] A much less researched question is whether a higher cost of funds actually discourages banks' risk-taking. Bliss and Flannery (2000) identify some beneficial market influences, but do not find strong evidence that equity and especially bond investors regularly influence managerial action.

We then empirically test banks' distances-to-default and subordinated bond spreads in relation to their capability of anticipating a material weakening in banks' financial condition. We use two different econometric models: a logit-model and a proportional hazard model. We find support in favour of using both indicators as leading indicators of bank fragility, regardless of our econometric specification. However, while we find robust predictive performance of the distance-to-default indicator between six and 18 months in advance, its predictive properties are quite poor closer to default. In contrast, subordinated debt spreads are found to have signal value, but only close to default. This is consistent with the predictions of theory. Our results also indicate that the subordinated debt-based signals are powerful predictors only for smaller banks, which are generally not implicitly insured against default. In contrast and as expected, the public safety net does not appear to affect the predictive power of the distance-to-default. We also find evidence that both indicators provide additional information relative to balance sheet data alone, but our results also suggest some complementarity between market and balance sheet data. Finally, we find support for our theoretical prediction that the two indicators together have more discriminatory power in predicting defaults than each alone.

A key issue for this as for any similar study is the definition of events of major financial problems at banks, as formal bank bankruptcies have been extremely rare in Europe. The study uses as such events downgradings of the Fitch/IBCA individual rating to category C or below indicating a severe concern. This is a sensible approach, because individual ratings exclude the effect of possible public support and focus on the true condition of the bank and, moreover, the majority of banks in our sample received public support or experienced a major restructuring after such a downgrading. Hence, the problems were severe enough to warrant major remedial action, even though there was no formal bankruptcy. The robustness of this definition and its possible implications are discussed at length later on. If anything, our approach should bias our findings against finding predictive power for the indicators.

The remainder of the paper is organised as follows: Section 2 examines the basic properties of the equity and bond market indicators and frames our empirical propositions. Section 3 defines our sample and the variables used in the empirical study. Section 4 contains descriptive analyses of the behaviour of the market indicators. Section 5 reports our econometric specifications and results. Section 6 presents some extensions and robustness checks. Finally, Section 7 concludes.

2. Properties of market indicators

In order to structure the analysis of the market indicators, we introduce two basic definitions:

Definition 1: Completeness

An indicator of bank fragility is called complete if it reflects three major determinants of default risk: (i) the market value of assets (*V*), reflecting all relevant information about earnings expectations; (ii) leverage (*L*), reflecting the contractual obligations the bank has to meet (defined as the book value of the total debt liabilities (*D*) per the given value of assets (*D*/*V*)); and (iii) the volatility of assets (σ), reflecting asset risk.

Definition 2: Unbiasedness

An indicator of bank fragility is called unbiased if it meets:

$$\begin{array}{ll} (i) & \dfrac{\partial Ind}{\partial V} < 0 \\ (ii) & \dfrac{\partial Ind}{\partial L} > 0 \\ (iii) & \dfrac{\partial Ind}{\partial \sigma} > 0 \end{array} \qquad (1)$$

where *Ind* may represent any fragility indicator. The conditions require the indicator to be decreasing in the earnings expectations, and increasing in the leverage and asset risk. Definition 1 follows the usual approach in the commercial applications to define default risk measures (eg KMV Corporation (1999)). Definition 2 is more novel in this context and requires that any fragility indicator be aligned with supervisors' conservative perspective. Hence, we would argue that only complete and unbiased indicators would be appropriate as early warning indicators of bank fragility, since only indicators with these two properties would fully and appropriately reflect the elements affecting default probabilities of banks.

We use option pricing theory and the valuation of equity and debt securities as a helpful tool to demonstrate some key properties of market-based fragility indicators. We consider a bank liability structure that consists of equity (E) and junior subordinated debt (J), and also some senior debt (I). This allows us to study the properties of the subordinated debt spreads directly. At the maturity date (T), payments can only be made to the junior claimants if the full promised payment has been made to the senior debt holders. To illustrate some of the basic concepts used below, suppose that both classes of debt securities are discount bonds and that the promised payments (book values) are I and J, respectively. ($D = I+J$) equals the total amount of debt liabilities. At the maturity date, the payoff profile of each security is as shown in Chart 1, depending on the asset value. To simplify notation, we assume that time to maturity equals T at the time of valuation of the equity and debt securities.

2.A Equity-based indicators

Equity holders have the residual claim on a firm's assets and have limited liability. As first realised by Merton (1977), equity can be modelled as a call option on the assets of the firm (here a bank), with a strike price equal to the total book value of the debt (see Chart 1). Thus, option pricing theory can be used to derive the market value and volatility of assets from the observable equity value (V_E) and volatility (σ_E), and D. Consider the basic Black and Scholes (1973) formula, valuing equity as:

$$\begin{aligned} V_E &= VN(d1) - De^{-rT}N(d2) \\ \sigma_E &= \left(\frac{V}{V_E}\right)N(d1)\sigma \\ d1 &\equiv \frac{\ln\left(\frac{V}{D}\right) + \left(r + \frac{\sigma^2}{2}\right)T}{\sigma\sqrt{T}} \\ d2 &\equiv d1 - \sigma\sqrt{T} \end{aligned} \tag{2}$$

where N represents the cumulative normal distribution, r the risk-free interest rate, and T the time to the maturity of the debt liabilities.

We can see from (2) that V_E is complete, since market prices reflect the relevant information for capturing default risk (V, D and σ). However, V_E is increasing in σ, which violates condition (iii) in (1). Therefore, an increase in the share price may not be consistent with a reduction in default risk.

However, as an alternative consider the negative of the distance-to-default ($-DD$),[4] which we derive from the Black-Scholes model in Appendix I:

$$(-DD) = -\frac{\ln\frac{V}{D} + \left(r - \frac{\sigma^2}{2}\right)T}{\sigma\sqrt{T}} = -\frac{\ln\left(\frac{1}{L}\right) + \left(r - \frac{\sigma^2}{2}\right)T}{\sigma\sqrt{T}} \tag{3}$$

V and σ are solved from the non-linear two equation system (2). DD indicates the number of standard deviations (σ) from the default point at maturity ($V = D$). From (3) we can obtain a first result:

[4] A similar measure is the basic conceptual ingredient in KMV Corporation's model for estimating default risk (see KMV Corporation (1999)).

Result 1

(-*DD*) is a complete and unbiased indicator of bank fragility for *V*>*V*' (given *D*). *V*' is defined as $De^{-(1/2\sigma^2+r)T}$.

Proof

(-*DD*) reflects *V*, *L* and σ; hence it is complete.

Clearly, $\frac{\partial(-DD)}{\partial V} < 0$ and $\frac{\partial(-DD)}{\partial L} > 0$. $\frac{\partial(-DD)}{\partial \sigma} = \frac{1}{2}\sqrt{T} + \sigma^{-2}T^{-1/2}\left(\ln\left(\frac{V}{D}\right) + rT\right) > 0$, when $V > De^{-(1/2\sigma^2+r)T}$.

(–*DD*) meets all the conditions in (1) when *V* is sufficiently large (given the amount of debt); hence, it is unbiased for *V*>*V*'.

(-*DD*) is unbiased for all positive values of *DD*, ie always when above the default point, since *DD*>0 when $V > De^{(1/2\sigma^2-r)T}$.[5] Hence, (-*DD*) is a complete and unbiased early warning indicator for all banks which are still solvent.

2.B Subordinated debt-based indicators

In determining the value of debt, it is important to explicitly account for subordination, since the payoff profile of the subordinated debt is different from the senior debt. Following Black and Cox (1976), the observable market value of subordinated debt (V_J) can be derived as a difference between two senior debt securities with the face values of (*I*+*J*) and *I*, and respective market values of (V_{I+J}) and (V_I) (see Chart 1):

$$V_J(V,D,\sigma,T) = V_{I+J}(V,I+J,\sigma,T) - V_I(V,I,\sigma,T). \tag{4}$$

The value of the individual senior debt securities can be expressed using the standard Merton (1990) option pricing formula. The value of the debt security (*I*+*J*) is affected by total leverage and equals:

$$V_{I+J} = (I+J)e^{-rT}\left(N(h_2(I+J)) + \frac{1}{Le^{-rT}}N(h_1(I+J))\right),$$

$$h_1(I+J) \equiv \frac{-1/2\sigma^2 T + \ln\left(\frac{(I+J)e^{-rT}}{V}\right)}{\sigma\sqrt{T}},\ h_2(I+J) \equiv \frac{-1/2\sigma^2 T - \ln\left(\frac{(I+J)e^{-rT}}{V}\right)}{\sigma\sqrt{T}}. \tag{5A}$$

The other senior security (*I*) is valued as:

$$V_I = Ie^{-rT}\left(N(h_2(I)) + \frac{V}{Ie^{-rT}}N(h_1(I))\right) \tag{5B}$$

with $h_1(I)$ and $h_2(I)$ analogous to (5.A). Finally, the yield to maturity ($y(T)$) is defined from:

$$e^{-y(T)T}J = V_J,\ \text{ie}\ y(T) = -\frac{1}{T}\ln\left(\frac{V_J}{J}\right) = -\frac{1}{T}\ln\left(\frac{V_{I+J}-V_I}{J}\right) \tag{6}$$

and the spread over and above the risk-free yield to maturity of the subordinated debt (*S*) equals $y(T)$-$r(T)$. *S* is equivalent to a credit risk premium, in the absence of any liquidity premia.

Based on (5) and (6) we can state a second result:

[5] Note that *V* can be somewhat less than *D* (*V*/*D* less than one) at the default point (*DD*=0) because of the drift and the interest rate effects at the time of valuation (<*T*).

Result 2

S is a complete and unbiased indicator of bank fragility for *V>V** (given *D*=I+J). *V** is defined as $[I(I+J)]^{1/2}e^{-(1/2\sigma^2+r)T}$.

Proof

By (5) and (6), *S* reflects *V*, *L* and σ; hence, it is complete.

Unbiasedness:

$\frac{\partial S}{\partial V}=\frac{\partial y(T)}{\partial V}=-\frac{J}{TV_J}\left[\frac{\partial V_{I+J}}{\partial V}-\frac{\partial V_I}{\partial V}\right]$. Following Merton (1990), the value of a senior debt security is an increasing function in the value of assets, and it turns out that $\frac{\partial V_{I+J}}{\partial V}=N(h_1(I+J))$, and $\frac{\partial V_I}{\partial V}=N(h_1(I))$. Thus, $\frac{\partial S}{\partial V}=-\frac{J}{TV_J}[N(h_1(I+J))-N(h_1(I))]$. The expression in the square brackets is always positive, because h_1 is increasing in the face value of debt. Since *J* and V_J are always positive, $\frac{\partial S}{\partial V}<0$ always.

Second, $\frac{\partial S}{\partial L}=-\frac{J}{TV_J}\left(\frac{\partial V_{I+J}}{\partial L}\right)$. Since $\frac{\partial V_{I+J}}{\partial L}=-N(h_1(I+J))L^{-2}<0$, $\frac{\partial S}{\partial L}>0$ always.

Third, $\frac{\partial S}{\partial \sigma}=-\frac{J}{TV_J}\left(\frac{\partial V_J}{\partial \sigma}\right)$. Thus, the sign of $\frac{\partial S}{\partial \sigma}$ is the opposite of the sign of $\frac{\partial V_J}{\partial \sigma}$. According to Black and Cox (1976, p 360), V_J is a decreasing (increasing) function of σ for *V* greater than (less than) the point of inflection, *V**. Thus, for *V>V**, $\frac{\partial S}{\partial \sigma}>0$.

Hence, *S* is unbiased for *V>V** as it meets all the conditions in (1), and biased for *V<V** by condition (iii).

*V** is a geometric average of (*I+J*) and *I* ("adjusted" for time to maturity, drift and interest rate effects), falling between the two face values (see Chart 1).[6] When the value of bank assets is high enough to cover both senior and junior debt, the interests of the senior and junior debt holders are aligned with each other and with the interests of the supervisor. Hence, when the bank is economically solvent (and equity has some value), the subordinated debt spread is an unbiased indicator of bank fragility. Since banks are likely to be monitored while being still sound enough to cover all debt, the spread can constitute a useful early indicator of deterioration in financial condition.

However, one should note that when the value of assets is lower than the threshold value *V** (which is to some extent below the total value of debt, depending on the amount of junior debt), the two groups of debt holders have conflicting interests. The junior claimants have interests similar to those of the equity holders to take on more asset risk, while the senior claimants' expected payoff is always decreasing in risk.[7]

The above investigation of the properties of the market signals is made in the context of a specific model: normal asset value diffusion and European option type (call for equity and put for debt). Namely, the market value of a debt instrument can also be expressed on the basis of the discounted value assuming no default risk and the value of a put option on the firm's assets (see Merton (1977) and Ron and Verma (1986)). The widespread use of the Merton model, also to generate quantitative probability of default estimates, speaks in favour of it. But unfortunately, the literature has not

---

6 Note that *V*<V'* as long as there is some junior debt outstanding.

7 This effect has an impact on the role of subordinated debt holders in disciplining banks' risk-taking: the contribution can be actually negative once the bank has entered the zone of de facto insolvency. In this zone, the sole right to approve business policies should lie with the senior debt holders (or supervisors) in order to avoid moral hazard. Levonian (2001) also makes the point that the incentives of the subordinated debt holders do not always side with those of the supervisors.

established general conditions under which the unbiasedness property could be established and verified for specific asset-liability structures (eg for banks). Thus, the performance of the market signals is ultimately an empirical issue.

Notwithstanding this general point, the crucial feature that, say, the call option value is (monotonically) increasing in V and decreasing in L seems to be a much more general result than the monotonic and increasing relationship between the option value and σ in the Merton model, which produces the often cited equity price bias. This result may not obtain for certain ranges of V under different (and possibly more plausible) distributional assumptions, eg based on bounded returns (Bliss (2000)), more complex liability structures, or under different option types, eg barrier options (Bergman et al (1996)).[8] Hence, alternative modelling assumptions would tend to question the universal biasedness of the simple equity price-based indicators, rather than the unbiasedness of the *DD* or *S*-measures.[9]

The main concern of this paper is indeed an empirical one: whether complete and unbiased market indicators (as derived from a specific model) are capable of signalling an increase in the default risk in a timely fashion.[10] Traditional accounting measures such as leverage ratios or earnings indicators are generally incomplete and therefore less useful as indicators of bank fragility. Thus, the key proposition whose validity we test is as follows:

Proposition 1

The equity market-based (-DD) and the bond market-based S constitute early indicators of a weakening in a bank's condition.

Finally, it is of interest to study how the subordinated debt spread behaves as a function of the asset value (or the distance-to-default) to see how the spread would be predicted to react to a deterioration in financial condition. According to Black and Cox (1976), the subordinated debt value is an increasing and concave function of V for $V>V^*$, like senior debt. Hence, the spread is a convex and decreasing function of V for $V>V^*$. This means that the spread would remain stable and close to zero for large intervals of changes in V and only react significantly relatively close to the default point.[11] This can be illustrated by plotting the spread as a function of the distance-to-default (varying V, holding I,J constant), under specific assumptions for the other parameters (see Chart 2). While the subordinated debt spread reacts earlier and more than the senior debt spread, it moves up significantly only when *DD* is relatively low.

Hence, the equity-based distance-to-default measure can be expected to provide an indication of a weakening financial condition earlier than the subordinated debt spread. This is a direct consequence of the different payoff structures of the equity and subordinated debt holders (for $V>V^*$). Debt holders care only about the left tail of the distribution of returns, while equity holders are interested in the whole distribution of returns. In a nutshell, the theory predicts that the two indictors have qualitatively different predictive properties, because the response of the spreads to an increase in default probability is non-linear. Therefore, the distance-to-default measure would be predicted to deliver an earlier signal of fragility than the spread. In the empirical analysis, we examine the performance of (-*DD*) and *S* with respect to different time leads under the proposition that:

---

8 There does not seem to be consensus about how to model the distribution of bank asset returns. Ritchken et al (1993) find some consistency between the behaviour of bank equity and the outcomes from a barrier option framework.

9 The analysis also relies on the idea that asset risk can be measured by asset variance, which seems to be relatively uncontested, while alternative approaches have also been proposed (foremost Harrison and Kreps (1979)).

10 Empirical evidence has suggested that the actual spreads are higher than suggested by Merton's model. Franks and Torous (1989) and Longstaff and Schwartz (1995) argue that an additional element in the spread is the expectation that equity holders and other junior claimants receive in the bankruptcy settlement more than what is consistent with absolute priority. In addition, Anderson and Sundaresan (1996) suggest that debt holders are forced to accept concessions to pay less than originally agreed prior to formal bankruptcy proceedings. Mella-Barral and Perraudin (1997) incorporate this strategic debt service into an option pricing-based model and show that the spread widening impact can be significant.

11 Bruche (2001) shows that the "hockey-stick" shape of the spread as a function of V can become more pronounced when one introduces into the basic pricing model asymmetric information and investors' coordination failure.

Proposition 2

The equity market-based (-DD) constitutes an earlier indicator of weakening in a bank's condition than S. S would react significantly only relatively close to the default point.

2.C Impact of the safety net

Following Merton (1977), the value of subordinated debt can be expressed in terms of two "no default risk" values for the senior debt securities (*I+J*) and *I* and two put option values (strike prices equalling the book values of debt as before).[12] A put option represents the value of the limited liability, ie equity holders' right of walking away from their debts in exchange for handing over the firm's assets to the creditors. In case of fully insured debt (like insured deposits), the put option component disappears, and the market value of the debt equals the "no-default-risk" value (and *S* is zero). There is no signal of fragility obtainable from the pricing of this debt. Hence, any market discipline requires that deposit insurance is explicitly restricted, leaving out some creditors with their money at stake (eg Gropp and Vesala (2001)).[13]

The literature (eg Dewatripont and Tirole (1993)) has also examined the problem related to the *credibility* of the restricted safety net. Losses from a failure of a significant bank might affect the banking system as a whole and, hence, imply systemic risk. In this case, it might be expected that the "systemic" banks would never be liquidated, or that the exposures of the systemically relevant debt holders (such as other banks) would always be covered, regardless of the features of the explicit safety net arrangements ("too big to fail"). If the implicit safety net is perceived to be unrestricted, the value of the put option is zero, since the debt holders would not face the risk of having to take over the assets of the bank. Thus, the market value of debt would again be equal to the "no default risk" value and all uninsured debt-based fragility indicators would be incomplete and fail to capture increased default risk.

The *perceived probability of bailout* will generally be less than one, since there is typically no certainty of public support under an explicitly restricted deposit insurance system. Authorities frequently follow a policy of constructive ambiguity in this regard. Under these circumstances debt-based indicators would have predictive power, but much less compared to a hypothetical completely uninsured case. In this context we take the existence of positive spreads on banks' uninsured debt issues as evidence that the perceived probability might be indeed less than one. However, the history of bank bailouts by the government (significant banks have not failed in Europe in recent history) suggests that spreads might nevertheless be substantially weakened in their power to lead banking problems as compared with the case where the absence of bailouts is fully credible. Gropp and Vesala (2001) find empirical support for this point. Their results suggest that banks' risk-taking in Europe was reduced in response to the introduction of explicit and restricted deposit insurance schemes. They also find evidence in favour of the notion that a number of banks are "too big to fail". In addition, Gropp and Richards (2001) find that banks' bond spreads do not appear to react to ratings announcements. Their findings could be interpreted as evidence in favour of widespread safety nets. After an extensive sensitivity analysis, they cannot exclude the possibility that bondholders expect to be insured against default risk in Europe.

As a rule, equity holders are not covered even in broad-based explicit safety nets. In addition, the existence of an implicit safety net would induce banks to take on increased leverage and asset risk, and these risk-taking incentives (moral hazard) would be the greater the more extensive the perceived safety net (see Gropp and Vesala (2001, Section 2)). While bond market indicators would not reflect this additional effect under a broad safety net, correctly specified equity indicators, such as (-*DD*), would.

---

12 For instance, $V_I = V_I e^{-y(T)T} = V_I^{RF} - V_{I,PO} = V_I e^{-r(T)T} - V_I e^{-r(T)T} N(-h2(I)) + VN(-h1(I))$, where $V_I^{RF} = Ie^{-r(T)T}$ denotes the "no-default-risk" value and V_{PO} the value of the put option.

13 The put option value also represents the value of the deposit insurance guarantee, since by guaranteeing the debt the guarantor has in fact issued the put option on the assets (see Merton (1977)). Hence, the deposit insurance value (V_{PO}) could also be used as an unbiased bank fragility indicator (see Bongini et al (2001)) with the same characteristics as the market value of debt-based indicators.

Hence, we can formulate an additional proposition:

Proposition 3

If a bank were covered by an implicit or explicit partial guarantee, the bond spread S would be a weaker leading indicator of bank fragility than the negative distance-to-default (-DD).

Whether equity and bond markets are able to effectively process the available information and send early signals which are informative of banks' default risk is investigated below in a sample of European banks. We evaluate the usefulness of the preferred (complete and unbiased) market indicators (*-DD* and *S*) for this purpose (Proposition 1). We also test whether the spread reacts later than (*-DD*) (Proposition 2), and whether a perception of the safety net dilutes the predictive power of the bond market signals, but leaves the equity market signals intact (Proposition 3).

3. Empirical implementation

Our data set consists of monthly observations from January 1991 to March 2001. The relatively high frequency of the data highlights one fundamental advantage of market-based indicators relative to balance sheet indicators. We decided to use monthly data, rather than an even higher frequency, in order to eliminate some of the noise in daily equity and bond prices. The data set consists of those EU banks for which the necessary rating, equity and bond market information is available. In the sample selection process we started from roughly 100 EU banks which had obtained a "financial strength" rating from Fitch/IBCA.[14] The sample size was then largely determined by the availability of market data. The two subsamples used in evaluating the equity and bond market signals consist of 84 and 59 banks, respectively (see Table 1). The samples contain banks from 14 (equity sample) and 12 (bond sample) EU countries.

3.A Measurement of bank "failures"

We were faced with the problem that no European banks formally declared bankruptcy during our sample period. In the absence of formal bank defaults, we considered a downgrade in the Fitch/IBCA "financial strength" to C or below as an event of materially weakened financial condition.[15] There are 25 such downgrades in the equity and 19 in the bond subsample, 32 in total (Table 2). We defend our definition of bank "failure" on two grounds: first, the "financial strength rating" is designed to exclude the safety net and, hence, should indicate the bank's true financial condition. A downgrade to the level of C or below signifies that there are significant concerns regarding profitability and asset quality, management and earnings prospects. In particular when the rating falls to the D/E category very serious problems are indicated, which either require or are likely to require external support. Second, in many cases after the downgrade to C or below, public support was eventually granted or a major restructuring was carried out to solve the problem. As detailed in Table 2, 11 banks received public support and eight banks underwent a major restructuring after the downgrading. The support or restructuring operations also generally took place relatively soon after these events (six to 12 months). In the remaining cases, no public support or substantial restructuring took place. In part this is a reflection of sample truncation in March 2001, as six of the remaining 13 downgrades took place in late 2000 or early 2001 and an eventual intervention cannot be excluded. Given that the downgrades precede the actions aimed at resolving the problem by quite some time, we would argue that our proxy for bank failures is quite sensible and generally should bias our results against finding predictive power of the indicators.

Our study is similar to the US studies investigating the relationship between market information and supervisory ratings (for example Evanoff and Wall (2000b), DeYoung et al (2000) and Berger et al (2000)), while we use the "individual" ratings as signals of banking problems. While we are concerned

---

[14] For an explanation of a "financial strength" rating see below.

[15] See Appendix 2 for the exact definitions of the Fitch/IBCA rating grades.

about our relatively small sample sizes (at least in terms of number of banks, not in terms of data points; see below) Evanoff and Wall (2000b), for example, consider 13 downgrades in supervisory CAMEL ratings in a sample of 557 US banks, constituting the default events. Hence, compared to the previous literature our sample appears reasonably large and fairly balanced. Further, rather than use the Fitch/IBCA ratings, it could be argued that we should use supervisory ratings (such as CAMEL ratings) instead. Unfortunately, we did not have access to historical supervisory information on individual banks and, in some European countries, comparable ratings by supervisors do not exist. Clearly, the supervisory ratings may be based on more detailed information relative to ratings by a ratings agency, including confidential information obtained at on-site inspections, but they may also be subject to forbearance.

3.B Market indicators

We calculated the negative of the distance-to-default (-*DD*) for each bank in the sample and for each time period (t) (ie month) using that period's equity market data. The system of equations in (2) was solved by using the generalised reduced gradient method to yield the values for V_A and σ_A, entering into the calculation of (-*DD*). Variable definitions are given in Table 3 and descriptive statistics in Table 4.

As to the inputs to the calculation of (-*DD*), we used monthly averages of the equity market capitalisation (V_E) from Datastream. The equity volatility (σ_E) was estimated as the standard deviation of the daily absolute equity returns and we took the six-month moving average (backwards) to reduce noise (as eg in Marcus and Shaked (1984)). The presumption is that the market participants do not use the very volatile short-term estimates, but more smoothed volatility measures. This is not an efficient procedure as it imposes the volatility to be constant (it is stochastic in Merton's original model). However, equity volatility is accurately estimated for a specific time interval, as long as leverage does not change substantially over that period (see for example Bongini et al (2001)). The total debt liabilities (V_L) are obtained from published accounts and are interpolated (using a cubic spline) to yield monthly observations. The time to the maturing of the debt (T) was set to one year, which is the common benchmark assumption without particular information about the maturity structure. Finally, we used the government bond rates as the risk-free rates (r).[16] The values solved for V and σ were not sensitive to changes in the starting values.

We largely followed convention when calculating the monthly averages of the secondary market subordinated debt spreads (S). We used secondary market spreads, rather than those from the primary market, as we would argue that secondary market spreads are more useful for the ongoing monitoring of bank fragility. In the absence of mandatory issuance requirements, such as those proposed by Calomiris (eg 1997), banks' new issuance could be too infrequent, or limited to periods when pricing is relatively advantageous. As we were concerned about too thin or illiquid bank bond markets in Europe, we only selected bonds with an issue size of more than €150 million. This figure seemed the best compromise between maintaining sample size and obtaining meaningful monthly price series from Bloomberg and Datastream, which were our main data sources. In addition, in order to minimise noise in the data series, we attempted to use straight fixed rate subordinated debt issues only. We were largely able to obtain such bonds, but in some cases we had to permit floating rate bonds into the sample. We used the standard Newton iterative method to calculate the bond yields to maturity.

For the larger countries, we were able to find bank bonds issued in the domestic currency which met our liquidity requirement. In the case of smaller countries, banks more frequently issued foreign than domestic currency denominated bonds prior to the introduction of the euro. Hence, we largely resorted to foreign currency issues (Deutsche Mark, euro, US dollar and, in two cases, yen) and matched them to government bonds issued in the same currency. We were able to construct risk-free yield curves for Germany, France and the United Kingdom and calculated spreads for banks in those countries relative

---

[16] Our (-*DD*) measure is subject to the Black-Scholes' assumption of a cumulative normal distribution (N) for the underlying asset values. As pointed out by Bliss (2000), this assumption may not hold in practice. He argues that the normal distribution does not take into account that closer to the default point adjustment in debt liabilities is likely to take place. Hence, empirically better formulas could be found, while delivering fragility indicators with similar qualitative characteristics as the standard (-*DD*).

to the corresponding point on those curves. For the other smaller countries, we were unable to obtain sufficient data to construct full risk-free yield curves. We therefore instead matched the remaining term to maturity and the coupon of the bank bond to a government bond issued by the government of the country of the bank's incorporation in the same currency.

3.C Expectation of public support

We use the "support rating" issued by Fitch/IBCA to indicate the likelihood of public support. We regard as cases of more likely public support the rating grades 1 or 2 (see Appendix 2). The former grade indicates existence of an assured legal guarantee, and the latter a bank for which in Fitch/IBCA's opinion state support would be forthcoming. This could be, for example, because of the bank's importance for the economy. Hence, the likelihood of support could depend on the size of the institution ("too big to fail"), but a bank could also be possibly "systemically" important for other reasons. The weaker "support ratings" (from 3 to 5) depend on the likelihood of private support from the parent organisation or owners, rather than from public sources. The share of banks with a "support rating" of 1 or 2 is quite high (around 65% in the equity sample and 80% in the bond sample). This is not surprising, since we are considering banks with a material securities market presence as an issuer. These banks tend to be significantly larger, again as expected, than those with a rating of 3 to 5. Their average amount of total debt liabilities is roughly 10 times higher.

3.D Sample selection

Before we present the results, it may be worthwhile to examine the sample in a little more detail, in particular with respect to sample selection issues. The first question that arises relates to the relevant universe of banks. For the bond sample, the universe is determined by those EU banks that were rated by Fitch/IBCA during the 10-year period under investigation.[17] Out of this total, those banks remained in the sample for which we were able to calculate bond spreads, ie for which sufficiently liquid and sizeable bonds were outstanding and the data were available in Bloomberg. Hence, relative to the universe of 103 rated banks, we were able to obtain meaningful bond price data for 59 banks. Sample selection issues may be a problem if the banks in the sample differ in their likelihood of failure relative to those in the universe of banks. In particular, we were concerned that we had tended to over-sample failures. It turns out that this is not the case. The probability of failure during the sample period is around 33% both in the universe and in the sample. Nevertheless, the banks in the sample may differ in other important criteria from those in the universe. For example, given our requirement that the bank must have substantial subordinated debt outstanding, the banks in the sample may be larger than those in the universe. This is the case, although the difference is not statistically significant. Finally, a bias may arise due to differences in data availability of the banks in the sample. If banks that eventually fail remain in the sample for only a relatively short period of time prior to failure, the proportional hazard model may overstate the predictive power of indicators. There could be a number of reasons for this problem. One, given that we chose a fixed starting point for our sample (1991) and given that naturally all failed banks drop out after failure, the time period that non-failed banks remain in the sample is longer. This by itself should not constitute a problem for the estimation. However, if failures occur disproportionately at the beginning of the sample period, ie in 1991-94, this could result in overstating the predictive power of our indicators in the proportional hazard model. However, the average time period in the sample for banks which eventually failed is 34 months. This should give us ample data to obtain unbiased estimates.[18]

In the case of the stock price sample, we would argue that the relevant universe is somewhat smaller. Again taking those banks which had obtained a rating from Fitch/IBCA as the starting point, the universe of banks is further reduced by banks which are not listed on a major European stock

---

[17] Clearly, this universe is substantially different from the notion of all EU banks. For small, non-traded banks, such as savings banks or cooperative banks, the idea of the importance of market indicators is clearly not relevant. In any event, we would argue that market indicators are precisely of most use in the case of large, complex financial institutions, because for these, balance sheet information may be more difficult to interpret.

[18] The average period in the sample of non-failing banks is, of course, longer with 76 months. Note that the maximum number of observations per bank is limited by our sampling period to 131 months.

exchange. It turns out that this concerns 11 banks. Of the remaining 92 banks, our sample contains 83 banks. The difference of nine banks is due to the unavailability of a stock price series in Datastream. The probability of failure in the sample is identical to that in the universe at one third. Again, we were concerned whether we observe the failing banks long enough to make meaningful inferences from the proportional hazard model. The average time of banks which eventually fail in the stock price sample is one month longer than those in the bond price sample, namely 35 months (non-failing banks: 73 months). Again, we feel that this should give us sufficient data to estimate the model.

4. Descriptive statistics

We constructed the sample for the empirical analysis as follows. For each month (t) of a downgrading ("default") event, we took all non-downgraded banks as a control sub-sample, and calculated all variables for both sub-samples with specified leads of x months.

As a first cut at the data, we conducted simple mean comparison tests to assess whether (-*DD*) and *S* are able to distinguish weaker banks within our data set. We also examined whether the indicators could lead the downgrading events by performing the mean comparison tests for various time leads (lead times of three, six, 12, 18 and 24 months). The results reported in Table 5 indicate that the banks that were downgraded had a significantly higher mean value of (-*DD*) than the non-downgraded banks up to and including 24 months prior to the downgrading events. We also find in the second panel that the banks that were downgraded had higher prior spreads (*S*) and that the spreads of the "defaulted" banks clearly increase as the "default" event is approached. However, the difference between "defaulted" and "non-defaulted" banks is never statistically significant when the full sample is considered. This suggests that *S* is a weaker leading indicator of bank fragility than (-*DD*).

The "default" indicators reflect two factors: first, the bank's ability to repay out of its own resources, and, second, the government's perceived willingness to absorb default losses on behalf of private creditors (see eg Flannery and Sorescu (1996)). Hence, in the third panel of Table 5 we limit the sample to those banks with a support rating of 3 or higher. We only present the t-tests up to x equals 12 months in order to maintain some sample size. Nevertheless, the figures given here should be interpreted with care, as even so sample sizes are small. The results offer further evidence that a safety net expectation can dilute the power of the spreads to reflect bank fragility, while there is no apparent impact on the distances-to-default. In this limited sample, there is now a significant difference in the mean values of *S* between "defaulted" and "non-defaulted" banks. Also in absolute terms, the difference in the average spreads is now higher.

5. Empirical estimation

5.A Estimation methods

We used two different econometric models to investigate the signalling properties of the market-based indicators of bank fragility. The first is a standard *logit-model* of the form:

$$\Pr[STATUS_t = 1] = \psi(\alpha_0 + \alpha_1 DI_{t-x} + \alpha_2 DSUPP_{t-x} * DI_{t-x}) \qquad (7)$$

where $\psi(\,)$ represents the cumulative logistic distribution, DI_{t-x} the fragility indicator at time t-x, and

$$STATUS_t = \begin{cases} 1 \text{ if bank was downgraded to C or below at time } t \\ \quad 0 \text{ otherwise} \end{cases}.$$

We estimate the model for different horizons separately, ie we investigate the predictive power of our two indicators three, six, 12, 18 and 24 months before the downgrading event. Generally, we would expect the predictive power to diminish as we move further away from the event. Significant and positive coefficients of the lagged market indicators (indicating a higher unconditional likelihood of problems when the fragility indicators have a high value) would support the use of (-*DD*) or *S* as early indicators of bank fragility (Proposition 1).

We created a dummy variable (DSUPP), equalling one when the Fitch/IBCA "support rating" is 1 or 2 in order to control for the government's perceived willingness to absorb default losses and to test for whether this dilutes the power of the market indicators. To this end, we interacted this variable with the market indicators. A significant and negative coefficient of (DSUPP**S*) and insignificant coefficient of (DSUPP*(-*DD*)) would support Proposition 3. Since we use several observations for the same bank in case the bank does not "default" during our sample period, our observations are not independent within banks, while they are independent across banks. Therefore, we adjusted the standard errors using the generalised method based on Huber (1967).

Our second model is a *Cox proportional hazard model* of the form:

$$h(t, DI, X) = h_0(t) e^{\beta_1 DI + \mathrm{B}_2 X} \tag{8}$$

where $h(t,DI,X)$ represents the proportional hazard function, $h_0(t)$ the baseline hazard, and X some control variables (see below). Again, we calculated robust standard errors, as we had multiple observations per bank and used Lin and Wei's (1989) adjustment to allow for correlation of the residuals within banks. The model parameters were estimated by maximising the partial log-likelihood function

$$\ln L = \sum_{j=1}^{D} \left\{ \sum_{r \in D_j} \left(\beta_1 DI_r + \mathrm{B}_2 X_r \right) - d_j \ln \left[\sum_{i \in R_j} \exp(\beta_1 DI_i + \mathrm{B}_2 X_i) \right] \right\} \tag{9}$$

where j indexes the ordered failure times $t(j)$ (j=1,2,…D), D_j is the set of d_j observations that "default" at $t(j)$ and R_j is the set of observations that are at risk at time $t(j)$. The model allows for censoring in the sense that, clearly, not all banks "default" during the sample period.[19]

The two models provide a robustness check whether equity and bond market indicators have signalling property as regards bank "defaults". In addition, they also provide insights into two distinct questions: the logit-model permits a test of the unconditional predictive power of the indicators with different lead times, whereas the proportional hazard model yields estimates of the impact of the market indicators on the *conditional probability of "defaulting".* The latter means that we obtain "default" probabilities, conditional on surviving to a certain point in time and facing a certain (-*DD*) or *S* in the previous period.

5.B Logit estimation results

Table 6A reports the results from estimating logit-models with different time leads. An increased (-*DD*) value tends to predict a greater likelihood of financial trouble. The respective coefficient is significant at the 10% level for the six-, 12- and 18-month leads. Hence, we find support for Proposition 1: (-*DD*) appears to have predictive properties of an increased (unconditional) likelihood of bank problems up to 18 months in advance. The coefficient ceases to be significant more than 18 months ahead of the event. However, we found the insignificance of the coefficient of the three-month lead somewhat puzzling. We suspect that the reason is increased noise in the -*DD* measure closer to the default, as evidenced by the higher standard error for the three- than the six-month leads. It may be the case that many eventually downgraded banks exhibit a lowering in the equity volatility just before the downgrading, which causes the derived asset volatility measure to decrease as well, reducing the (-*DD*) value.

Turning back to Table 6A, we find that the coefficient of DSUPP*($-DD_{t-x}$), measuring the impact of the safety net, is never statistically significant. Moreover, the hypothesis that the coefficient of ($-DD_{t-x}$) is zero for the banks with a strong expectation of government support is rejected for all lead times, except for x=24. The safety net does not appear to be important for the predictive power of the distance-to-default as an indicator of bank fragility.

[19] For more details on estimating hazard models see Kalbfleisch and Prentice (1980).

The results for the bond spreads, *S*, strongly support Proposition 1 as well (see Table 6B). The coefficients for lead times of up to 18 months are significant at least at the 5% level. The results also highlight that it is important to control for the expectation of public support in the case of spreads. The coefficient of the interacted term (DSUPP*S_{t-x}) is significant and negative, and a joint hypothesis test reveals that the coefficient on the spread is zero for the banks with a high (a rating of 1 or 2) expectation of public support. This finding is in contrast to the results using -*DD* as an indicator of bank fragility.

A convenient way to summarise the results of the logit models just described is given in Chart 3. The chart presents the coefficients from Tables 6A and 6B, normalised, such that the maximum effect is equal to one. It reveals that the maximum predictive power of spreads occurs quite shortly before default, around six to 12 months before. In contrast, *DD* has relatively little predictive power close to the event, but instead reaches its maximum no less than 18 months ahead of the default. These patterns correspond closely to the theoretical predictions of the option pricing framework discussed in Section II.

The results of discrete choice models may be quite sensitive to the underlying distributional assumptions, in particular in cases where the distribution of the dependent variable is as skewed as in this sample. Only 4% of the bond sample and 3% of the stock sample were "defaulting" observations. As a simple robustness check, we estimated the corresponding Probit-models and found essentially unchanged results, both in terms of magnitude and significance.[20]

5.C Hazard estimation results

Tables 7 and 8 give the *hazard ratios* and corresponding P-values for a model without additional control variables for both (-*DD*) and *S*. Only (-*DD*) is significant (at the 5% level); both indicators have the expected positive signs. The hazard ratios, indicating a greater conditional likelihood of "default", are increasing in the values of the fragility indicators, which is consistent with the logit results.

The tables also show the results for a test of the proportional hazard assumption (ie the zero-slope test), which amounts to testing whether the null hypothesis of a constant log hazard function over time holds for the individual covariates as well as globally. For (-*DD*), this assumption is violated. Hence, we present in Table 9 results from an alternative model specification, in which we use a dummy variable of the following form

$$ddind = \begin{cases} 1 \; if \; (-DD) > -3.2 \\ 0 \; otherwise \end{cases} \tag{10}$$

where -3.2 represents the 25th percentile of the distribution of (-*DD*). Hence, in this specification, we investigate whether banks with "short" distances-to-default are more likely to fail compared to all other banks. We find that the indicator significantly (at the 1% level) increases the hazard of a bank "defaulting", as before, and the model is no longer rejected due to the violation of the proportional hazard assumption.

We also examined the weaker performance of *S* than -*DD* in the baseline specification (as given in Tables 7 and 8). In the logit-model, we found that two factors significantly affect the predictive power of the spread: the presence of a safety net and whether or not the bank resides in the United Kingdom. Table 10 shows that the coefficient of the spread significantly improves when controlling for the United Kingdom by means of a dummy variable. *S* now is significant at the 1% level. In addition, the dummy for the United Kingdom is significant at the 5% level: higher spreads in the United Kingdom are associated with a significantly lower hazard ratio, ie a significantly lower likelihood of failure. For -*DD* the inclusion of the safety net dummy or the UK dummy do not materially affect the results, as in the logit specification, and are not reported here. Further, the logit results suggested that for banks which are likely to benefit from public support in case of trouble, the predictive power of bond spreads is reduced to zero. This finding is confirmed in Table 11.

[20] The results are available from the authors upon request.

The most convenient way to interpret the results is to consider the Nelson-Aalen *survivor functions*, which are depicted in Chart 4. The cumulative hazard functions display the probability of survival, given that the bank survived to period *t* and had a fragility indicator of a certain level. For convenience of presentation, we split the sample into those banks that have a default indicator in the top 25th percentile and all other banks. We can then test whether the survivor functions are significantly different and read the difference in the "default" probability at each point in time, given that the bank survived to that point. Using a log-rank test for both the distance-to-default and the spread, we can reject the equality of the survivor functions for the two groups at the 5% level. Excluding UK banks (the second part of the lower panel in Chart 4), we can reject equality at any significance level. Note that comparing the survivor functions with and without UK banks, excluding the UK banks results in a downward shift of both curves. Hence, excluding UK banks, all banks with a high spread (greater than 98 basis points) fail during the sample period. Only UK banks survive the entire sample period with a high spread. In this paper we will not explore this issue further. We only conclude that a UK spread puzzle remains, which we cannot explain.[21]

Even more interesting, we can immediately read off the difference in the survivor probability, given that a bank has remained in one or the other group. For (-*DD*), we find no difference in the hazard even after two years (24 months). Differences only arise subsequently: after 36 months, a bank which had a (-*DD*) > -3.2 for that period of time has a failure probability that is 20 percentage points higher relative to a bank that was consistently in the control group. This is consistent with the findings in the logit-model: (-*DD*) is found to be an indicator which has better leading properties for events further in the future. In contrast, spreads react only relatively shortly before default. Given survival, spreads essentially lose all their discriminating power after one year. The results also highlight that the prevalence of indicators matters, which suggests that the use of hazard models adds new insights relative to standard logit-models. Logit-models are unable to yield predictions which are conditional on default indicators having prevailed for periods of time.

Hence, in line with Proposition 2, the spread reacts more closely to the "default" point than (-*DD*). Put differently, banks may "survive" substantially longer with a short distance-to-default, but the likelihood of quite immediate problems is very high if they exhibit a high spread (in our definition of 100 basis points or above). As we show in the earlier part of this paper, the strong reaction of the spreads only close to the default point is explained by the non-linear payoff profile of subordinated debt holders.

Finally we present log-rank tests of the equality of survivor functions for those banks with an implicit safety net ("support rating" of 1 or 2) in Table 12. We find that the distance-to-default has more predictive power for banks which are likely to benefit from governmental support, and little predictive power for those that do not.[22] More importantly, Table 12 shows the importance of UK banks, as well as the safety net, for the predictive qualities of bond spreads. With UK banks included, we find only weak discriminating power of spreads even for banks which are not likely to receive public support in case of problems. Without UK banks, however, we find that spreads perform significantly better in case of banks with little or no public support, confirming our earlier results and Proposition 3.

6. Robustness and extensions

As an extension, it is interesting to examine whether the market indicators contain information which is not already summarised in ratings. To this end, we controlled for the "individual rating" at the time the market indicators were observed. The results given in Table 13A for the (-*DD*) measure are fairly similar to those reported in Table 7A, although the significance of the (-*DD*) indicator is somewhat reduced. Overall, they suggest that the (-*DD*) indicator adds to the information obtainable from (Fitch/IBCA) ratings and the more the longer the time leads. The results are even stronger for the

---

[21] Gropp and Olters (2001) attempt an explanation using a political economy model. They argue that as the United Kingdom has a market-based financial system as opposed to continental Europe, which is bank-based, a political majority to bail out banks is more difficult to obtain in the United Kingdom. Investors, therefore, want to be compensated for this additional default risk and require higher spreads.

[22] This somewhat puzzling finding, which we would not want to oversell, may in fact have to do with sample composition.

spreads (see Table 13B). We conclude that both of the indicators analysed in this paper appear to contain additional information from ratings, at least in terms of their ability to predict bank "failures".

This also addresses the specific issue raised by our definition of "failures". Namely, there is the possibility that we would be using market indicators to predict rating downgradings, which could be based on the same set of information of the probability of default. However, as we find that the market indicators contain additional information compared to prevailing ratings, this concern does not seem to be warranted. However, even if the ratings contained completely similar information to our market indicators, we would find support in our standard logit and hazard models for using market indicators: high-frequency market data have leading properties over discrete bank problem events reflected in their individual ratings.

We also checked whether the distance-to-default measure performs better in terms of its (unconditional) predictive property than simpler equity-based indicators. First, we estimated the logit-models using the equity volatility as the fragility indicator. However, it turned out to be a significantly weaker predictor of "default". The coefficients of $\sigma_{E,t-x}$ were never statistically significant. The composite nature of the (*-DD*) apparently improves predictive performance and reduces noise. We found similar results for a simple leverage measure (V_E/V_L).[23]

Next, we wanted to explore whether our market indicators add information to that already available from banks' balance sheets. Conceptually, this is obvious: market-based indicators should fully reflect past balance sheet information as well as forward-looking expectations about the prospects of the bank. First note that we were unable to estimate the hazard model with balance sheet variables, as they are not available at a monthly frequency. Hence, we estimated logit models only.[24] Clearly, the choice of which balance sheet variables to use is arbitrary. We followed the previous literature (see eg Sironi (2000), Flannery and Sorescu (1996)) and considered a set of balance sheet indicators emulating the categories of CAMEL ratings (Capital adequacy, Asset quality, Management, Earnings, Liquidity).[25] Then, we calculated a composite score based on the bank's position in each year's distribution for every indicator.[26] In this way, we were able to consider the correlation between the different indicators, ie whether a bank is "strong" or "weak" by more than one indicator. We re-estimated the model containing only the market indicators, in order to ensure comparability given the reduced sample size. Second, we estimated a model only with balance sheet indicators, and third, a model combining market and balance sheet indicators. Here, we only report results for the 12 months time lead.

Results for the distance-to-default indicator (Table 14A) show that it adds some information to that already available from balance sheet data. In the model combining the distance-to-default and the balance sheet indicators, the distance-to-default indicator is significant (at the 5% level), and the model fit, as measured by the pseudo-R^2, increases from 0.20 to 0.24 over the one containing only balance sheet variables.[27] In addition, the significance of the distance-to-default indicator improves in the combined model, when compared with the model with only the distance-to-default indicator. This suggests that the distance-to-default indicator provides additional information to that of balance sheet

[23] The results are available from the authors upon request.

[24] Even for the logit-models we were faced with a significant reduction in sample size. Since balance sheet data are available only on an annual basis, we used only end-year market indicators, rather than utilising all available monthly observations with the same horizon as in the earlier specifications.

[25] In order to maintain a sufficient sample size in the set of failed banks, we had to consider only four out of five indicators. Hence, the liquidity indicator was taken out of the analysis.

[26] The composite score is calculated in the following way:

- we considered the percentile ranking of the bank in each year distribution for every indicator;
- we divided the ranking distributions into four quartiles, and assigned a score varying from 0 (best) to 3 (worst) to the position of the bank in the rankings;
- we obtained the composite score by simply summing up the scores for each indicator, yielding a variable ranging from zero (a bank in good condition with all indicators) to 15 (a bank in bad condition with all indicators).

The FDIC uses a broadly similar approach for its CAMEL model (see FDIC (1994)).

[27] The likelihood-ratio test rejects the hypothesis of no significance of the distance-to-default indicator.

variables, but it does not replace the balance sheet indicators. In other words, the distance-to-default and the balance sheet indicators are both useful for the monitoring of banks and play a complementary role.

Empirical estimates from the same exercise for the spreads indicator are presented in Table 14B. They suggest that spreads also add some information to that already available from balance sheet data, although the evidence is weaker. As before, the model combining the spreads and the balance sheet indicators has a slightly better fit (in terms of pseudo-R^2) over the one containing only balance sheet variables. However, by itself spreads are not significant, even for the banks that are not expected to be supported. Our interpretation is that spreads are highly correlated with the balance sheet information and, hence, to some extent simply appear to reflect backward-looking information, rather than information about the future performance of the bank.

Clearly, tests of the sort presented here have the drawback that they can always be criticised on the basis of omitted variable bias, ie that some other balance sheet indicator may be more relevant. In order to alleviate this criticism, we have taken care to use variables in line with the previous literature and have also tried to emulate a CAMEL approach, which is used by many regulators. The most important result based on this exercise may be that we find some complementarity between market and balance sheet indicators.

Finally, we wondered whether the two market indicators might not provide complementary information to each other. In particular, in the previous section, we demonstrated that the two indicators have very different predictive properties through time. Spreads react late, but lose predictive power further away from the event. The distance-to-default is not a very strong indicator close to default, but has strong leading properties around two years out.[28] Table 15 gives the results from a model with both indicators included simultaneously. We find that both variables are significant at least at the 5% level.

Based on this finding, we can ask two further questions. One, which combination of spread and distance-to-default gives us the most discriminatory power? And, second, is this an improvement over using one or the other indicator alone? In Chart 5 we attempt to shed some light on both questions. In the top panel we have given the survivor functions for banks which are above the median in at least one of the indicators and are in the top 75th percentile in the other versus all other banks. We find that the survivor functions are not significantly different from one another. In the bottom panel, we have plotted the survivor functions for banks that are above the median in both indicators versus all other banks. Now the survivor functions are statistically significantly different at the 5% level. It turns out that the “above median in both indicators” criterion gives us maximum discriminatory power.

Further, comparing the lower panel of Chart 5 to Chart 4, we find that the combination of both indicators provides us with better discriminatory power than either indicator alone. In comparison to the distance-to-default (top panel of Chart 4), we have significantly more discriminatory power closer to the default, which we would attribute to the addition of information contained in spreads. Looking at the lower panel of Chart 4, we find that the addition of information contained in the distance-to-default to spreads reduces type one error dramatically. We are missing significantly fewer defaults when using a combination of both indicators, which is evident from the much flatter curvature of the top line in Chart 5 compared to Chart 4 (lower panel). Overall, we conclude that the market indicators appear to provide useful information not only relative to balance sheet information and ratings, but also to each other.

7. Conclusion

In this paper, we present evidence in favour of using market price-based measures as early indicators of bank fragility. We first argue that sensible indicators of bank fragility should be both *complete*, in that they should reflect all potential sources of default risk, and *unbiased*, in that they should reflect these risks correctly. We then demonstrated that it is possible to derive indicators satisfying both

[28] The simple correlation coefficient between the spread and the distance-to-default is -0.034, in itself suggesting that the two indicators measure different things. Note also that the sample sizes in Table 15 are reduced somewhat relative to earlier models, as they contain only those observations with both bond and stock market data during the same period.

qualities from equity as well as from debt prices. We find that the negative distance-to-default is a preferred indicator over other equity price-based indicators, since it is unbiased in the sense that it will correctly flag an increase in asset volatility. The standard bond spread also satisfies our conditions. We show that both indicators perform quite well as leading indicators for bank fragility in a sample of EU banks. Due to the absence of banks declaring formal bankruptcy, we measured a bank “failure” as a downgrading in the Fitch/IBCA “financial strength rating” to C or below. We argue that this measure of bank fragility may be sensible as in virtually all cases there was government support or a major restructuring in the wake of the event.

Specifically, we estimate both a logit and a proportional hazard model. The logit-model estimates suggest that both bond spreads and distances-to-default have predictive power up to 18 months in advance of the event. This was corroborated by the estimates obtained using the hazard model. The results, however, also point towards significant differences between the two indicators. One, the negative distance-to-default exhibits poor predictive power close to the event. Similarly, our results show that banks might “survive” relatively long periods of time with short distances-to-default. In contrast, bond spreads have a tendency to only react close to the default, ie they only react when the situation of the bank has already become quite desperate. This implies that banks tend to survive only relatively short periods of time with high spreads. These findings are consistent with the theoretical properties of the respective indicators, which we analyse in an option pricing framework. Second, we present some evidence that bond spreads predict financial difficulties only in the case of (smaller) banks which do not benefit from a stronger expectation of a public bailout. We measured this expectation in terms of the “support rating”, indicating the likelihood of public intervention. The equity-based distance-to-default measure was not found sensitive to the expectation of an implicit safety net, which is in line with our priors. Finally, we demonstrate that, given the different properties of the bond and equity-based indicators, they also provide complementary information to each other, in particular with respect to reducing type I errors.

We interpret our findings in a way to suggest that supervisors (and possibly the literature) may want to devote more attention to the equity market when considering the use of the information embedded in the market prices of the securities issued by banks. Equity market data could provide supervisors with useful complementary information. The information may be complementary with respect both to balance sheet data and to bond-based market indicators.

As an important caveat, it should be stressed that there might be considerable practical difficulties in using either of the indicators proposed in this paper. For example, the distance-to-default measure, apart from its relative computational complexity, may be sensitive to shifts in derived asset volatility. This, in turn, may be due to irregularities in the equity trading in the period closer to default. Further, the measure is quite sensitive to the measure of equity volatility used and distributional assumptions about equity returns. Similarly, the calculation of bond spreads may be difficult in practice, because of relatively illiquid bond markets, resulting in noisy price data for bank bonds and the lack of reliable risk-free benchmarks (especially in smaller countries).

Tables

Table 1

Composition of banks by country and availability of equity and bond data

	Equity	Bond		Equity	Bond
Belgium	4	1	Italy	20	7
Denmark	2		Netherlands	3	4
Germany	10	16	Austria	3	2
Greece	3		Portugal	4	1
Spain	7	2	Finland	1	2
France	8	9	Sweden	3	3
Ireland	4	2	United Kingdom	11	10
			Total	**84**	**59**

Table 2

Downgrading events (to "individual rating" C or below) in the sample

Bank	Downgrading	Support / restructuring / other	Timing
A. Cases of public support			
Banco Español de Credito**	Jun 93	Public financial support	Dec 93
Banco di Napoli**	Jan 95	Public capital injection	Early 96
Banca Nazionale del Lavoro	Jun 97	Public capital support in the form of a transfer of Artigiancassa	During 96
Bankgesellschaft Berlin	Jun 99	Recapitalisation (partly government-owned bank)	During 01
CPR	Nov 98	Support from the parent group (CA)	End-98
Credit Lyonnais*	Jun 94	Public financial support	Spring 95
Credit Foncier de France	First rating (D) Apr 00	Public financial support	Apr 96
Erste Bank der Oesterreichischen Sparkassen	Feb 00	Capital injection (from the savings banks' system)	Oct 00
Okobank	Oct 94	Public capital injection	Oct 93-end-95
Skandinaviska Enskilda Banken	Jul 92	Government guarantee	Dec 92
Svenska Handelsbanken	Dec 92	Government guarantee	Dec 92

B. Cases of substantial restructuring			
Banca Popolare di Novara**	Oct 95	Major restructuring, eg new management	During 96
Bank Austria	Jun 96	Absorbed by West-Deutsche LB	May 97
Banque Natexis	Nov 96	Merger (Credit National and Banque Federal de BP)	Jan 97
Banque Worms	Nov 99	Sold to Deutsche Bank	Oct 00
CIC Group	Aug 95	Fully privatised	During 96
Commercial Bank of Greece	Dec 98	Sale of significant parts of operations (Ionian and Popular Bank)	Early 99
Entenial	Mar 99	Merger with Banque La Hénin-Epargne Crédit (BLH).	
Creditanstalt	Jan 97	Takeover by Bank Austria	Jan 97
C. Other cases			
Banca Commerciale Italiana	Jun 00	Weak performance and asset quality	
Banca di Roma	Nov 96	Depressed profitability and asset quality eg due to several acquisitions	
Banca Popolare di Intra**	Feb 01	Weak performance and asset quality	
Banca Popolare di Lodi	Jun 00	Weak performance and asset quality	
Banca Popolare di Milano**	Nov 95	Weak performance and asset quality	
Banca Popolare di Sondrio**	Mar 00	Weak performance and asset quality	
Banco Zaragozano**	Mar 95	Weak performance and asset quality	
Bayerische Landesbank*	Dec 99	Weak capital adequacy and asset quality	
Credito Valtellinese	Feb 01	Weak performance and asset quality	
Deutsche Genossenschaftsbank*	Nov 00	Weak performance and asset quality	
HSBC Bank*	May 91	Weak performance and asset quality	
Standard Chartered*	Jun 90	Weak performance and asset quality	
Westdeutsche Landesbank*	Nov 98	Exposures to Russia, weak capitalisation	

Source: Fitch/IBCA. * Only in the bond sample. ** Only in the equity sample.

Table 3

Definition of variables

Variable	Definition
Market value of equity (V_E)	Monthly average equity market capitalisation (millions of euros)
Equity volatility (σ_E)	6-month moving average (backwards) of daily absolute equity returns (%)
Book value of debt liabilities (D)	Total debt liabilities (interpolated monthly observations) (millions of euros)
Market value of assets (V)	Derived (equations (2)) monthly average of the total asset value (millions of euros)
Volatility of assets (σ)	Derived (equations (2)) monthly estimate of the asset value volatility (%)
Negative of the distance-to-default (-*DD*)	Monthly average (-*DD*) calculated from V_A, σ_A, and V_L(equation (3))
Spread (S)	Calculated monthly average subordinated debt spread of the yield to maturity over the risk-free yield to maturity
Dummy indicating expected public support (DSUPP)	Dummy variable equalling one if Fitch/IBCA support rating 1 or 2 (zero otherwise)
Status variable (STATUS)	Binary variable equalling one if a bank experiences a downgrading in Fitch/IBCA “individual rating” to C or below (zero otherwise)

Table 4

Descriptive statistics

Variable	*t-x*	Nobs	Mean	Std dev	Min	Max[1]
Market value of equity (V_E) (millions of euros)	x = 3 months	1043	10,212	17,452	13.64	191,638
	x = 6 months	1043	10,047	17,305	11.80	229,167
	x = 12 months	1040	9,043	15,597	13.79	183,195
	x = 18 months	1039	8,363	14,509	13.64	129,555
	x = 24 months	1036	7,377	13,226	11.84	104,839
Equity volatility (σ_E)	x = 3 months	1043	0.27	0.14	0.01	2.01
	x = 6 months	1043	0.27	0.14	0.01	2.01
	x = 12 months	1040	0.28	0.14	0.01	0.71
	x = 18 months	1039	0.27	0.15	0.01	2.06
	x = 24 months	1036	0.25	0.14	0.01	2.06
Book value of debt liabilities (*D*) (millions of euros)	x = 3 months	1043	94,862	117,375	464.95	715,825
	x = 6 months	1043	91,921	113,277	441.31	688,596
	x = 12 months	1040	86,908	106,286	397.59	636,515
	x = 18 months	1039	82,799	100,645	358.20	556,785
	x = 24 months	1036	79,308	95,969	305.34	490,866
Market value of assets (*V*) (millions of euros)	x = 3 months	1043	99,500	120,350	568.99	735,885
	x = 6 months	1043	96,617	116,403	519.16	710,957
	x = 12 months	1040	90,818	108,557	484.66	652,365
	x = 18 months	1039	85,963	102,492	365.65	569,511
	x = 24 months	1036	81,478	96,825	312.37	499,827
Volatility of assets (σ)	x = 3 months	1043	0.04	0.05	0.00	0.65
	x = 6 months	1043	0.04	0.05	0.00	0.65
	x = 12 months	1040	0.04	0.04	0.00	0.28
	x = 18 months	1039	0.04	0.05	0.00	0.73
	x = 24 months	1036	0.03	0.04	0.00	0.73
Negative of the distance-to-default (*-DD*)	x = 3 months	1043	– 5.64	6.00	– 87.71	0.99
	x = 6 months	1043	– 5.60	5.71	– 91.12	0.99
	x = 12 months	1040	– 5.28	5.01	– 71.71	– 1.20
	x = 18 months	1039	– 5.62	6.57	– 133.89	1.05
	x = 24 months	1036	– 5.90	6.46	– 130.44	1.05
Spread (*S*) (%)	x = 3 months	478	0.89	1.14	– 0.49	6.02
	x = 6 months	474	0.87	1.15	– 0.40	6.08
	x = 12 months	457	0.79	1.04	– 0.27	6.07
	x = 18 months	432	0.75	1.04	– 0.82	6.32
	x = 24 months	407	0.70	1.06	– 0.62	6.23

[1] The large max values for equity and asset volatility are due to Banca Popolare dell'Emilia Romagna, which had very high volatility levels from December 1996 to May 1997. This observation was not found to affect the econometric results.

Table 5

Ability of (-*DD*) and *S* to distinguish weaker banks: mean value tests, all banks

	Status	Nobs	Mean	Std error	Difference[1]	Difference < 0[2]
Equity			($-DD_{t-x}$)			
x = 3 months	0	1018	– 5.68	0.19	– 1.58	– 3.490 ***
	1	25	– 4.10	0.41		
x = 6 months	0	1018	– 5.64	0.18	– 1.79	– 5.335 ***
	1	25	– 3.85	0.28		
x = 12 months	0	1018	– 5.31	0.16	– 1.62	– 4.887 ***
	1	22	– 3.69	0.29		
x = 18 months	0	1018	– 5.66	0.21	– 1.93	– 5.181 ***
	1	21	– 3.72	0.31		
x = 24 months	0	1018	– 5.93	0.20	– 1.55	– 2.823 ***
	1	18	– 4.38	0.51		
Bond			S_{t-x}			
x = 3 months	0	457	0.88	0.05	– 0.19	– 0.68
	1	21	1.07	0.27		
x = 6 months	0	454	0.86	0.05	– 0.18	– 0.55
	1	20	1.04	0.32		
x = 12 months	0	438	0.79	0.05	– 0.10	– 0.37
	1	19	0.89	0.26		
x = 18 months	0	417	0.74	0.05	– 0.12	– 0.43
	1	15	0.86	0.27		
x = 24 months	0	393	0.70	0.05	– 0.03	– 0.13
	1	14	0.73	0.26		
Bond[3]			S_{t-x}			
x = 3 months	0	78	0.24	0.02	– 0.55	– 1.997 **
	1	5	0.79	0.25		
x = 6 months	0	72	0.22	0.02	– 0.36	– 2.90 **
	1	5	0.58	0.12		
x = 12 months	0	67	0.22	0.02	– 0.38	– 1.556 *
	1	4	0.60	0.25		

Note: Two sub-sample *t*-tests (unequal variances) are reported for the difference in mean values of ($-DD_{t-x}$) and S_{t-x} in the sub-samples of downgraded (SATUS=1) and non-downgraded banks (STATUS=0). *, **, *** indicate statistical significance at the 10%, 5% and 1% levels, respectively.

[1] Mean (STATUS=0) – Mean (SATUS=1). [2] *t*-statistics for testing the hypothesis that difference is negative. [3] Banks with low public support expectation and excluding UK banks.

Table 6A

Predictive performance of the distance-to-default indicator: logit-estimations, all banks

x = 3 months				
Explanatory variable	**Coefficient**	**Robust std error[2]**	***z***	***P*>| *z* |**
Constant	– 2.803 ***	0.454	– 6.170	0.000
$(-DD_{t-3})$	0.113	0.091	1.240	0.216
DSUPP*$(-DD_{t-3})$	0.158	0.105	1.510	0.130
Number of observations	1043		Log likelihood	– 114.35
F-test[1]	5.22 **		Pseudo R^2	0.0307
x = 6 months				
Explanatory variable	**Coefficient**	**Robust std error[2]**	***z***	***P*>| *z* |**
Constant	– 2.620 ***	0.440	– 5.950	0.000
$(-DD_{t-6})$	0.182 *	0.096	1.890	0.058
DSUPP*$(-DD_{t-6})$	0.112	0.109	1.030	0.302
Number of observations	1043		Log likelihood	– 114.04
F-test[1]	6.44 **		Pseudo R^2	0.0333
x =12 months				
Explanatory variable	**Coefficient**	**Robust std error[2]**	***z***	***P*>| *z* |**
Constant	– 2.889 ***	0.451	– 6.400	0.000
$(-DD_{t-12})$	0.212 **	0.105	2.030	0.043
DSUPP*$(-DD_{t-12})$	0.018	0.117	0.150	0.880
Number of observations	1040		Log likelihood	– 103.96
F-test[1]	3.78 **		Pseudo R^2	0.0247
x = 18 months				
Explanatory variable	**Coefficient**	**Robust std error[2]**	***z***	***P*>| *z* |**
Constant	– 2.686 ***	0.541	– 4.960	0.000
$(-DD_{t-18})$	0.287 *	0.149	1.920	0.054
DSUPP*$(-DD_{t-18})$	– 0.014	0.126	– 0.110	0.913
Number of observations	1039		Log likelihood	–102.742
F-test[1]	4.29 **		Pseudo R^2	0.0322
x = 24 months				
Explanatory variable	**Coefficient**	**Robust std error[2]**	***z***	***P*>| *z* |**
Constant	– 3.301 ***	0.594	– 5.560	0.000
$(-DD_{t-24})$	0.171	0.130	1.320	0.188
DSUPP*$(-DD_{t-24})$	– 0.034	0.113	– 0.300	0.761
Number of observations	1036		Log likelihood	– 89.315
F-test[1]	1.11		Pseudo R^2	0.0163

Note: All models are estimated using the binary variable STATUS as the dependent variable. *, **, *** indicate statistical significance at the 10%, 5% and 1% levels, respectively.

[1] F-test for the hypothesis that the sum of the coefficients of $(-DD_{t-x})$ and DSUPP*$(-DD_{t-x})$ is zero (ie that the coefficient of $(-DD_{t-x})$ is zero for banks with a greater expectation of public support). $\chi 2$ values reported. [2] Standard errors adjusted.

Table 6B

Predictive performance of the spread indicator: logit-estimations, all banks

Explanatory variable	Coefficient	Robust std error[2]	z	$P>\|z\|$
x = 3 months				
Constant	– 3.361***	0.387	– 8.680	0.000
(S_{t-3})	2.838***	1.120	2.530	0.010
DSUPP*(S_{t-3})	– 2.546**	1.100	– 2.310	0.021
Number of observations	364		Log likelihood	– 69.854
F-test[1]	1.41		Pseudo R^2	0.064
x = 6 months				
Constant	– 3.497***	0.421	– 8.300	0.000
(S_{t-6})	4.073***	1.555	2.620	0.009
DSUPP*(S_{t-6})	– 3.745***	1.513	– 2.480	0.010
Number of observations	361		Log likelihood	– 66.464
F-test[1]	1.86		Pseudo R^2	0.071
x = 12 months				
Constant	– 3.416***	0.402	– 8.500	0.000
(S_{t-12})	3.186**	1.311	2.430	0.015
DSUPP*(S_{t-12})	– 2.781**	1.286	– 2.160	0.031
Number of observations	348		Log likelihood	– 64.379
F-test[1]	2.09		Pseudo R^2	0.052
x = 18 months				
Constant	– 3.528***	0.437	– 8.070	0.000
(S_{t-18})	2.706**	1.112	2.430	0.015
DSUPP*(S_{t-18})	– 2.402**	1.088	– 2.210	0.027
Number of observations	328		Log likelihood	– 52.302
F-test[1]	0.67		Pseudo R^2	0.044
x = 24 months				
Constant	– 3.433	0.470	– 7.300	0.000
(S_{t-24})	2.305	2.280	1.010	0.312
DSUPP*(S_{t-24})	– 2.062	2.194	– 0.940	0.347
Number of observations	310		Log likelihood	– 50.013
F-test[1]	0.29		Pseudo R^2	0.015

Note: All models are estimated using the binary variable STATUS as the dependent variable and excluding UK banks. *, **, *** indicate statistical significance at the 10%, 5% and 1% levels, respectively.

[1] F-test for the hypothesis that the sum of the coefficients of (S_{t-x}) and DSUPP*(S_{t-x}) is zero (ie the coefficient of (S_{t-x}) is zero for banks with a greater expectation of public support). $\chi 2$ values reported. [2] Standard errors adjusted.

Table 7

Performance of the distance-to-default indicator: proportional hazard estimation, all banks

Explanatory variable	Hazard ratio	Robust std error	z	$P>\|z\|$
(-*DD*)	0.728**	0.115	2.02	0.04
Number of subjects	84	Time at risk		5365
Number of failures	25	Starting log likelihood		– 100.49
Number of observations	5365	Final log likelihood		– 96.71
Wald χ^2	4.08**	Zero-slope test		7.66***

Note: Estimated using Cox regression. Log-likelihood given in the text. Standard errors are corrected for clustering using Wei and Lin's (1989) method. *, **, *** indicate statistical significance at the 10%, 5% and 1% levels, respectively.

Table 8

Performance of the bond spread: proportional hazard estimation, all banks

Explanatory variable	Hazard ratio	Robust std error	z	$P>\|z\|$
S	1.00	0.002	0.75	0.455
Number of subjects	59	Time at risk		3604
Number of failures	19	Starting log likelihood		– 69.76
Number of observations	3604	Final log likelihood		– 69.54
Wald χ^2	0.56	Zero-slope test		0.40

Note: Estimated using Cox regression. Log-likelihood given in the text. Standard errors are corrected for clustering using Wei and Lin's (1989) method. *,**, *** indicate statistical significance at the 10%, 5% and 1% levels, respectively.

Table 9

Performance of the distance-to-default indicator: proportional hazard estimation using a dummy variable, all banks

Explanatory variable	Hazard ratio	Robust std error	z	$P>\|z\|$
Dummy for (-*DD*) >-3.2	2.69***	1.034	2.57	0.01
Number of subjects	84	Time at risk		5365
Number of failures	25	Starting log likelihood		– 100.49
Number of observations	5365	Final log likelihood		– 97.86
Wald χ^2	6.62***	Zero-slope test		1.52

Note: Estimated using Cox regression. Log-likelihood given in the text. Standard errors are corrected for clustering using Wei and Lin's (1989) method. *,**, *** indicate statistical significance at the 10%, 5% and 1% levels, respectively.

Table 10

Performance of the bond spread:
proportional hazard estimation controlling for the UK, all banks

Explanatory variable	Hazard ratio	Robust std error	*z*	*P*>| *z* |
S	1.01***	0.002	2.74	0.006
Dummy for UK	– 0.065**	0.080	– 2.25	0.025
Number of subjects	59	Time at risk		3604
Number of failures	19	Starting log likelihood		– 69.76
Number of observations	3604	Final log likelihood		– 65.18
Wald χ^2	8.76***	Zero-slope test (global test)		1.86

Note: Estimated using Cox regression. Log-likelihood given in the text. Standard errors are corrected for clustering using Wei and Lin's (1989) method. *,**, *** indicate statistical significance at the 10%, 5% and 1% levels, respectively.

Table 11

Performance of the bond spread:
proportional hazard estimation controlling the level of support, UK banks excluded

Explanatory variable	Hazard ratio	Robust std error	*z*	*P*>| *z* |
S	1.02***	0.005	3.79	0.000
Dummy "high support"**S*	– 0.99***	0.005	– 2.71	0.007
Number of subjects	49	Time at risk		2720
Number of failures	18	Starting log likelihood		– 61.51
Number of observations	2720	Final log likelihood		– 57.07
Wald χ^2	16.38***	Zero-slope test (global test)		2.58

Note: Estimated using Cox regression. Log-likelihood given in the text. Standard errors are corrected for clustering using Wei and Lin's (1989) method. *,**, *** indicate statistical significance at the 10%, 5% and 1% levels, respectively.

Table 12

The role of the safety net and the UK location: log-rank tests for equality of survivor functions, all banks

	χ^2	$P > \chi^2$
(-*DD*)		
Dummy “high support” equal to 1	4.94**	0.03
Dummy “high support” equal to 0	0.90	0.34
S		
Dummy “high support” equal to 1	1.95	0.16
Dummy “high support” equal to 0	3.30*	0.07
***S*; excluding UK banks**		
Dummy “high support” equal to 1	7.81***	0.005
Dummy “high support” equal to 0	30.19***	0.000

Note: Estimated using the Cox regression in Tables 9 and 11. *,**, *** indicate statistical significance at the 10%, 5% and 1% levels, respectively.

Table 13A

Predictive performance of the distance-to-default indicator: logit-estimations, controlling for the Fitch/IBCA individual rating before the event

x = 6 months

Explanatory variable	Coefficient	Robust std Error	*z*	*P*>| *z* |
Constant	– 3.888***	0.709	– 5.490	0.000
$(-DD_{t-6})$	0.186	0.117	1.590	0.112
DSUPP*$(-DD_{t-6})$	0.092	0.127	0.730	0.468
INDRAT_{t-6}	0.357**	0.168	2.120	0.034
Number of observations	959		Log likelihood	– 105.237
F-test	4.52**		Pseudo R^2	0.0916

x = 12 months

Explanatory variable	Coefficient	Robust std Error	*z*	*P*>| *z* |
Constant	– 3.954***	0.663	– 5.960	0.000
$(-DD_{t-12})$	0.208*	0.120	1.730	0.084
DSUPP*$(-DD_{t-12})$	0.022	0.136	0.160	0.873
INDRAT_{t-12}	0.321**	0.151	2.120	0.034
Number of observations	931		Log likelihood	– 96.997
F-test	3.22*		Pseudo R^2	0.0685

x = 18 months

Explanatory variable	Coefficient	Robust std Error	*z*	*P*>| *z* |
Constant	– 3.431***	0.754	– 4.550	0.000
$(-DD_{t-18})$	0.290*	0.163	1.780	0.075
DSUPP*$(-DD_{t-18})$	0.017	0.151	0.110	0.913
INDRAT_{t-18}	0.277*	0.150	1.850	0.064
Number of observations	909		Log likelihood	– 93.172
F-test	4.25**		Pseudo R^2	0.0669

Note: Logit-estimations are reported for the sample of downgraded and non-downgraded banks, controlling for the individual rating before the event. *, **, *** indicate statistical significance at the 10%, 5% and 1% levels, respectively.

See notes to Table 7A.

Table 13B

Predictive performance of the spread indicator: logit-estimations, controlling for the Fitch/IBCA individual rating before the event

x = 3 months

Explanatory variable	Coefficient	Robust std error	*z*	*P*>| *z* |
Constant	– 9.659**	3.954	– 2.440	0.015
($S_{t\text{-}3}$)	2.277***	0.797	2.860	0.004
DSUPP*($S_{t\text{-}3}$)	– 1.994***	0.747	– 2.670	0.008
$\text{INDRAT}_{t\text{-}6}$	1.610	1.015	1.590	0.113
Number of observations	305		Log likelihood	– 36.639
F-test[1]	1.48		Pseudo R^2	0.355

x = 6 months

Explanatory variable	Coefficient	Robust std error	*z*	*P*>| *z* |
Constant	– 8.366***	2.990	– 2.800	0.005
($S_{t\text{-}6}$)	3.364**	1.555	2.160	0.030
DSUPP*($S_{t\text{-}6}$)	– 3.068**	1.458	– 2.100	0.035
$\text{INDRAT}_{t\text{-}6}$	1.253	0.809	1.550	0.122
Number of observations	295		Log likelihood	– 36.458
F-test[1]	1.38		Pseudo R^2	0.316

x = 12 months

Explanatory variable	Coefficient	Robust std error	*z*	*P*>| *z* |
Constant	– 7.837***	2.874	– 2.730	0.006
($S_{t\text{-}12}$)	3.078***	1.169	2.630	0.008
DSUPP*($S_{t\text{-}12}$)	– 2.790***	1.092	– 2.560	0.010
$\text{INDRAT}_{t\text{-}12}$	1.158	0.810	1.430	0.153
Number of observations	283		Log likelihood	– 35.293
F-test[1]	0.62		Pseudo R^2	0.283

Note: Logit-estimations are reported for the sample of downgraded and non-downgraded banks, controlling for the individual rating before the event and excluding UK banks. *, **, *** indicate statistical significance at the 10%, 5% and 1% levels, respectively.

See notes to Table 7B.

Table 14A

Information content of the distance-to-default indicator: logit-estimations, all banks

Model with only the distance-to-default indicator
x = 12 months

Explanatory variable	Coefficient	Robust std error[2]	*Z*	*P*>| *z* |
Constant	– 1.790***	0.492	– 3.640	0.000
(DD_{t-x})	0.249**	0.121	2.070	0.039
DSUPP*(DD_{t-x})	0.005	0.119	0.040	0.970
Number of observations	408		Log likelihood	– 82.626
F-test[1]	3.97**		Pseudo R^2	0.035

Model with only balance sheet indicators
x = 12 months

Explanatory variable	Coefficient	Robust std error[2]	*Z*	*P*>| *z* |
Constant	– 7.105***	1.082	– 6.570	0.000
SCORE	0.574***	0.121	4.740	0.000
Number of observations	408		Log likelihood	– 68.588
			Pseudo R^2	0.199

Model with the distance-to-default indicator and balance sheet indicators
x = 12 months

Explanatory variable	Coefficient	Robust std error[2]	*Z*	*P*>| *z* |
Constant	– 6.232***	1.155	– 5.390	0.000
(DD_{t-x})	0.242**	0.110	2.200	0.028
DSUPP*(DD_{t-x})	– 0.044	0.127	– 0.340	0.732
SCORE	0.585***	0.125	4.670	0.000
Number of observations	408		Log likelihood	– 65.360
F-test[1]	3.03*		Pseudo R^2	0.238

Note: All models are estimated using the binary variable STATUS as the dependent variable. *, **, *** indicate statistical significance at the 10%, 5% and 1% levels, respectively. SCORE is a synthetic variable summarising the ranking of the bank with regard to four indicators representing respectively capital adequacy, asset quality, efficiency and profitability.

See notes to Table 7B.

Table 14B

Information content of the spreads indicator:
logit-estimations, all banks

Model with only the spreads indicator
x = 12 months

Explanatory variable	Coefficient	Robust std error[2]	*z*	*P*>| *z* |
Constant	– 2.451***	0.405	– 6.060	0.000
(S_{t-x})	2.999**	1.353	2.220	0.027
DSUPP*(S_{t-x})	– 2.575*	1.328	– 1.940	0.053
Number of observations	144		Log likelihood	– 49.388
F-test[1]	2.00		Pseudo-R^2	0.055

Model with only balance sheet indicators
x = 12 months

Explanatory variable	Coefficient	Robust std error[2]	*z*	*P*>| *z* |
Constant	– 6.272***	1.269	– 4.940	0.000
SCORE	0.548***	0.142	3.850	0.000
Number of observations	144		Log likelihood	– 40.260
			Pseudo-R^2	0.230

Model with the spreads indicator and balance sheet indicators
x = 12 months

Explanatory variable	Coefficient	Robust std error[2]	*z*	*P*>| *z* |
Constant	– 6.305***	1.233	– 5.110	0.000
(S_{t-x})	2.079	1.627	1.280	0.201
DSUPP*(S_{t-x})	– 1.662	1.600	– 1.040	0.299
SCORE	0.514***	0.138	3.730	0.000
Number of observations	144		Log likelihood	– 39.136
			Pseudo-R^2	0.251

Note: All models are estimated using the binary variable STATUS as the dependent variable. *, **, *** indicate statistical significance at the 10%, 5% and 1% levels, respectively. SCORE is a synthetic variable summarising the ranking of the bank with regard to four indicators representing respectively capital adequacy, asset quality, efficiency and profitability. The models exclude UK banks.

See notes to Table 7B.

Table 15

Performance of the distance-to-default and the bond spread: proportional hazard estimation, UK banks excluded

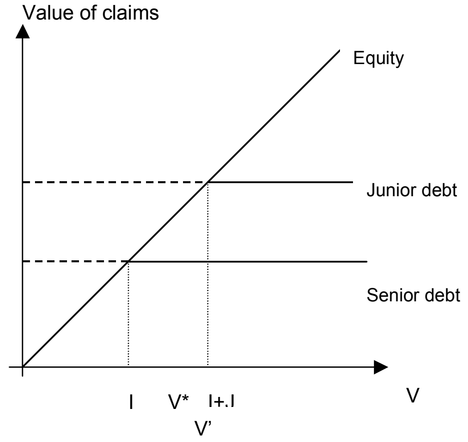
Explanatory variable	Hazard ratio	Robust std error	*z*	*P*>| *z* |
Dummy for (-*DD*) >-3.2	4.01**	2.55	2.19	0.029
S	1.01***	0.004	2.77	0.006
Dummy "high support"**S*	– 0.99**	0.005	– 2.41	0.016
Number of subjects	34	Time at risk		1494
Number of failures	10	Starting log likelihood		– 31.17
Number of observations	1494	Final log likelihood		– 27.94
Wald χ^2	12.90***	Zero-slope test (global test)		2.65

Note: Estimated using Cox regression. Log-likelihood given in the text. Standard errors are corrected for clustering using Wei and Lin's (1989) method. *,**, *** indicate statistical significance at the 10%, 5% and 1% levels, respectively.

Charts

Chart 1

Payoff profiles at maturity of equity, senior and junior debt



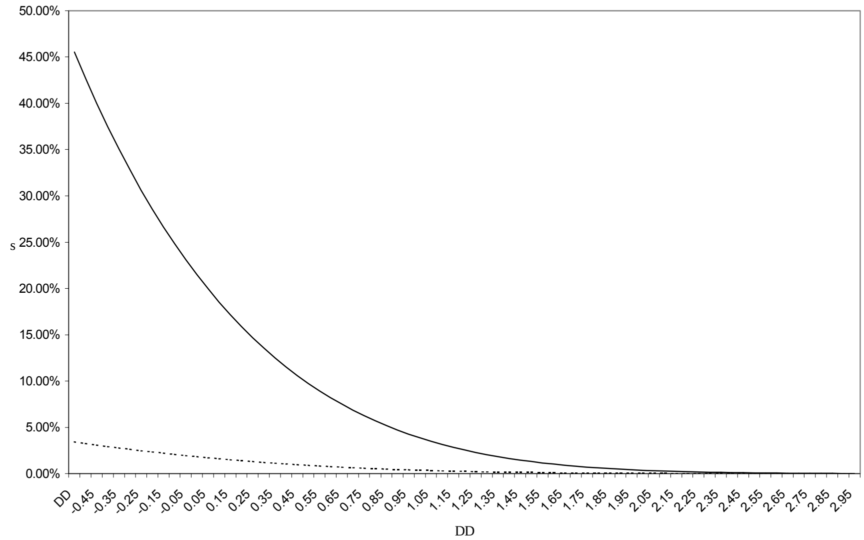
Payoffs at maturity

	V < I	I < *V* < I+J	*V* > I+J
Equity	0	0	*V*-I-J
Junior debt	0	*V*-I	J
Senior debt	*V*	I	I

Chart 2

Predicted spread (Black-Cox) (% of face value) as a function of distance-to-default

Subordinated debt (solid line), senior debt (dashed line)



Parameter assumptions: σ=0.05, r=0.05, T=1, I+J=1, I/(I+J)=0.9.

Chart 3

Summary of logit-estimation results

The chart displays the pattern of coefficients on the two indicators from Tables 6A and 6B with different horizons. The coefficients were normalised, such that the largest effect is equal to unity.

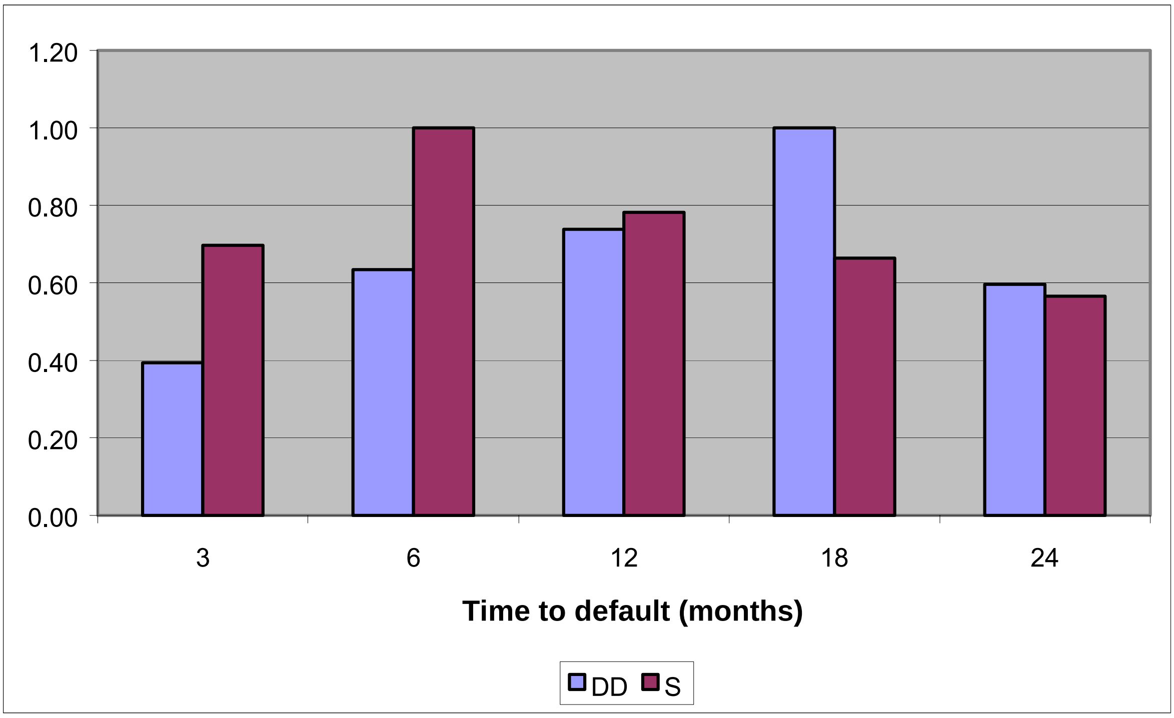
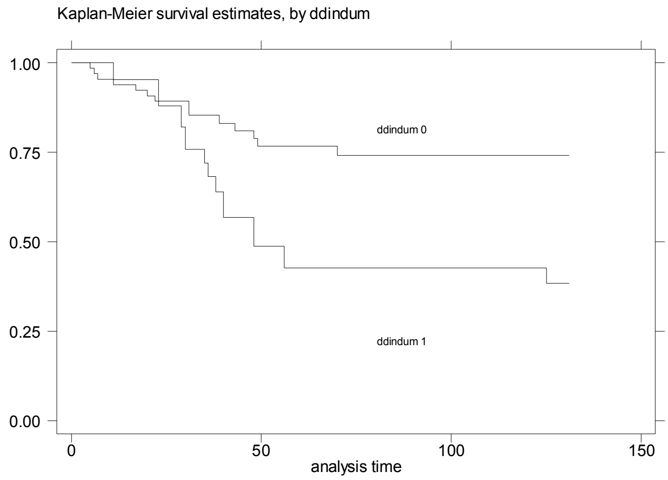


Chart 4

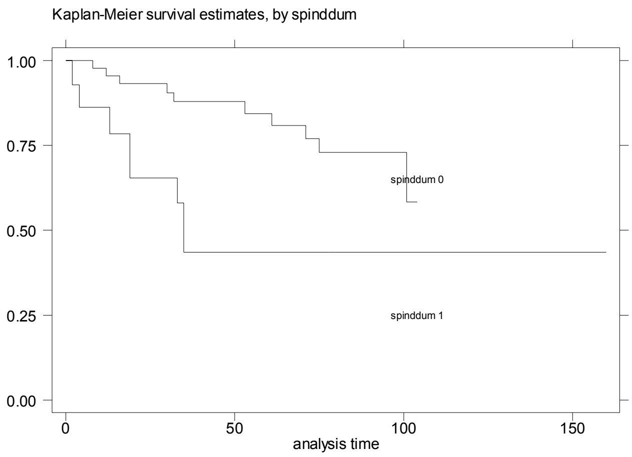
Survivor functions for the distance-to-default and spread

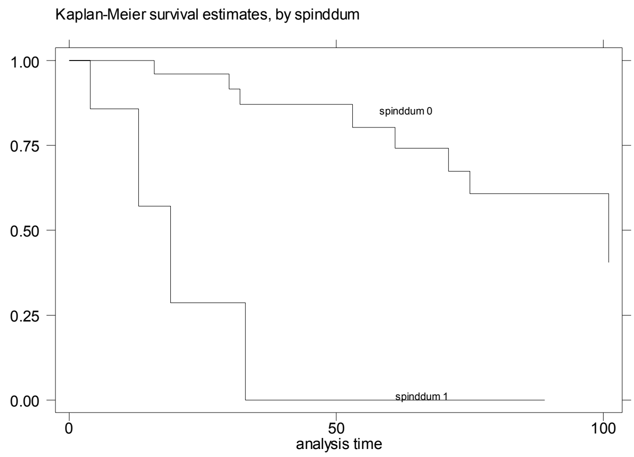
A. Distance-to-default



ddindum=1 if (–*DD*) > -3.2 and 0 otherwise. Analysis time is measured in months. Log-rank test for equality (χ^2 distributed) is equal to 6.08, which rejects equality at the 5%-level.

B. Spreads

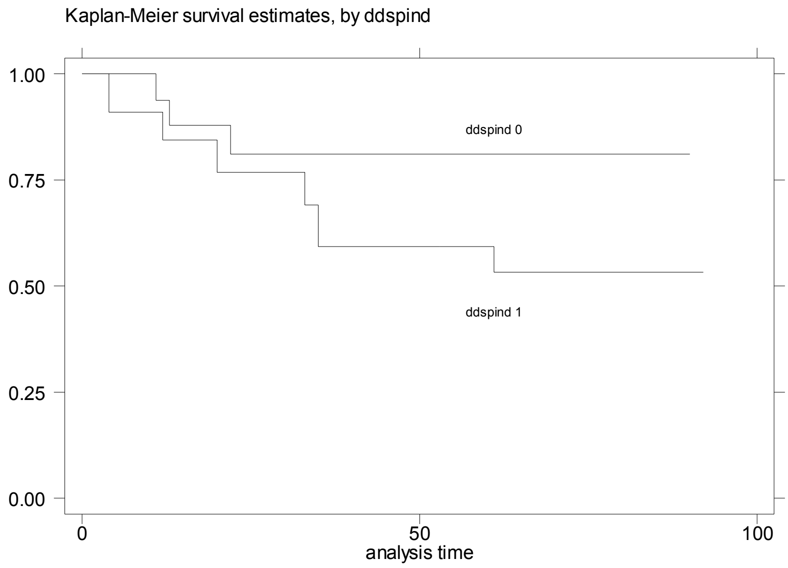




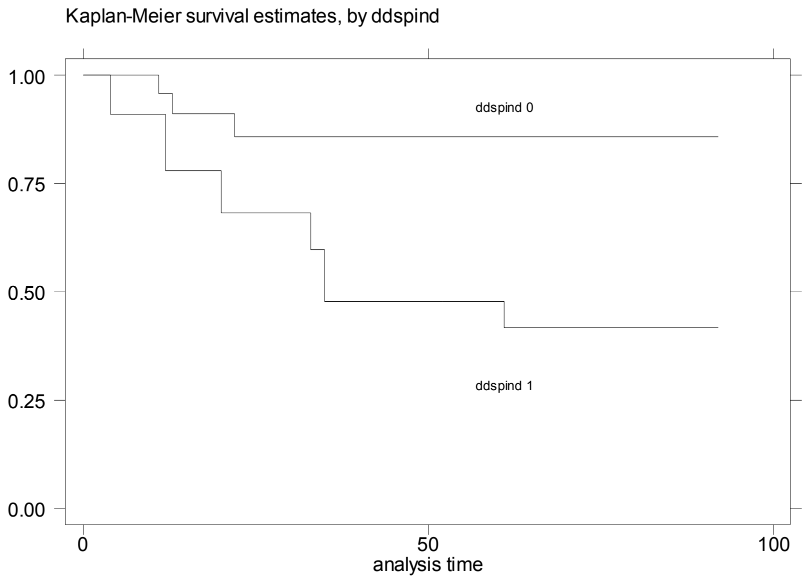
spinddum=1 if *S*>98 basis points and 0 otherwise. Panel B excludes UK banks. Analysis time is measured in months. Log-rank test for equality (χ^2 distributed) is equal to 4.73 and 25.9, respectively. Equality is rejected at the 5% (with UK banks) and at any significance level (without UK banks).

Chart 5

Survivor functions for the distance-to-default and spread, both indicators in the same model



At least one of the two indicators in top half and the other in top 75th percentile. Survival functions are not statistically significantly different (Chi squared of 1.04).



Both indicators in top half of the respective distributions. Survival functions are statistically significantly different at the 5% level (Chi squared of 4.1).

Appendix 1:
Distance-to-default according to the Black and Scholes formula[29]

In the BS model the time path of the market value of assets follows a stochastic process:

$$\ln V_A^T = \ln V_A + \left(r - \frac{\sigma_A^2}{2}\right)T + \sigma_A \sqrt{T}\,\varepsilon,$$

which gives the asset value at time T (ie maturity of debt), given its current value (V_A). ε is the random component of the firm's return on assets, which the BS model assumes normally distributed, with zero mean and unit variance, $N(0,1)$.

Hence, the current distance d from the default point (where $\ln V = \ln D$) can be expressed as:

$$d = \ln V^d - \ln D = \ln V \;\; + (r - \frac{\sigma^2}{2})T + \sigma\sqrt{T}\,\varepsilon - \ln D <=>$$

$$\frac{d}{\sigma\sqrt{T}} = \frac{\ln\left(\frac{V}{D}\right) + \left(r - \frac{\sigma^2}{2}\right)T}{\sigma\sqrt{T}} + \varepsilon.$$

That is, the distance-to-default (*DD*)

$$DD \equiv \frac{d}{\sigma\sqrt{T}} - \varepsilon = \frac{\ln\left(\frac{V}{D}\right) + \left(r - \frac{\sigma^2}{2}\right)T}{\sigma\sqrt{T}}$$

represents the number of standard deviations (σ_A) that the firm is from the default point.

The implied probability of default (*IPD*) can be defined as the probability that the asset value is less or equal to the book value of debt liabilities when the debt matures:

$$IPD \equiv \Pr\left[\ln V^T \leq \ln D\right] <=> \Pr\left[\ln V + \left(r - \frac{\sigma^2}{2}\right)T + \sigma\sqrt{T}\,\varepsilon \leq \ln D\right], \text{ie}$$

$$IPD = \Pr\left[-\frac{\ln\frac{V}{D} + \left(r - \frac{\sigma^2}{2}\right)T}{\sigma\sqrt{T}} \leq \varepsilon\right] = \Pr\left[(-DD) \leq \varepsilon\right].$$

Given that ε is normally distributed, *IPD*=*N*(-*DD*).

[29] See KMV Corporation (1999) for a similar derivation and more ample discussion.

Appendix 2:
Ratings definitions used by Fitch/IBCA

Fitch/IBCA's individual ratings attempt to assess how a bank would be viewed if it were entirely independent, and could not rely on external support. These ratings are designed to assess a bank's exposure to, appetite for and management of risk, and thus represent the view on the likelihood that it would run into significant difficulties. The principal factors analysed to evaluate the bank and determine these ratings include profitability and balance sheet integrity, franchise, management, operating environment and prospects.

Fitch/IBCA distinguishes among the following categories:

A. **A very strong bank.** Characteristics may include outstanding profitability and balance sheet integrity, franchise, management, operating environment or prospects.

B. **A strong bank.** There are no major concerns regarding the bank. Characteristics may include strong profitability and balance sheet integrity, franchise, management, operating environment or prospects.

C. **An adequate bank which, however, possesses one or more troublesome aspects.** There may be some concerns regarding its profitability and balance sheet integrity, franchise, management, operating environment or prospects.

D. **A bank which has weaknesses of internal and/or external origin.** There are concerns regarding its profitability and balance sheet integrity, franchise, management, operating environment or prospects.

E. **A bank with very serious problems which either requires or is likely to require external support.**

Note that, in addition, there are gradations between these five rating categories, ie A/B, B/C, C/D, and D/E.

The **support ratings** do not assess the quality of a bank. Rather, they are Fitch/IBCA's assessment of whether the bank would receive support should this be necessary:

1. A bank for which there is a clear legal guarantee on the part of the state, or a bank of such importance both internationally and domestically that, in Fitch/IBCA's opinion, support from the state would be forthcoming, if necessary. The state in question must clearly be prepared and able to support its principal banks.

2. A bank for which, in Fitch/IBCA's opinion, state support would be forthcoming, even in the absence of a legal guarantee. This could be, for example, because of the bank's importance to the economy or its historical relationship with the authorities.

3. A bank or bank holding company which has institutional owners of sufficient reputation and possessing such resources that, in Fitch/IBCA's opinion, support would be forthcoming, if necessary.

4. A bank for which support is likely but not certain.

5. A bank, or bank holding company, for which support, although possible, cannot be relied upon.

Literature

Altman, E and A Saunders (2000): "An analysis and critique of the BIS proposal on capital adequacy and ratings", mimeo, New York University Stern School of Business, January.

Aharoney, J and I Swary (1996): "Contagion effects of bank failures: evidence from capital markets", *Journal of Business* 56, pp 305-22.

Anderson, R and S Sundaresan (1996): "Design and valuation of debt contracts", *Review of Financial Studies* 9, pp 37-68.

Berger, A, S Davies and M Flannery (2000): "Comparing market and supervisory assessments of bank performance: who knows what when?", *Journal of Money Credit and Banking* 32.

Bergman, Z Yaacov, B Grundy and Z Weiner (1996): "General properties of option prices", *Journal of Finance* 51, 5, 1573-610.

Black, F and J Cox (1976): "Valuing corporate securities: some effects of bond indenture provisions", *Journal of Finance* 2, pp 351-67.

Black, F and M Scholes (1973): "The pricing of options and corporate liabilities", *Journal of Political Economy* 3, pp 637-54.

Bliss, R (2000): "The pitfalls in inferring risk from financial market data", mimeo, Federal Reserve Bank of Chicago, December.

Bliss, R and M Flannery (2000): "Market discipline in the governance of US bank holding companies", *Federal Reserve Bank of Chicago Working Paper* 03.

Board of the Governors of the Federal Reserve System (1999): "Using subordinated debt as an instrument of market discipline", Report of a study group, Washington DC, December.

Bongini, P, L Laeven and G Majnoni (2001): "How good is the market at assessing bank fragility? A horse race between different indicators", mimeo, World Bank, January.

Bruche, M (2001): "Bond pricing with co-ordination failure: implications for the use of spreads as an indicator of financial health", mimeo, European Central Bank.

Calomiris, C (1997): *The post-modern bank safety net: lessons from developed and developing countries*, American Enterprise Institute, Washington DC.

Demsetz, R, M Saidenberg and P Strahan (1996): "Banks with something to lose: the disciplinary role of franchise value", *Federal Reserve Bank of New York Economic Policy Review* 2, pp 1-14.

Dewatripont M and J Tirole (1993): *The prudential regulation of banks*, Editions Payot, Lausanne.

DeYoung, R, M Flannery, W Lang and S Sorescu (2000): "The information content of bank exam ratings and subordinated debt prices", *Journal of Money Credit and Banking*.

Docking, D S, M Hirschey and E Jones (1997): "Information and contagion effects of bank loan-loss reserve announcements", *Journal of Financial Economics* 43(2), pp 219-39, February.

Evanoff, D and L Wall (2000a): "Subordinated debt as bank capital: a proposal for regulatory reform", mimeo, Federal Reserve Bank of Chicago.

——— (2000b): "Sub-debt yields as bank risk measures", mimeo, Federal Reserve Bank of Chicago, November.

Federal Deposit Insurance Corporation (1994): "EMS/CAEL user manual", FDIC Division of Supervision Analysis and Monitoring Section, Washington DC, February.

Flannery, M (1998): "Using market information in prudential banking supervision: A review of US evidence", *Journal of Money Credit and Banking* 3, pp 273-305.

——— (2000): "The faces of market discipline", mimeo, University of Florida, October.

Flannery, M and S Sorescu (1996): "Evidence of bank market discipline in subordinated debenture yields: 1983-1991", *Journal of Finance* 4, pp 1347-77.

Franks, J and W Torous (1989): "An empirical investigation of US firms in renegotiation", *Journal of Finance* 44, pp 747-69.

Gilbert R A, A P Meyer and M D Vaughan (1999): "The role of supervisory screens and econometric models in off-site surveillance", *Federal Reserve Bank of St. Louis Review*, pp 31-56.

Goh, J and L Ederington (1993): "Is a bond rating downgrade bad news, good news, or no news for stockholders?" *Journal of Finance* 48, pp 2001-8.

Gropp, R and J-P Olters (2001): "The political economy of bank bail-outs", mimeo, European Central Bank, June.

Gropp, R and A Richards (2001): "Rating agency actions and the pricing of debt and equity of European banks: what can we infer about private sector monitoring of bank soundness?", *European Central Bank Working Paper* 76.

Gropp, R and J Vesala (2001): "Deposit insurance and moral hazard: does the counterfactual matter?", *European Central Bank Working Paper* 47.

Hancock, D and M L Kwast (2001): "Using subordinated debt to monitor bank holding companies: is it feasible?", *Journal of Financial Services Research* 20(2-3), pp 147-87, October-December.

Hand, J, R Holthausen and R Leftwich (1992): "The effect of bond rating agency announcements on bond and stock prices", *Journal of Finance* 57, pp 733-52.

Harrison, J and D Kreps (1979): "Call options and the risk of underlying securities", *Journal of Financial Economics* 13, 425-34.

Huber, P (1967): "The behaviour of maximum likelihood estimates under non-standard conditions", in Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, CA, University of California Press, pp 221-33.

Jagtiani, J, G Kaufman and C Lemieux (2000): "Do markets discipline banks and bank holding companies? Evidence from debt pricing", *Federal Reserve Bank of Chicago Emerging Issues Series* 3R.

Kalbfleisch, J and R Prentice (1980): "The statistical analysis of failure time data", *Wiley Series in Probability and Mathematical Statistics*, John Wiley and Sons, New York.

Keeley, M (1990): "Deposit insurance, risk, and market power in banking", *American Economic Review* 80, pp 1183-200.

KMV Corporation (1999): *Modelling default risk*, San Francisco.

Levonian, M (2001): "Subordinated debt and the quality of market discipline in banking", mimeo, Federal Reserve Bank of San Francisco, May.

Lin, D and L Wei (1989): "The robust inference for the Cox Proportional Hazard Model", *Journal of the American Statistical Association* 84, pp 1074-78.

Longstaff, F and E Schwartz (1995): "A simple approach to valuing risky and floating rate debt", *Journal of Finance 50*, pp 789-819.

Marcus, A and I Shaked (1984): "The valuation of FDIC deposit insurance using option-pricing estimates", *Journal of Money Credit and Banking* 16, 446-60.

Mella-Barral, P and W Perraudin (1997): "Strategic debt service", *Journal of Finance* 2, pp 531-56.

Merton, R (1977): "An analytic derivation of the cost of deposit insurance and loan guarantees", *Journal of Banking and Finance* 1, pp 3-11.

——— (1990): *Continuous time finance*, Blackwell, Oxford and Malden.

Morgan, D and K Stiroh (2001): "Bond market discipline of bank: is the market tough enough?", *Federal Reserve Bank of New York Working Paper*.

Musumeci, J and J Sinkey (1990): "The international debt crisis, investor contagion, and bank security returns in 1987: the Brazilian experience", *Journal of Money Credit and Banking* 22, pp 209-20.

Ritchken, P, J Thomson, R DeGennari and A Li (1993): "On flexibility, capital structure and investment decisions for the insured bank", *Journal of Banking and Finance*, 17, 1133-46.

Ron, E and A Verma (1986): "Risk-based capital adequacy standards for a sample of 43 major banks", *Journal of Banking and Finance* 13, pp 21-9.

Sironi, A (2000): "Testing for market discipline in the European banking industry: evidence from subordinated debt spreads", Board of Governors of the Federal Reserve System, *Finance and Economics Discussion Series*, No 2000-40, July.

Smirlock, M and H Kaufold (1987): "Bank foreign lending, mandatory disclosure rules, and the reaction of bank stock prices to the Mexican debt crisis", *Journal of Business* 60, pp 347-64.

The effect of VaR-based risk management on asset prices and the volatility smile[1]

Arjan Berkelaar, World Bank,[2] Phornchanok Cumperayot, Erasmus University, Rotterdam, and Roy Kouwenberg, Aegon Asset Management, The Hague

Abstract

Value-at-risk (VaR) has become the standard criterion for assessing risk in the financial industry. Given the widespread usage of VaR, it becomes increasingly important to study the effects of VaR-based risk management on the prices of stocks and options. We solve a continuous-time asset pricing model, based on Lucas (1978) and Basak and Shapiro (2001), to investigate these effects. We find that the presence of risk managers tends to reduce market volatility, as intended. However, in some cases VaR risk management undesirably raises the probability of extreme losses. Finally, we demonstrate that option prices in an economy with VaR risk managers display a volatility smile.

1. Introduction

Many financial institutions and non-financial firms nowadays publicly report value-at-risk (VaR), a risk measure for potential losses. Internal uses of VaR and other sophisticated risk measures are on the rise in many financial institutions, where, for example, a bank's risk committee may set VaR limits, both amounts and probabilities, for trading operations and fund management. At the industrial level, supervisors use VaR as a standard summary of market risk exposure.[3] An advantage of the VaR measure, following from extreme value theory, is that it can be computed without full knowledge of the return distribution. Semi-parametric or fully non-parametric estimation methods are available for downside risk estimation. Furthermore, at a sufficiently low confidence level the VaR measure explicitly focuses risk managers' and regulators' attention on infrequent but potentially catastrophic extreme losses.

Given the widespread use of VaR-based risk management, it becomes increasingly important to study the effects on the stock market and the option market of these constraints. For example, institutions with a VaR constraint might be willing to buy out-of-the-money put options on the market portfolio in order to limit their downside risk. If multiple institutions follow the same risk management strategy, then this will clearly lift the equilibrium prices of these options. Also the shape of the stock return distribution in equilibrium will be affected by the collective risk management efforts. As a result, it might even be the case that the distribution of stock returns will become more heavy-tailed. This would imply that the attempt to handle market risk, and thus to reduce default risk, has adversely raised the probability of such events.

Recently, Basak and Shapiro (2001) have derived the optimal investment policies for investors who maximise utility, subject to a VaR constraint, and found some surprising features of VaR usage. They show, in a partial equilibrium framework, that a VaR risk manager often has a higher loss in extremely

---

[1] This article was first published in *European Financial Management*, vol 8, issue 2, June 2002, pp 139-64. The copyright holder is Blackwell Publishers Ltd.

[2] Corresponding author: World Bank, Investment Management Department (MC7-300), 1818 H Street NW, Washington DC 20433, USA, tel: +1 202 473 7941, fax: +1 202 477 9015, e-mail: aberkelaar@worldbank.org. This paper reflects the personal views of the authors and not those of the World Bank. We would like to thank Suleyman Basak and Alex Shapiro for their helpful comments.

[3] The Bank for International Settlements (BIS) mandates internationally active financial institutions in the G10 countries to report VaR estimates and to maintain regulatory capital to cover market risk.

bad states than a non-risk manager. The risk manager reduces his losses in states that occur with (100 – α)% probability, but seems to ignore the α% of states that are not included in the computation of VaR. Starting from this equilibrium framework based on the Lucas pure exchange economy, in this paper we aim to further investigate Basak and Shapiro's (2001) very interesting and relevant question regarding the usefulness of VaR-based risk management.

In our economic setup, agents maximise the expected utility of intermediate consumption up to a finite planning horizon *T* and the expected utility of terminal wealth at the horizon. A portion of the investors in the economy are subject to a VaR risk management constraint, which restricts the probability of losses at the planning horizon *T*. As a result of our setup, asset prices do not drop to zero at the planning horizon and, moreover, we can ignore the unrealistic jump in asset prices that occurs just after the horizon of the VaR constraint, as in Basak and Shapiro (2001). We find that the VaR agents' investment strategies, depending on the state of nature, directly determine market volatility, the equilibrium stock price and the implied volatilities of options. In general VaR-based risk management tends to reduce the volatility of the stock returns in equilibrium and hence the regulation has the desired effect. In most cases the stock return distribution has a relatively thin left tail and positive skewness, which reduces the probability of severe losses relative to a benchmark economy without risk managers.

However, we also find that in some cases VaR-based risk management adversely amplifies default risk through a relatively heavier left tail of the return distribution. In very bad states the VaR risk managers switch to a gambling strategy that pushes up market risk. The adverse effects of this gambling strategy are typically strong when the investors consume a large share of their wealth, or when the VaR constraint has a relatively high maximum loss probability α. Additionally, we study option prices in the VaR economy. We find that the presence of VaR risk managers tends to reduce European option prices, and hence the implied volatilities of these options. Moreover, we find that the implied volatilities display a smile, as often observed in practice, unlike the benchmark economy, where implied volatility is constant.

We conclude that VaR regulation performs well most of the time, as it reduces the volatility of the stock returns and it limits the probability of losses. However, in some special cases, the VaR constraint can also adversely increase the likelihood of extremely negative returns. This negative side effect typically occurs if the investors in the economy have a strong preference for consumption instead of terminal wealth, or when the VaR constraint is rather loose (ie with high α). Note that the negative consequences of VaR-based risk management are mainly due to the "all or nothing" gambling attitude of the optimal investment strategy in case of losses, which might seem rather unnatural. In this paper we argue that the gambling strategy of a VaR risk manager might not be that unnatural for many investors, as it is closely related to the optimal strategy of loss-averse agents with the utility function of prospect theory.

Prospect theory is a framework for decision-making under uncertainty developed by the psychologists Kahneman and Tversky (1979), based on behaviour observed in experiments. The utility function of prospect theory is defined over gains and losses, relative to a reference point. The function is much steeper over losses than over gains and also has a kink in the reference point. Loss-averse agents dislike losses, even if they are very small, and therefore their optimal investment strategy tries to keep wealth above the reference point.[4] Once a loss-averse investor's wealth drops below the reference point, he tries to make up his previous losses by following a risky investment strategy. Hence, similar to a VaR agent, a loss-averse agent tries to limit losses most of the time, but starts taking risky bets once his wealth drops below the reference point. The optimal investment strategy under a VaR constraint might therefore seem rather natural for loss-averse investors. Or, conversely, one could argue that a VaR constraint imposes a minimum level of "loss aversion" on all investors affected by the regulation.

This paper is organised as follows: in Section 2, we define our dynamic economy and the market-clearing conditions required in order to solve for the equilibrium prices. Individual optimal investment decisions are also discussed. The general equilibrium solutions and analysis are presented in Section 3. We focus on the total return distribution of stocks and the prices of European options in the presence of VaR risk managers. Section 4 investigates the similarity between risk management

---

[4] This behaviour is induced by the kink in the utility function, ie first-order risk aversion; see Berkelaar and Kouwenberg (2001a).

policies based on VaR and the optimal investment strategy of loss-averse investors. Section 4 finally summarises the paper and presents our conclusions.

2. Economic setting

2.1 A dynamic economy

In this section, the pure exchange economy of Lucas (1978) is formulated in a continuous-time stochastic framework. Suppose in a finite horizon, $[0,T]$, economy, there are heterogeneous economic agents with constant relative risk aversion (CRRA). The agents are assumed to trade one riskless bond and one risky stock continuously in a market without transaction costs.[5] There is one consumption good, which serves as the numeraire for other quantities, ie prices and dividends are measured in units of this good. The bond is in zero net supply, while the stock is in constant net supply of 1 and pays out dividends at the rate $\delta(t)$, for $t \in [0,T]$. The dividend rate is presumed to follow a Geometric Brownian motion:[6]

$$d\delta(t) = \mu_\delta \delta(t)dt + \sigma_\delta \delta(t)dB(t) \tag{1}$$

with $\mu_\delta > 0$ and $\sigma_\delta > 0$ constant.

The equilibrium processes of the riskless money market account $S_0(t)$ and the stock price $S_1(t)$ are the following diffusions, as will be shown in Section 3.1:

$$dS_0(t) = r(t)S_0(t)dt\ , \tag{2}$$

$$dS_1(t) + \delta(t) = \mu(t)S_1(t)dt + \sigma(t)S_1(t)dB(t),$$

where the interest rate $r(t)$, the drift rate $\mu(t)$ and the volatility $\sigma(t)$ are adapted processes and possibly path-dependent.

As we assume a dynamically complete market, these price processes ensure the existence of a unique state price density (or pricing kernel) $\xi(t)$, following the process

$$\frac{d\xi(t)}{\xi(t)} = -r(t)dt - \kappa(t)dB(t),\ \xi(0) = 1, \tag{3}$$

where $\kappa(t) = (\mu(t) - r(t))/\sigma(t)$ denotes the process for the market price of risk (Sharpe ratio).

Following from the law of one price, the pricing kernel $\xi(t)$ relates future dividend payments $\delta(s)$, $s \in (t,T]$ to today's stock price $S_1(t)$:

$$S_1(t) = \frac{1}{\xi(t)} E_t\left[\int_t^T \xi(s)\delta(s)ds\right]. \tag{4}$$

Intuitively the stock price is the price you pay to achieve a certain dividend in each state at each time t. Equation (4) is simply an over-time summation of the Arrow-Debreu security prices, discounting the future dividend payouts to today's value. The state price density process will therefore play an important role in deriving the equilibrium prices.

---

5 Basak and Shapiro (2001) assume N risky assets. However, our results are robust to the number of assets.

6 All mentioned processes are assumed to be well defined and satisfy the appropriate regularity conditions. For technical details, see Karatzas and Shreve (1998).

2.2 Preferences, endowments and risk management

Suppose there are two groups of agents in the economy: non-risk-managing and risk-managing agents. Agents belonging to the former group freely optimise their investment strategy, ie without risk management constraints, whereas the latter group is obligated to take a VaR restriction as a side constraint when structuring portfolios.[7] We assume that a proportion λ of the agents is not regulated, while the remaining proportion $(1 - \lambda)$ is. Each agent is endowed at time zero with initial wealth $W_i(0)$. We use subscript $i = 1$ for the unregulated agents and $i = 2$ for the risk managers. For both groups of agents we define a non-negative consumption process $c_i(t)$ and a process for the amount invested in stock $\pi_i(t)$. The wealth $W_i(t)$ of the agents then follows the process below:

$$dW_i(t) = r(t)W_i(t)dt + (\mu(t) - r(t))\pi_i(t)dt - c_i(t)dt + \sigma(t)\pi_i(t)dB(t), \tag{5}$$

for $i = 1, 2$; $\forall t \in [0,T]$.

As in the case of asset prices, today's wealth can be related to future consumption and terminal wealth through the state price density process $\xi(t)$:

$$W_i(t) = \frac{1}{\xi(t)} E_t\left[\int_t^T \xi(s)c_i(s)ds + \xi(T)W_i(T)\right]. \tag{6}$$

The agents maximise their utility from intertemporal consumption in $[0,T]$ and terminal wealth at the planning horizon T, which are represented by $U_i(c_i(t))$ and $H_i(W_i(T))$ respectively. The parameter $\rho_1 > 0$ determines the relative importance of utility from terminal wealth compared to utility from consumption. The planning problem for an unregulated agent then is:

$$\begin{aligned}
&\max_{c_1,\pi_1} \quad E\left[\int_0^T U_1(c_1(s))ds + \rho_1 H_1(W_1(T))\right]. \\
&\text{s.t.} \quad dW_1(t) = r(t)W_1(t)dt + (\mu(t) - r(t))\pi_1(t)dt - c_1(t)dt + \sigma(t)\pi_1(t)dB(t), \\
&\qquad W_1(t) \geq 0, \text{ for } \forall t \in [0,T].
\end{aligned} \tag{7}$$

Additionally, in order to limit the likelihood of large losses, the risk managers have to take a VaR constraint into account. Based on the practical implementation of VaR and its interpretation by Basak and Shapiro (2001), at the horizon T the maximum likely loss with probability $(1 – \alpha)$% over a given period, namely VaR(α), is mandated to be equal to or below a prespecified level. More precisely, the agents are allowed to consume continuously but make sure that, only with probability α% or less, their wealth $W_2(T)$ falls below the critical floor level $\underline{W}$. Therefore, the second group of agents faces the following optimisation problem with the additional VaR constraint:

$$\begin{aligned}
&\max_{c_2,\pi_2} \quad E\left[\int_0^T U_2(c_2(s))ds + \rho_2 H_2(W_2(T))\right] \\
&\text{s.t.} \quad dW_2(t) = r(t)W_2(t)dt + (\mu(t) - r(t))\pi_2(t)dt - c_2(t)dt + \sigma(t)\pi_2(t)dB(t), \\
&\qquad W_2(t) \geq 0, \text{ for } \forall t \in [0,T], \\
&\qquad P[W_2(T) \geq \underline{W}] \geq 1 - \alpha.
\end{aligned} \tag{8}$$

We assume that all agents have constant relative risk aversion over intertemporal consumption $U_i(c_i(t)) = V_{CRRA}(c_i(t))$ and over terminal wealth $H_i(W_i(T)) = V_{CRRA}(W_i(T))$ for $i = 1, 2$, where $V_{CRRA}(\cdot)$ is a power utility function:

$$V_{CRRA}(x) = \frac{1}{1-\gamma} x^{1-\gamma}, \text{ for } \gamma > 0;\ x > 0. \tag{9}$$

---

7 It should be noted that the superfluous risk management critique (see Modigliani and Miller (1958), Stiglitz (1969a,b and 1974), DeMarzo (1988), Grossman and Vila (1989) and Leland (1998)), does not hold at the individual level. The critique states that risk management is irrelevant for institutions and firms since individuals can undo any financial restructuring by trading in the market. This paper considers individual agents, and hence this line of reasoning is invalid here.

Note that the power utility function (9) is increasing and strictly concave and hence agents are assumed to be risk-averse. By assuming a common power utility function for intertemporal consumption and terminal wealth, we can isolate the effect of VaR-based risk management on asset prices. Our main purpose is to study the influence of risk management on the equilibrium price of the risky asset and on option prices.

In this paper, the VaR horizon coincides with the investment horizon, which is different from Basak and Shapiro's (2001) work. In their work, agents are concerned with the optimal consumption path over their lifetime, while obligated to a one-time-only risk evaluation. The VaR condition is supposed to be satisfied at some intermediate time, before the end of the agent's life. As a consequence of this setup, a severe jump in the price level occurs when the VaR condition is lifted. In addition, since agents consume everything at the planning horizon *T* and wealth drops to zero at that time, the corresponding asset prices go to zero.

In this paper, the horizon is just a subperiod of the lifetime in Basak and Shapiro (2001). Thus, agents can evaluate their VaR performance at the end of each period, eg a 10-day or an annual report, along the way maximising the utility from their intertemporal consumption as well as their terminal wealth. In our perspective, this adjustment makes the model more realistic. At the horizon agents may end up with claims on the assets and prices do not necessarily drop to zero. Moreover, within our setup we can ignore the jump in asset prices that occurs directly after the VaR horizon in Basak and Shapiro (2001). Note that this jump in asset prices occurs because all regulated investors drop the VaR constraint collectively, at a prespecified point in time. We think that such a coordinated abandonment of risk management policies is rather hypothetical and therefore we do not analyse the consequences of a jump in asset prices in this paper.

2.3 Equilibrium conditions and optimal decisions

In order to investigate the equilibrium asset prices in an economy with VaR risk managers, in this section we discuss the conditions that should be satisfied in any general equilibrium. In equilibrium, each agent optimises his individual consumption-investment problem. Moreover, as the consumption good cannot be stored, it follows that aggregate consumption in the economy has to equal aggregate dividends at each time $t \in [0,T]$. Additionally, from Walras law it follows that all markets have to clear, eg the good and the riskless securities markets, given that the stock market is in equilibrium at each time $t \in [0,T]$. Combined, this gives the following set of equilibrium conditions:

$$\lambda c_1^*(t) + (1-\lambda)c_2^*(t) = \delta(t), \tag{10}$$

$$\lambda \pi_1^*(t) + (1-\lambda)\pi_2^*(t) = S_1(t),$$

$$\lambda W_1^*(t) + (1-\lambda)W_2^*(t) = S_1(t),$$

where $c_1^*(t)$ and $\pi_1^*(t)$ are the optimal consumption and investment decisions for each type of agent i = 1, 2 and $W_i^*(t)$ is the corresponding optimal wealth process.

Typically the optimal policies for agents with power utility can be derived with dynamic programming, as in Merton (1969), as well as with the martingale methodology of Cox and Huang (1989), Karatzas et al (1987) and Pliska (1986). However, for the risk managers, the binding VaR restriction induces non-concavity into the optimisation problem (through the wealth function). Following Basak and Shapiro (2001), we derive the optimal policies for the regulated agents with power utility by applying the martingale methodology. The dynamic optimisation problem melts down to the following problem:

$$\max_{c_i,\pi_i} \quad E\left[\int_0^T U_i(c_i(s))ds + \rho_i H_i(W_i(T))\right]$$

$$\text{s.t.} \quad E\left[\int_0^T \xi(s)c_i(s)ds + \xi(T)W_i(T)\right] \leq \xi(0)W_i(0), \tag{11}$$

$$W_i(T) \geq 0,$$

for i = 1, 2, with an additional constraint for the risk managers:

$$P[W_2(T) \geq \underline{W}] \geq 1-\alpha. \tag{12}$$

In the next proposition we characterise the optimal consumption paths and terminal wealth profiles for both groups of agents:

Proposition 1

For any state price density following the process (3), the optimal intertemporal consumption policies $c_1^*(t)$ and terminal wealth profiles $W_i^*(T)$ of both groups of agents *i* = 1, 2 are

$$c_1^*(t) = I_1(y_1\xi(t)) = (y_1\xi(t))^{-1/\gamma}, \tag{13}$$

$$c_2^*(t) = I_2(y_2\xi(t)) = (y_2\xi(t))^{-1/\gamma},$$

$$W_1^*(T) = (y_1\xi(T)/\rho_1)^{-1/\gamma},$$

$$W_2^*(T) = \begin{cases} I_2(y_2\xi(T)/\rho_2) & \text{for } \xi(T) < \underline{\xi} \\ \underline{W} & \text{for } \underline{\xi} \le \xi(T) \le \overline{\xi} \\ I_2(y_2\xi(T)/\rho_2) & \text{for } \overline{\xi} \le \xi(T) \end{cases}$$

where $I_i(z)$ denotes the inverse of the marginal utility function $z = U_i'(x)$, $\underline{\xi} = \rho_2 U_2'(\underline{W})/y_2$, $\overline{\xi}$ is such that $P[\xi(T) > \overline{\xi}] = \alpha$ and $y_i \ge 0$ is a Lagrange multiplier that satisfies

$$E\left[\int_0^T \xi(s)c_i^*(s)ds + \xi(T)W_i^*(T)\right] = \xi(0)W_i(0), \text{ for } i = 1, 2. \tag{14}$$

In order to facilitate the derivation of equilibrium prices in the following sections, we additionally use the following proposition proved by Karatzas et al (1990) and Basak (1995). It provides a different representation of the equilibrium conditions by applying the martingale methodology.

Proposition 2

If there exists a state price density process $\xi(t)$ satisfying

$$\delta(t) = \lambda I_1(y_1\xi(t)) + (1-\lambda)I_2(y_2\xi(t)), \tag{15}$$

where y_1 and y_2 are Lagrange multipliers defined in (14), then the equilibrium conditions (10) are satisfied by the corresponding optimal consumption and investment policies.

Before we actually derive the equilibrium prices, we will first discuss the optimal investment strategies in partial equilibrium of the unrestricted agents (benchmark agents), the regulated agents (the VaR risk managers) and the portfolio insurers. Portfolio insurers are VaR risk managers that do not tolerate any losses below $\underline{W}$, ie they represent the extreme case $\alpha = 0$.[8] It is noteworthy that portfolio insurers fully insure against all states of nature at the minimal wealth level $\underline{W}$, whereas VaR risk managers with $\alpha > 0$ only partially insure. VaR risk managers with $\alpha > 0$ do not necessarily insure in expensive states that occur with very low probability.

In Figure 1, we display the optimal terminal wealth profiles for the three types of agents. All agents have initial endowment of $W(0) = 1$ and power utility over consumption and over terminal wealth with risk aversion parameter $\gamma = 1$, which is in limit equivalent to log utility. We set the trade-off between consumption and terminal wealth at $\rho_1 = \rho_2 = 10$, to avoid excessive consumption. The maximum loss probability α of the risk managers and the portfolio insurers is equal to 1% and 0 respectively. The critical wealth $\underline{W}$ at time $T = 1$ is 0.95. The interest rate and the Sharpe ratio are given by 5% and 0.4. Endogenous thresholds, $\underline{\xi}$ and $\overline{\xi}$, are then calculated.

8 See Basak (1995), Basak and Shapiro (1999), Grossman and Vila (1989) and Grossman and Zhou (1996).

To be more specific, throughout the paper, states of the world are roughly classified into three ranges. First, good states are the states in which the VaR constraint is not binding since the optimal (terminal) wealth, as a consequence of individual optimisation, is likely to be above or equal to the minimal wealth threshold. Second, intermediate states $\left(\underline{\xi},\overline{\xi}\right]$ are the states in which the unrestricted individual's wealth is likely to end up below the threshold. Thus, VaR risk management becomes binding in these states. Last, extremely bad states are the small area of worst cases with a total probability mass of α. In these states the VaR managers will no longer try to keep wealth above $\underline{W}$. Note that sizes of these ranges depend on the minimal wealth threshold $\underline{W}$ and the probability of loss α for the VaR managers.

Figure 1 shows that the optimal terminal wealth function of the benchmark agents is decreasing smoothly from good states of the world (low $\xi(T)$) to bad states (high $\xi(T)$), whereas the partial insurers and full insurers behave differently. In the good states region (low $\xi(T)$), they all have smoothly decreasing curves. However, for the portfolio insurers to guarantee terminal wealth in intermediate and extreme states, from $\underline{\xi}$ onwards, they have to pay based on the pricing kernel, ie the worse the state implies the higher the price to obtain a fixed amount of wealth in that particular state. Therefore, the portfolio insurers consume relatively less in good states in order to pay for the insurance in bad states.

Unlike the portfolio insurers, the VaR-based risk managers are not obligated to insure in the most expensive states (high $\xi(T)$). In very bad states, from $\overline{\xi}$ onwards, terminal wealth dramatically drops and remains lower than the benchmark level. In terms of probability, the VaR managers have high probability of achieving the critical wealth threshold that they insure in intermediate states. However, lower payoff in bad states leads to a heavier tail in the lower quantiles of the wealth distribution, compared to other types of investors. Note that the area on the right of $\overline{\xi}$ contains only α% of probability mass. Hence, the area where the VaR risk manager tolerates losses is less probable than it might seem in the figure.

Next we analyse the dynamic (state-varying) optimal investment policies of the VaR managers relative to the benchmark agents and the portfolio insurers. In order to do so, we assume for now Geometric Brownian motions for the dividend rate and the stock price with constant interest rate r = 5% and constant market price of risk κ = 0.4. Later on we will see that in the general equilibrium framework the interest rate and the market price of risk are indeed constant. As in the previous example, initial wealth is $W(0) = 1$ and $T = 1$, while the current time is $t = 0.75$.

In Figure 2, the optimal wealth of three kinds of investors is shown at an intermediate point in time ($t = 0.75$). The figure demonstrates the common feature of a steeply decreasing wealth function, with a slower descent as the states are worsening. As in Figure 1, the benchmark agent's wealth is a decreasing convex function, while the portfolio insurers always keep their wealth above the level $\exp(-r(T-t))\underline{W}$. The time-$t$ wealth profile of the VaR risk manager with α = 1% lies in between these two extremes. It looks like a smoothed version of the terminal wealth profile at the horizon T, without the abrupt jump at $\overline{\xi}$.

Figure 3 displays the optimal portfolio weight of stocks relative to the benchmark. It reveals how VaR-based investors change their investment strategy to insure their terminal wealth. In good states the VaR risk-managing agents follow the benchmark behaviour. In the middle range, when the state price is relatively low they replicate the portfolio insurers' strategy, which is to invest in riskless assets. In bad states the VaR agents greatly increase their exposure to the risky asset, as they try to maximise the probability of their wealth staying above the critical threshold $\underline{W}$. However, to avoid bankruptcy, the risk managers eventually reduce their risk exposure again.[9]

Concluding: the VaR restriction reduces the exposure to risky assets in good states relative to the benchmark, while in extremely bad states it stimulates risky-asset holding in a desperate attempt to make up for low wealth (gambling). Thus, outside the area where the VaR constraint is binding, logically the VaR agents' wealth is always lower than in the benchmark case. In the following section,

[9] Please note that states with a pricing kernel higher than three in Figure 3 have almost no probability of occurring and are only shown for illustrative purposes.

we will investigate the impact of this investment strategy on market risk, the equilibrium asset prices and the total return distribution. The usefulness of VaR regulation, implicitly or explicitly, will be discussed in Section 5.

3. General equilibrium with VaR risk managers

3.1 Closed-form solutions

In this section, we expand the individual optimisation problem of the previous section to a Lucas general equilibrium model, based on the market clearing conditions (10). In the economy, there is a fraction λ of agents, who are unrestricted, and the remaining fraction $1-\lambda$, who apply VaR-based risk management. Our economy is formulated such that, in some subperiod, agents maximise their utility from intermediate consumption, $t \in [0,T]$, and from terminal wealth at the horizon *T*. The VaR restriction is imposed on the risk managers at time *T*, coinciding with their planning horizon.

As a first step in solving the equilibrium model, it is convenient to derive the equilibrium state price density as a function of aggregate dividends by inverting equation (15). Given the stochastic process for the state price density in (3), as a second step we can infer the equilibrium interest rate and market price of risk. The following proposition summarises these general results:

Proposition 3

In any economy with $0 \le \lambda \le 1$ equilibrium exists and the state price density is given by

$$\xi(t) = (v(y_1, y_2)\delta(t))^{-\gamma}, \tag{16}$$

where

$$v(y_1, y_2) = \left(\lambda y_1^{-1/\gamma} + (1-\lambda) y_2^{-1/\gamma}\right)^{-1}. \tag{17}$$

The equilibrium interest rate and market price of risk processes are constant:

$$r(t) = \gamma\left(\mu_\delta - 1/2(1+\gamma)\sigma_\delta^2\right), \tag{18}$$

$$k(t) = \gamma\sigma_\delta .$$

An important conclusion from proposition (3) is that the interest rate *r* and the market price of risk *k* are constant in equilibrium. Furthermore, the fraction of risk managers in the economy does not affect the interest rate and the market price of risk. Hence, in our setup the presence of risk managers will only have an impact on the price level of the stock, on its drift rate μ and volatility σ. Note, however, that the mean and volatility always have to move in lockstep due to the constant market price of risk. In addition, based on the assumption of a Geometric Brownian for the dividend process, equation (16) implies a log-normal state price density.

The constant interest rate and market price of risk arise due to the assumption that both groups of agents share an identical power utility function over intertemporal consumption. Although this economic setup leads to some inflexibility, a major advantage is that we can derive closed-form solutions for the equilibrium prices and hence fully analyse the economic problem at hand. Basak (1995) and Basak and Shapiro (2001) also impose equivalent assumptions in order to study equilibrium with portfolio insurers and value-at-risk regulation respectively. Without an identical utility function over consumption for both groups of agents, we would have to resort to numerical techniques as in Grossman and Zhou (1996).[10]

---

10 The same holds if we leave out consumption and only assume utility over terminal wealth.

Given the state price density of proposition (3), we can derive the equilibrium stock price from the equilibrium conditions (10). Once a closed-form expression has been obtained, the drift rate μ and volatility σ of the stock price process follow straightforwardly from Ito's lemma. Below we first present the equilibrium price in a benchmark economy with unrestricted investors only ($\lambda = 1$), before we consider our general results in the presence of risk managers:

Proposition 4

The equilibrium price of the risky asset in an economy with unregulated agents only ($\lambda = 1$) is

$$S_1(t) = \delta(t)\left(a(t) + \rho_1^{1/\gamma} e^{\eta(T-t)}\right), \tag{19}$$

with

$$a(t) = \frac{1}{\eta}\left(e^{\eta(T-t)} - 1\right),\ \eta = (1-\gamma)\mu_\delta - 1/2\gamma(1-\gamma)\sigma_\delta^2. \tag{20}$$

The stock price follows a Geometric Brownian motion with constant drift rate and volatility given by

$$\mu(t) = \gamma\left(\mu_\delta + 1/2(1-\gamma)\sigma_\delta^2\right),\ \sigma(t) = \sigma_\delta . \tag{21}$$

In the case of unregulated agents with power utility, the interest rate is constant and the stock price follows a Geometric Brownian motion, resembling the familiar Black-Scholes assumptions for option pricing. As a result, it has a log-normal (Gaussian-class) distribution. If we additionally introduce VaR-based risk managers ($\lambda < 1$), then the equilibrium stock price process changes quite drastically:

Proposition 5

The equilibrium price of the risky asset in an economy with both unregulated agents and VaR risk managers is

$$\begin{aligned} S_1(t) &= a(t)\delta(t) + \lambda(y_1/\rho_1)^{-1/\gamma} v(y_1, y_2) e^{\eta(T-t)}\delta(t) \\ &+ (1-\lambda)(y_2/\rho_2)^{-1/\gamma} v(y_1, y_2) e^{\Gamma(t)}\delta(t)\left(1 - N\left(-d\left(\delta(t), t, \underline{\xi}\right)\right) + N\left(-d\left(\delta(t), t, \overline{\xi}\right)\right)\right) \\ &+ (1-\lambda)\underline{W} e^{-r(T-t)}\left(N\left(-f\left(\delta(t), t, \underline{\xi}\right)\right) - N\left(-f\left(\delta(t), t, \overline{\xi}\right)\right)\right) \end{aligned} \tag{22}$$

with

$$f(\delta, t, x) = \frac{\log(x) + \gamma\left(\log(\delta(t)) + \log(v(y_1, y_2))\right) + \left(r - \frac{1}{2}\kappa^2\right)(T-t)}{\kappa\sqrt{(T-t)}}, \tag{23}$$

$$d(\delta, t, x) = f(\delta, t, x) + \frac{1}{\gamma}\kappa\sqrt{(T-t)} \tag{24}$$

$$\Gamma(t) = \frac{1-\gamma}{\gamma}\left(r + \frac{1}{2}\kappa^2\right)(T-t) + \frac{1}{2}\left(\frac{1-\gamma}{\gamma}\right)^2 \kappa^2 (T-t) \tag{25}$$

and $N(\cdot)$ is the standard-normal cumulative distribution function.

In an economy with VaR-regulated investors, the drift rate μ_t and the volatility σ_t of the stock price process are no longer constant, which can be easily verified by applying Ito's lemma to the stock price formula (22). Figures 4 and 5 show that at a point in time the expected drift rate and the volatility can be either high or low relative to the benchmark case, depending on the state of the world. In the next section we will analyse these drift rates and the volatility processes. Moreover, the impact of VaR-based risk management on the equilibrium stock price, its total return distribution and the equilibrium option prices will be examined respectively.

3.2 Drift rate, volatility and prices in equilibrium

For the numerical examples, we set the parameters of the dividend process as $\mu_\delta = 0.056$ and $\sigma_\delta = 0.115$, based on monthly S&P 500 index data from 1980 up to 1999. All agents have initial wealth $W(0) = 1$ and maximise a power utility over intertemporal consumption and terminal wealth with risk aversion $\gamma = 1$. We set the trade-off between consumption and terminal wealth at $\rho_1 = \rho_2 = 10$. The maximum loss probability α is 100%, 1% and 0 respectively for the unregulated agents, the risk managers and the portfolio insurers. The critical wealth $\underline{W}$ threshold at the planning horizon $T = 1$ is 0.95, while the current time is $t = 0.75$. The insurer economy and the risk manager economy both contain 50% of the benchmark agents and 50% of their own population (ie $\lambda = 1/2$).

In order to take a closer look at the stock price process, first we present the expected drift rate and the volatility of the stock returns as functions of the pricing kernel. In Figures 4 and 5, one can see that in all three types of economies the market volatility and the drift rate move in lockstep as a consequence of the constant market price of risk. In the benchmark economy, the mean and the volatility of the stock price are constant in every state at all times, corresponding to the Black-Scholes assumptions for option pricing. In the portfolio insurance economy, the expected returns and the volatility are lower in most states, compared to the benchmark. Intuitively, the portfolio insurers want to hold less equity. In order to clear the market, the attractiveness of stocks has to be increased and, hence, the equilibrium volatility is reduced. The portfolio insurance strategy therefore reduces volatility and stabilises the economy.

For exactly the same reasons, the volatility is reduced in intermediate states in the partially insured VaR economy. However, in bad states the VaR agents abandon the portfolio insurance strategy and instead they start to increase their demand for risky assets, as can been seen in Figure 3.[11] The overwhelming demand caused by this gambling policy leads to a relatively high volatility and drift rate, in order to clear the market. Eventually, in very bad states, the portfolio weight of the VaR agents returns to normal levels in order to avoid bankruptcy. As a result, the equilibrium volatility and drift rate are then pulled back to the benchmark level.

We will now analyse the consequences of the presence of VaR risk managers on the equilibrium stock price. Figure 6 displays the equilibrium price of the risky asset in different states of nature, ranging from good to bad, for the benchmark economy with $\lambda = 1$ and the mixed economies with $\lambda = 1/2$, at the intermediate time $t = 0.75$. For all economies, the price of the risky asset is generally high in good states of the world (low $\xi(t)$) and low in bad states (high $\xi(t)$). This follows from the inverse relation between the state price density and aggregate dividends in proposition 3, namely a higher pricing kernel level implies lower aggregate dividends and consequently a lower stock price.

We start by explaining the price function in the portfolio insurance economy (the dotted line). In good states the price function in the economy with portfolio insurers has a similar shape as in the benchmark case; however, it lies at a lower level. This follows from the fact that the constrained agents have less wealth and therefore they demand less stocks: the equilibrium price has to be lower in order to clear the market. In intermediate and bad states, the prices in the portfolio insurance economy are higher than in the benchmark economy. In these states the equilibrium volatility in the portfolio insurance economy is considerably lower than in the benchmark case. The low volatility increases the demand for stocks and hence also increase the equilibrium price.

The price function in the economy with VaR risk managers is similar to the pattern in the portfolio insurance economy in good and intermediate states. However, in bad states, the VaR risk managers switch from the portfolio insurance strategy to a risky gambling strategy and this alters the shape of the price function. Due to the increased demand for stocks in bad states, the volatility increases in order to clear the market. While volatility increases and makes the stock less attractive, the price function drops rather fast and ends up below the benchmark case. Finally, in extremely bad states, the VaR agents reduce their risk exposure to avoid bankruptcy, volatility returns to regular levels and the shape of the price function becomes similar to the benchmark case again.

---

[11] Note that Figure 3 is not used for a quantitative analysis but for qualitative interpretation only, since the drift rate and volatility do not coincide with the equilibrium case in Figures 4 and 5.

3.3 The return distribution and option prices in equilibrium

Figures 7, 8 and 9 show the annualised total return distribution of the stock, including the estimated divided yield. As the dividend payout rate changes stochastically through time, an exact calculation of the total dividend payments would require extensive simulation. Instead we approximate the stochastic dividend payout $\int_0^t \delta(z)dz$ between time zero and t by $(\delta(t)-\delta(0))t$. As the dividend process is the same in each economy, we do not think that the error inherent in this approximation will have a serious impact on our analysis or conclusions.

In the benchmark economy, the returns are log-normally distributed with slight positive skewness and kurtosis close to three, as the stock price follows a Geometric Brownian motion. Figure 7 shows the effect on the stock return distribution of a VaR constraint applied to 50% of the investors (ie λ = 50%), with critical wealth level $\underline{W}$ = 0.95 and maximum loss probability α = 0%, 1% and 5%. In an economy with VaR risk managers or portfolio insurers, the volatility of the stock return distribution is clearly lower. Moreover, the probability of negative returns up to –25% has also decreased substantially. Hence, the VaR restriction seems to stabilise the economy. Note, however, that the probability of negative returns in excess of –25% has increased in the case of a loose VaR constraint with α = 5%, due to increased risk-taking at lower levels of wealth.

Figure 8 shows the return distribution in economies with a VaR constraint with α = 1%, at different critical wealth levels $\underline{W}$ = 0.90, 0.95 and 0.99. The volatility of the stock returns tends to decrease as the critical wealth level becomes higher and hence the VaR constraint becomes tighter. Note that in the case of a very tight constraint, with $\underline{W}$ = 0.99, the left tail of the distribution becomes thicker and extreme returns below –30% are more likely than in the unregulated economy. In general, though, the economies with a VaR constraint are less volatile and the regulation has the intended effect.

Another important parameter affecting the economy is ρ_i, which determines the trade-off between the utility of intertemporal consumption and the utility of terminal wealth. The previous computations were made with $\rho_1 = \rho_2 = 10$, putting emphasis on terminal wealth. Figure 9 shows the return distribution in an economy where intertemporal consumption is more important, with $\rho_1 = \rho_2 = 1$. In this case the left tail of the distribution in the VaR economy becomes quite thick. The investors consume a large share of their wealth and therefore have more difficulties meeting the VaR constraint in case of losses: this leads to a more risky investment strategy and hence a heavy left tail.

So far, we have found that the presence of risk managers has a profound impact on the equilibrium stock price and its return distribution. Now we will concentrate on the option market for European call and put contracts. The prices of these options can be computed easily by discounting the payoffs at maturity with the pricing kernel, as in no-arbitrage equation (4). The option payoffs are a function of the equilibrium stock price, which we have derived in closed form. We only have to compute the expectation in equation (4), which can be implemented straightforwardly as the pricing kernel follows a Geometric Brownian motion. Once we have calculated the option prices for a wide range of strike prices, we transform them into implied volatilities with the Black-Scholes formula, in order to facilitate interpretation and comparisons.

Figure 10 shows the implied volatility in economies with a VaR constraint (α = 0%, 1% and 5%) and in the benchmark economy. In the benchmark case we observe a constant implied volatility, following from the Geometric Brownian motion of the stock price and the constant interest rate. In the economies with a VaR constraint a remarkable option smile can be recognised, as often observed in practice. Note that the implied volatility in economies with a VaR constraint is lower than in the benchmark case most of the time. Moreover, as the VaR constraint becomes more strict (ie from α = 5% to α = 0%), the implied volatility decreases. Similar volatility smiles can be observed if we change other parameters of the VaR economy such as $\underline{W}$ and ρ_2. Option prices are generally lower in VaR economies, as a result of the reduced volatility of the stock return distribution.

4. VaR risk management and loss aversion

The equilibrium analysis in the previous section shows that, in general, VaR-based risk management reduces the volatility of stock returns. However, in some cases the VaR constraint adversely increases the probability of extreme losses. The negative consequences of VaR-based risk management are mainly due to the "all or nothing" gambling attitude of the optimal investment strategy in case of losses, which might seem rather unnatural. In this section our aim is to demonstrate that the gambling strategy of a VaR risk manager might not be that unnatural for many investors, as it is closely related to the optimal strategy of loss-agents with the utility function of prospect theory.

Prospect theory was proposed by Kahneman and Tversky (1979) as a descriptive model for decision-making under uncertainty, given the strong violations of the traditional utility paradigm observed in practice. In experiments, Kahneman and Tversky (1979) found that people are concerned about changes in wealth, rather than the level of wealth itself. Moreover, individuals treat gains and losses relative to their reference point differently: the pain of a loss is felt much more strongly than the payoff of an equivalent gain. Furthermore, in the domain of gains people are risk-averse, while they become *risk-seeking* in the domain of losses.

Kahneman and Tversky (1979) quantified these empirical findings in prospect theory: individuals maximise an S-shaped value function (26), which is convex for losses and concave for gains relative to the reference point θ:

$$U_{LA}(x) = \begin{cases} -A(\theta - x)^{\gamma_L}, & \text{for } x \le \theta \\ +B(x-\theta)^{\gamma_G}, & \text{for } x > \theta, \end{cases} \tag{26}$$

where A > 0 and B > 0 to ensure that $U_{LA}(\cdot)$ is an increasing function and $0 < \gamma_L \le 1$, $0 < \gamma_G \le 1$. Moreover, $A > B$ holds in the case of loss aversion. The parameters of the loss-averse utility function were estimated by Tversky and Kahneman (1992) as $\gamma_L = \gamma_G = 0.88$ and $A/B = 2.25$, based on experiments. An illustration of the value function can be found in Figure 11. We denote agents who maximise this value function with $A > B$ as "loss-averse".

Berkelaar and Kouwenberg (2001a) derive the optimal wealth profile of a loss-averse investor, in a similar dynamic economy as in Section 2 of this paper:

Proposition 6

For any state price density following the process (3), the optimal intertemporal consumption policy $c^*_{LA}(t)$ and terminal wealth profiles $W^*_{LA}(T)$ of a loss-averse agent are

$$c^*_{LA}(t) = I(y_{LA}\xi(t)) = (y_{LA}\xi(t))^{1/\gamma - 1}, \tag{27}$$

$$W^*_{LA}(T) = \begin{cases} \theta + \left(\dfrac{y_{LA}\xi(T)}{\rho_{LA}B\gamma_G}\right)^{1/(\gamma_G - 1)} & \text{for } \xi(T) < \overline{\xi}/y_{LA} \\ 0 & \text{for } \xi(T) \ge \overline{\xi}/y_{LA}, \end{cases} \tag{28}$$

where $I(z)$ denotes the inverse of the marginal utility function over consumption $z = U'(x)$, and $\overline{\xi}$ solves $f(\overline{\xi}) = 0$ with

$$f(x) = \frac{1-\gamma_G}{\gamma_G}\left(\frac{1}{x}\right)^{\gamma_G/(1-\gamma_G)}(\rho_{LA}B\gamma_G)^{1/(1-\gamma_G)} - \theta x + \rho_{LA}A\theta^{\gamma_L} \tag{29}$$

and $y_{LA} \ge 0$ is a Lagrange multiplier, satisfying

$$E\left[\int_0^T \xi(s)c^*_{LA}(s)ds + \xi(T)W^*_{LA}(T)\right] = \xi(0)W_{LA}(0). \tag{30}$$

Figure 12 shows the portfolio weights corresponding to this optimal wealth profile for a loss-averse investor that puts more emphasis on terminal wealth than on consumption ($\rho_{LA} = 100$), assuming a constant interest rate r = 5% and a constant Sharpe ratio κ = 0.4. Note that the investment strategy of the loss-averse agent is qualitatively similar to the optimal policy of a VaR constrained agent. The strategy is cautious in good states of the world with a low pricing kernel, and more risky in bad states of the world with a high pricing kernel. Loss-averse agents dislike losses, even if they are very small, and therefore their optimal investment strategy tries to keep wealth above the reference point in good states of the world by investing more wealth in the riskless asset.[12]

Once a loss-averse investor's wealth drops below the reference point, he tries to make up his losses by following a risky investment strategy. The risk-seeking behaviour stems from the convex shape of the utility function of prospect theory below the reference point. Hence, similar to a VaR agent, a loss-averse agent tries to limit losses most of the time, but starts taking risky bets once his wealth drops below the reference point. The optimal investment strategy under a VaR constraint might therefore seem rather natural for loss-averse investors. Or, conversely, one could argue that a VaR constraint imposes a minimum level of "loss aversion" on all investors affected by the regulation.

Not surprisingly, loss-averse agents have a similar impact on asset prices in equilibrium as VaR-constrained agents. We refer to Berkelaar and Kouwenberg (2001b) for a closed-form solution of the asset price in a dynamic economy with loss-averse agents. Figure 13 compares the stock return distribution in an economy with 50% loss-averse agents to an economy with 50% VaR-constrained agents, while Figure 14 shows the volatility smile in the economies. Both figures demonstrate a clear resemblance in prices. We conclude that asset prices in an economy with a VaR constraint correspond qualitatively to prices in an economy with loss-averse investors who put emphasis on terminal wealth.

5. Conclusions

In order to investigate the effect of VaR-based risk management on asset prices in equilibrium, we adopted the continuous-time equilibrium model of Basak and Shapiro (2001). The asset pricing model is modified such that agents maximise their expected utility from intermediate consumption and terminal wealth at the planning horizon. While dynamically optimising their consumption-investment plan, the VaR restriction is imposed on the investors at the planning horizon. The VaR horizon thus coincides with the investment horizon and can be interpreted as a subperiod of the investor's lifetime (eg an official reporting period of 10 days or one year). Within this setup we can ignore two unrealistic price movements that occur in the model of Basak and Shapiro (2001); (1) a jump in prices just after the VaR horizon, and (2) prices dropping to zero at the planning horizon.

We derived the closed-form equilibrium solutions for the asset prices in the model. Our main findings are as follows: the presence of VaR risk managers generally reduces market volatility. However, in very bad states the optimal investment strategy of VaR risk managers is to take a large exposure to stocks, pushing up market risk and creating a hump in the equilibrium price function. In some special cases this strategy can adversely increase the probability of extreme negative returns. This effect is amplified in economies where investors consume a large share of their wealth, in economies where the VaR constraint has a high maximum loss probability α and in economies with a high VaR threshold $\underline{W}$.

We also derived the prices of European options in our economy and found that implied volatility is typically lower in the presence of VaR risk managers. Moreover, VaR risk management creates an implied volatility smile in our economy, as often observed in practice. As a final conclusion, we would like to stress that VaR-based risk management has a stabilising effect on the economy as a whole for most parameter settings in our model. However, it is important to note that the VaR restriction in some cases might worsen catastrophic states that occur with a very small chance, due to the gambling strategy of the VaR risk managers in bad states of the world.

The gambling strategy of VaR-based risk managers in bad states of the world is optimal within a standard dynamic investment model, but might seem rather unnatural for investors in the real world. In the last section of the paper we demonstrated that this gambling strategy in bad states is shared by loss-averse investors, who maximise the utility function of prospect theory (Kahneman and Tversky (1979)). As the optimal investment strategies of loss-averse agents and VaR risk managers are quite similar, it might be relatively easy for the group of loss-averse investors to adopt VaR-based risk management.

[12] See footnote 4.

Appendix

Proof of proposition 1

We refer to Cox and Huang (1989), Karatzas et al (1987) and Pliska (1986) for the optimal consumption and investment policies for agents with power utility. Basak and Shapiro (2001) derive the optimal policies for the portfolio choice problem with an additional VaR constraint.

Proof of proposition 2

This proof can be found in Karatzas et al (1990) and Basak (1995).

Proof of proposition 3

If we substitute the optimal consumption policies (13) into equilibrium relationship (15), then we find:

$$\delta(t) = \lambda(y_1\xi(t))^{-1/\gamma} + (1-\lambda)(y_2\xi(t))^{-1/\gamma} = \left(\lambda y_1^{-1/\gamma} + (1-\lambda)y_2^{-1/\gamma}\right)\xi(t)^{-1/\gamma}, \tag{31}$$

and hence the state price density in equilibrium is

$$\xi(t) = \left(\lambda y_1^{-1/\gamma} + (1-\lambda)y_2^{-1/\gamma}\right)^{\gamma}\delta(t)^{-\gamma} = (v(y_1,y_2)\delta(t))^{-\gamma}. \tag{32}$$

By applying Ito's lemma, we can derive that $\xi(t)$ follows the stochastic process below:

$$\begin{aligned} d\xi(t) &= -\left(\gamma\mu_\delta - 1/2\,\gamma(\gamma+1)\sigma_\delta^2\right)(v(y_1,y_2)\delta(t))^{-\gamma}dt - \gamma\sigma_\delta(v(y_1,y_2)\delta(t))^{-\gamma}dB(t) \\ &= -\left(\gamma\mu_\delta - 1/2\,\gamma(\gamma+1)\sigma_\delta^2\right)\xi(t)dt - \gamma\sigma_\delta\xi(t)dB(t). \end{aligned} \tag{33}$$

Equating the processes (3) and (33), we can determine the constant interest rate r and the constant market price of risk κ.

Proof of proposition 4

The equilibrium stock price in an economy with unregulated agents is a special case of proposition 5 with $\lambda = 1$ (see proof below). The drift rate μ(t) and volatility σ(t) of the process can be derived by applying Ito's lemma to the stock price.

Proof of proposition 5

The price of the risky asset can be derived from the third equilibrium condition in (10):

$$S_1(t) = \lambda W_1^*(t) + (1-\lambda)W_2^*(t). \tag{34}$$

Given the optimal policies of a normal agent and the process for $\delta(t)$, we can derive:

$$\begin{aligned} W_1^*(t) &= \frac{1}{\xi(t)}E_t\left[\int_t^T \xi(s)c_1^*(s)ds + \xi(T)W_1^*(T)\right] \\ &= \frac{1}{\xi(t)}\left(E_t\left[\int_t^T \xi(s)(y_1\xi(s))^{-1/\gamma}ds\right] + E_t\left[\xi(T)(y_1\xi(T)/\rho_1)^{-1/\gamma}\right]\right) \\ &= \frac{1}{\xi(t)}y_1^{-1/\gamma}v(y_1,y_2)^{1-\gamma}\left(E_t\left[\int_t^T \delta(t)^{1-\gamma}ds\right] + \rho_1^{1/\gamma}E_t\left[\delta(T)^{1-\gamma}\right]\right) \\ &= y_1^{-1/\gamma}v(y_1,y_2)\delta(t)^{\gamma}\left(a(t)\delta(t)^{1-\gamma} + \rho_1^{1/\gamma}e^{\eta(T-t)}\delta(t)^{1-\gamma}\right) \\ &= y_1^{-1/\gamma}v(y_1,y_2)\left(a(t) + \rho_1^{1/\gamma}e^{\eta(T-t)}\right)\delta(t). \end{aligned} \tag{35}$$

Similarly, we find for the VaR constrained agents:

$$W_2^*(t) = \frac{1}{\xi(t)} E_t\left[\int_t^T \xi(s) c_2^*(s) ds + \xi(T) W_2^*(T)\right] \tag{36}$$

$$= y_2^{-1/\gamma} v(y_1, y_2) a(t) \delta(t) + \frac{1}{\xi(t)} E_t\left[\xi(T) W_2^*(T)\right]$$

$$\begin{aligned} &= y_2^{-1/\gamma} v(y_1, y_2) a(t) \delta(t) \\ &\quad + \frac{1}{\xi(t)} E_t\left[\xi(T)(y_2 \xi(T)/\rho_2)^{-1/\gamma} 1_{\{\xi(T) < \underline{\xi} \cup \xi(T) \geq \overline{\xi}\}} + \xi(T) \underline{W} 1_{\{\underline{\xi} \leq \xi(T) < \overline{\xi}\}}\right] \end{aligned} \tag{37}$$

$$\begin{aligned} &= y_2^{-1/\gamma} v(y_1, y_2) a(t) \delta(t) \\ &\quad + (y_2, \rho_2)^{-1/\gamma} v(y_1, y_2) e^{\Gamma(t)} \delta(t)\left(1 - N\left(-d\left(\delta(t), t, \underline{\xi}\right)\right) + N\left(-d\left(\delta(t), t, \overline{\xi}\right)\right)\right) \\ &\quad + \underline{W} e^{-r(T-t)}\left(N\left(-f\left(\delta(t), t, \underline{\xi}\right)\right) - N\left(-f\left(\delta(t), t, \overline{\xi}\right)\right)\right), \end{aligned} \tag{38}$$

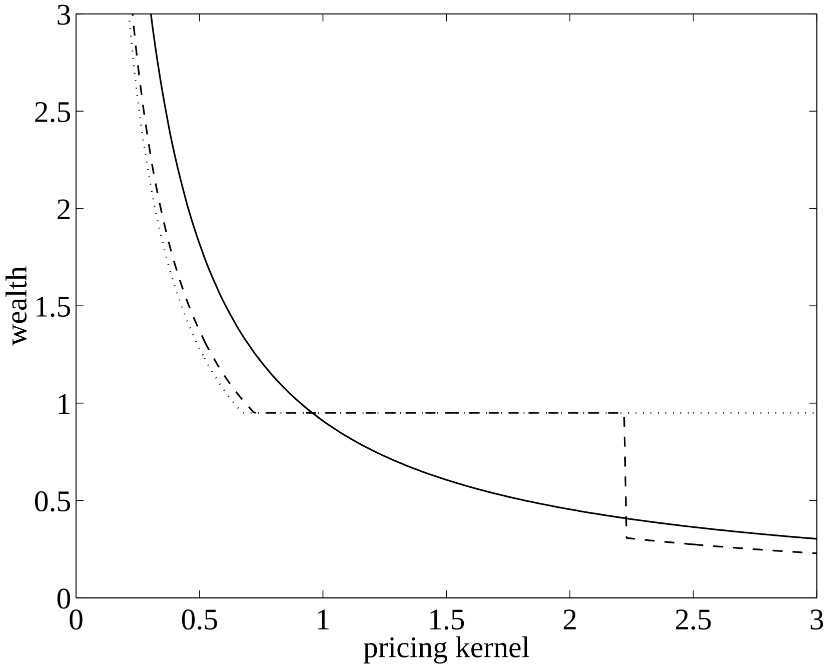
where $1_{\{\xi(T) \geq \overline{\xi}\}}$ denotes the indicator function. Finally, by substituting (35) and (36) into (34), we obtain the equilibrium price.

Proof of proposition 6

This proof can be found in Berkelaar and Kouwenberg (2001b).

Figure 1

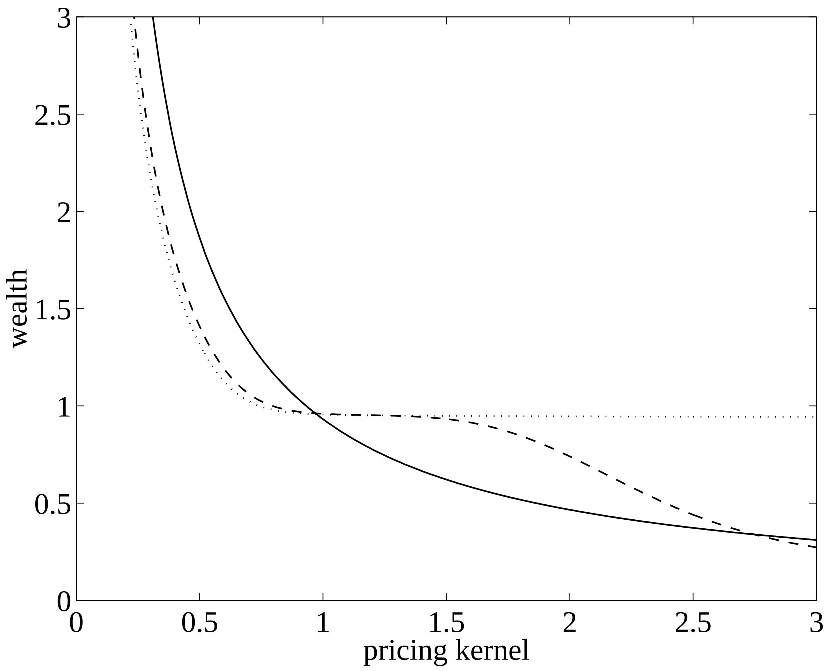
Wealth at time *T* = 1



This figure shows the optimal wealth at time $T = 1$, for the unregulated agent (solid line), the VaR risk manager (dashed line) and the portfolio insurer (dotted line).

Figure 2

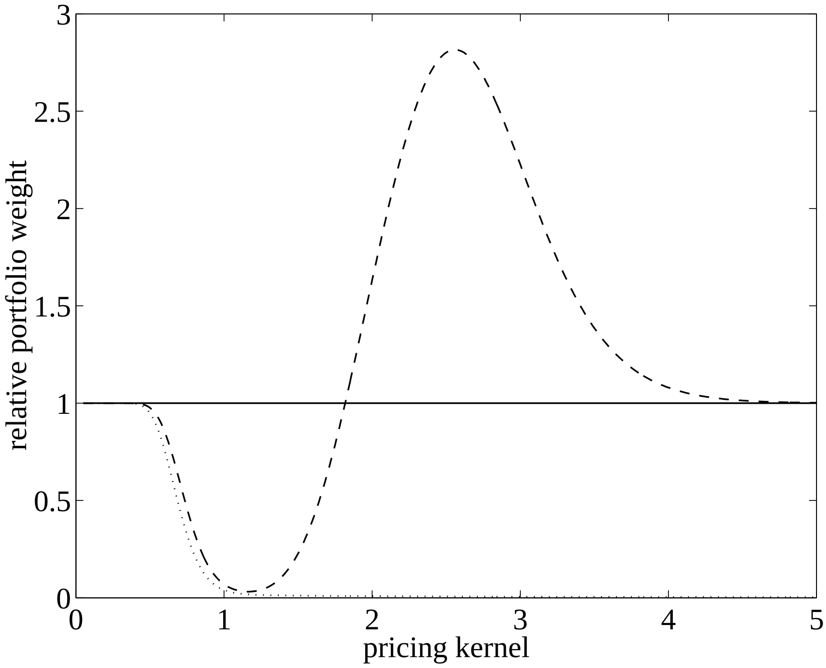
Wealth at time *t* = 0.75



This figure shows the optimal wealth at time $t = 0.75$, for the unregulated agent (solid line), the VaR risk manager (dashed line) and the portfolio insurer (dotted line).

Figure 3

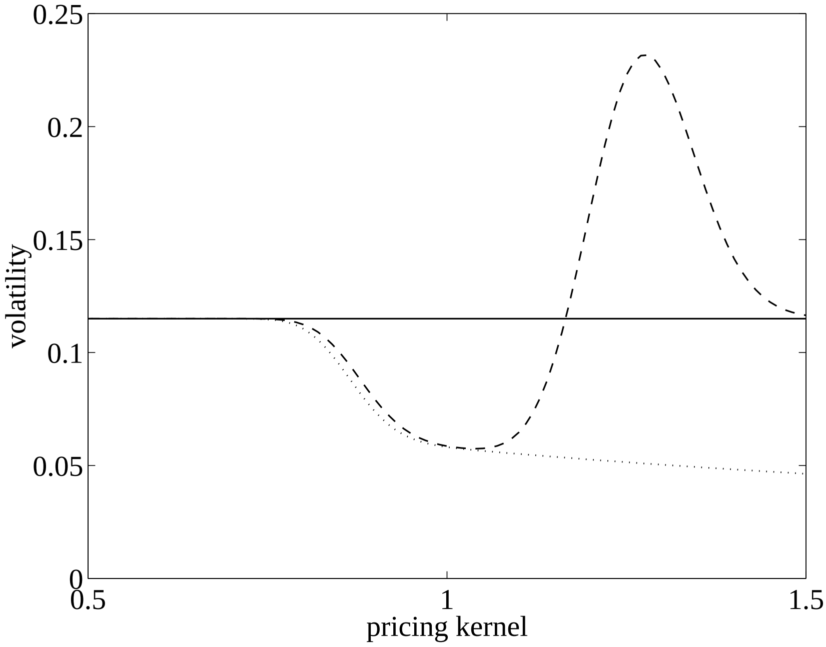
Relative portfolio weight of stocks



This figure shows the portfolio weight of stocks relative to the unregulated agent at time $t = 0.75$, for the unregulated agent (solid line), the VaR risk manager (dashed line) and the portfolio insurer (dotted line).

Figure 4

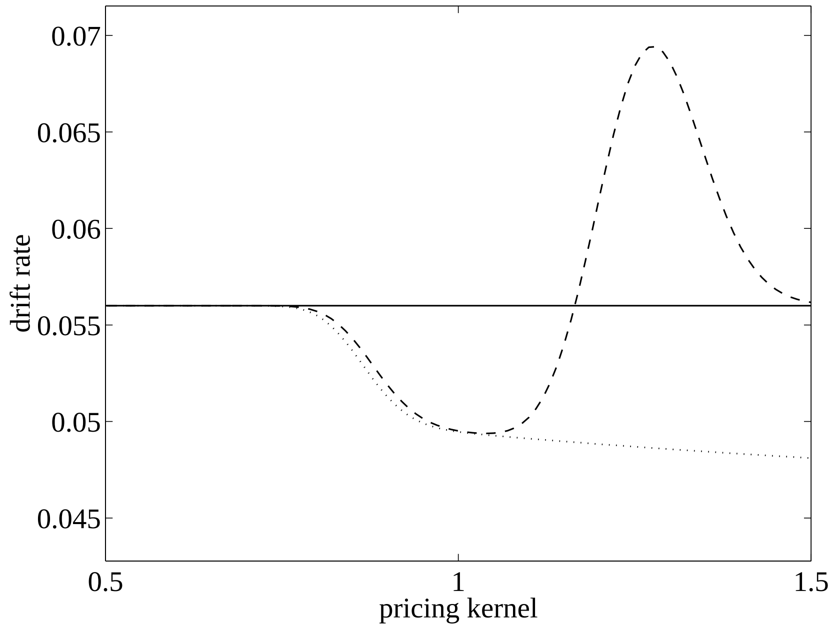
Volatility in equilibrium



This figure shows the volatility of the stock price process in equilibrium at time $t = 0.75$, for the economy with unregulated agents (solid line), the economy with 50% VaR risk managers (dashed line) and the economy with 50% portfolio insurers (dotted line).

Figure 5

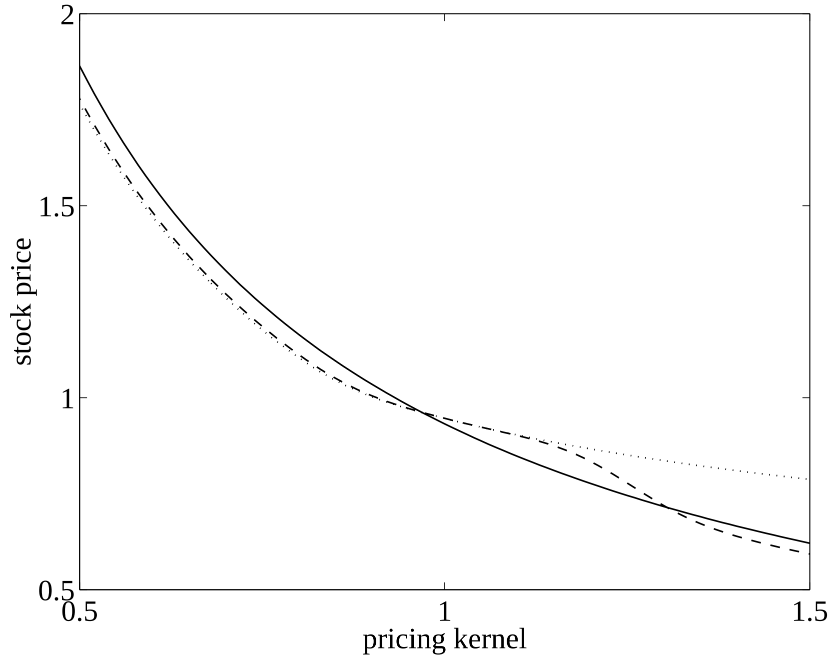
Drift rate in equilibrium



This figure shows the drift rate of the stock price process in equilibrium at time t = 0.75, for the economy with unregulated agents (solid line), the economy with 50% VaR risk managers (dashed line) and the economy with 50% portfolio insurers (dotted line).

Figure 6

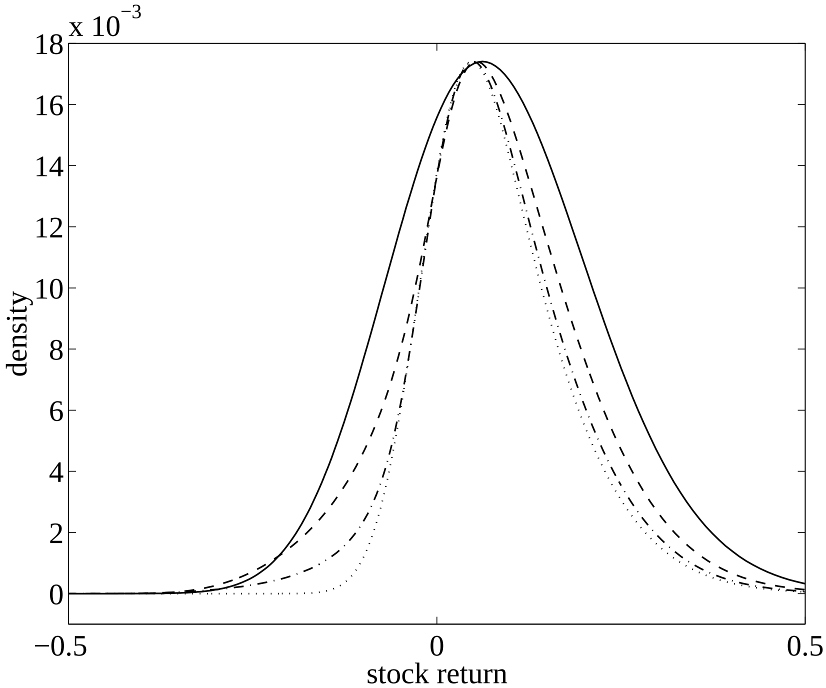
Stock price in equilibrium



This figure shows the stock price in equilibrium at time t = 0.75, for the economy with unregulated agents (solid line), the economy with 50% VaR risk managers (dashed line) and the economy with 50% portfolio insurers (dotted line).

Figure 7

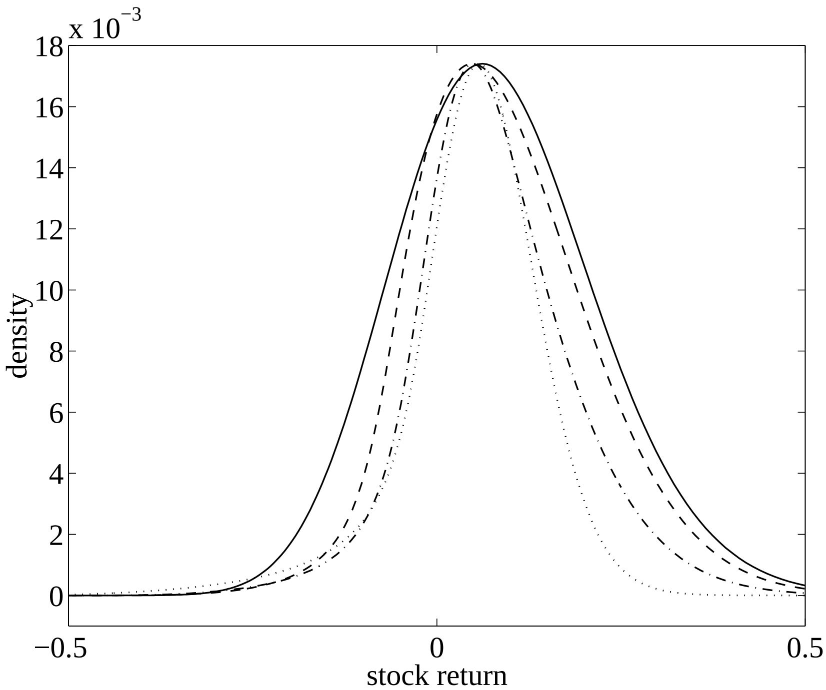
Stock return distribution in equilibrium



This figure shows the return distribution of stocks in equilibrium at time *t* = 0.75, for the economy with unregulated agents (solid line), the economy with 50% VaR risk managers with α = 5% (dashed line), the economy with 50% VaR risk managers with α = 1% (dashed-dotted line), and the economy with 50% portfolio insurers (dotted line).

Figure 8

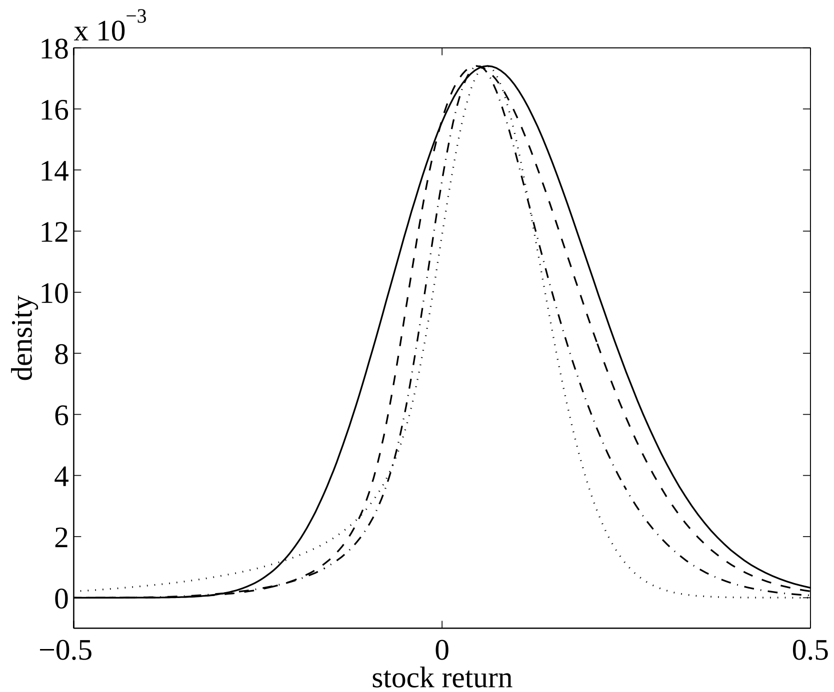
Stock return distribution in equilibrium



This figure shows the return distribution of stocks in equilibrium at time *t* = 0.75, for the economy with unregulated agents (solid line), the economy with 50% VaR risk managers with *W* = 0.90% (dashed line), the economy with 50% VaR risk managers with *W* = 0.95% (dashed-dotted line), and the economy with 50% VaR risk managers with *W* = 0.99% (dotted line).

Figure 9

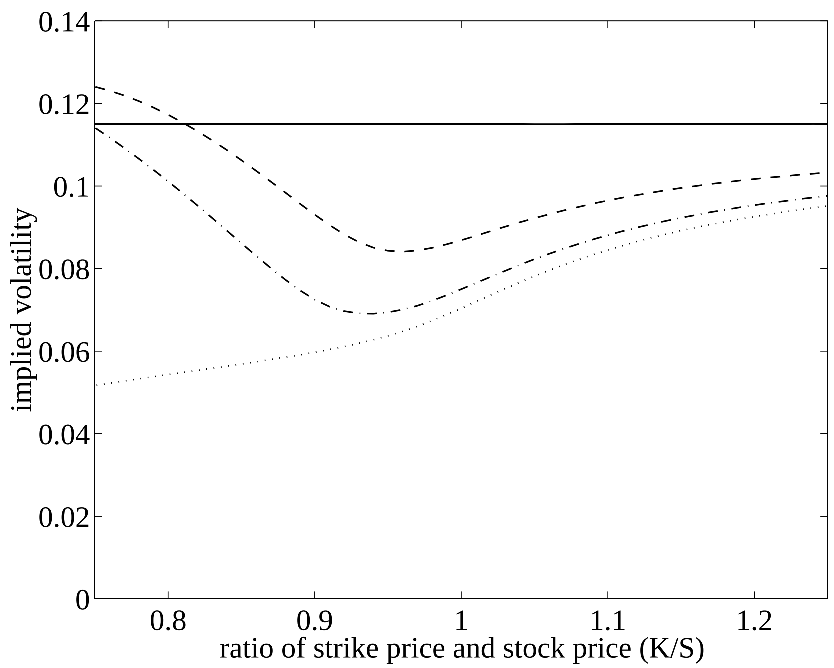
Stock return distribution in equilibrium



This figure shows the return distribution of stocks in equilibrium at time t = 0.75, for the economy with unregulated agents (solid line), the economy with 50% VaR risk managers with $\rho_1 = \rho_2 = 20$ (dashed line), the economy with 50% VaR risk managers with $\rho_1 = \rho_2 = 10$ (dashed-dotted line), and the economy with 50% VaR risk managers with $\rho_1 = \rho_2 = 1$ (dotted line).

Figure 10

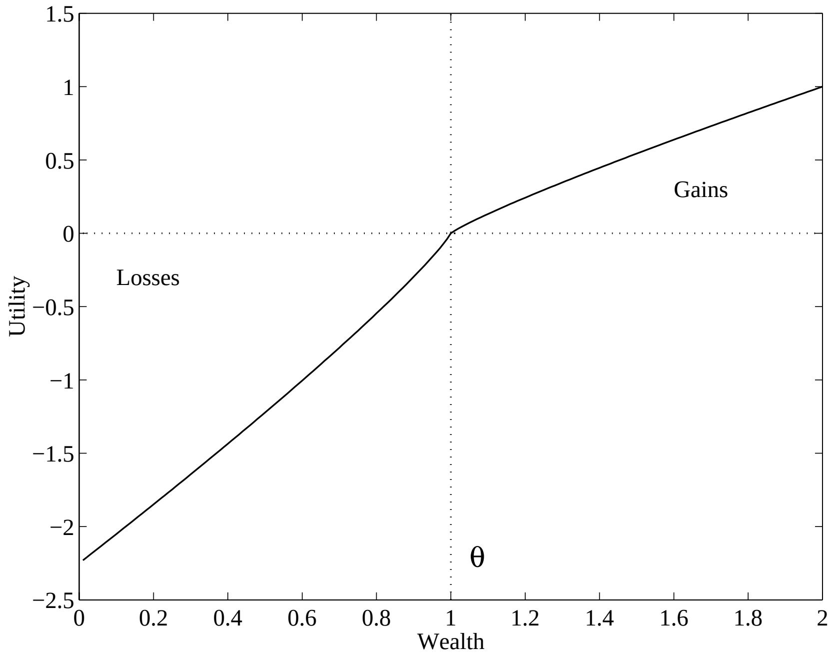
Implied volatility of equilibrium option prices



This figure shows the implied volatility of option prices at time t = 0. The call and put options are of the European type, with maturity at time t = 0.75. The figure displays equilibrium implied volatilities in the economy with unregulated agents (solid line), the economy with 50% VaR risk managers with α = 5% (dashed line), the economy with 50% VaR risk managers with α = 1% (dashed-dotted line), and the economy with 50% portfolio insurers (dotted line).

Figure 11

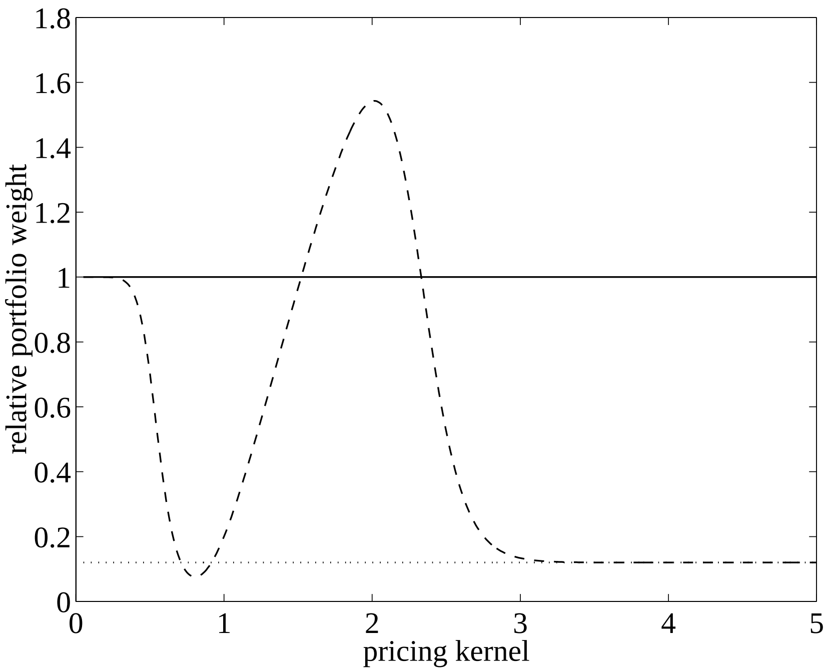
Utility function of prospect theory



This figure shows the utility function of prospect theory, with parameter values $\gamma_1 = \gamma_2 = 0.88$, $A = 2.25$, $B = 1.0$ and $\theta = 1.0$.

Figure 12

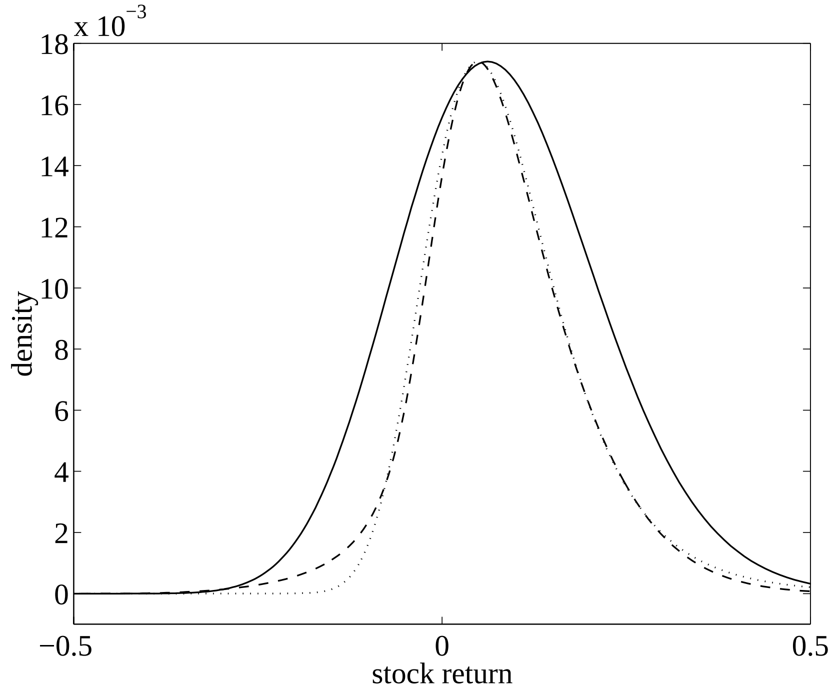
Relative portfolio weight of stocks



This figure shows the portfolio weight of stocks relative to an unregulated agent with $\gamma = 0.88$ at time $t = 0.75$ (solid line), for a loss-averse agent with $\rho_{LA} = 100$ (dashed line) and for an unregulated agent with $\gamma = 0$ (dotted line).

Figure 13

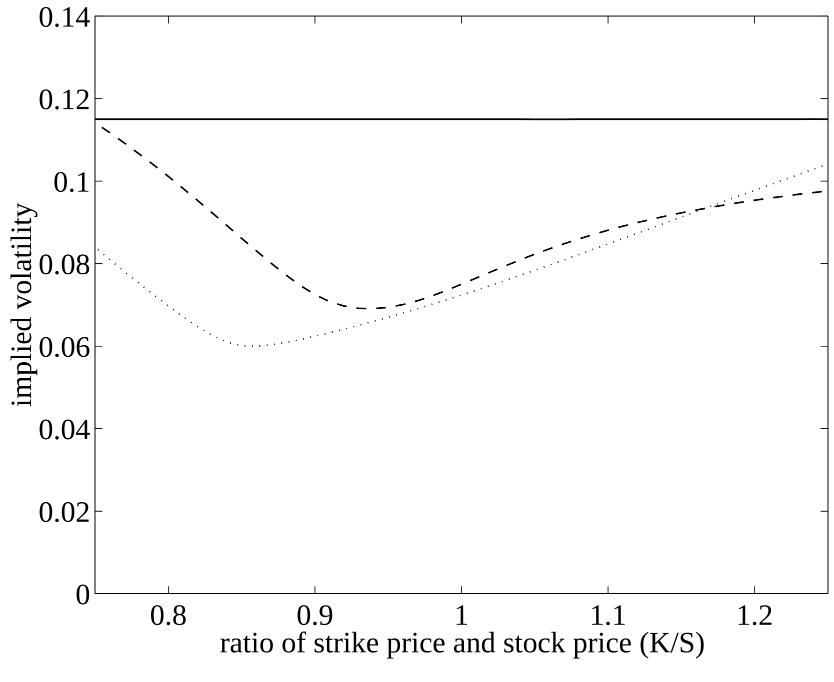
Stock return distribution in equilibrium



This figure shows the return distribution of stocks in equilibrium at time t = 0.75, for the economy with unregulated agents (solid line), the economy with 50% VaR risk managers with α = 1% (dashed line), and the economy with 50% loss-averse agents with ρ_{LA} = 100 (dotted line).

Figure 14

Implied volatility of equilibrium option prices



This figure shows the implied volatility of option prices at time t = 0. The call and put options are of the European type, with maturity at time 0.75. The figure displays equilibrium implied volatilities in the economy with unregulated agents (solid line), the economy with 50% VaR risk managers with α = 1% (dashed line), and the economy with 50% loss-averse agents with ρ_{LA} = 100 (dotted line).

Bibliography

Artzner, P, F Delbaen, J-M Eber and D Heath (1999): "Coherent measures of risk", *Mathematical Finance* 9, 203-28.

Basak, S (1995): "A general equilibrium model of portfolio insurance", *Review of Financial Studies* 8, 1059-90.

Basak, S and A Shapiro (2001): "Value-at-risk based risk management: optimal policies and asset prices", *Review of Financial Studies* 14, 371-405.

Berkelaar, A and R Kouwenberg (2001a): "Optimal portfolio choice under loss aversion", Working Paper.

——— (2001b): "From boom til bust: how loss aversion affects asset prices", Working Paper.

Black, F and M Scholes (1973): "The pricing of options and corporate liabilities", *Journal of Political Economy* 3, 637-54.

Cox, J and C Huang (1989): "Optimum consumption and portfolio policies when asset prices follow a diffusion process", *Journal of Economic Theory* 49, 33-83.

DeMarzo, P M (1988): "An extension of the Modigliani-Miller theorem to stochastic economies with incomplete markets and interdependent securities", *Journal of Economic Theory* 45, 353-69.

Grossman, S J and J-L Vila (1989): "Portfolio insurance in complete markets: a note", *Journal of Business* 62, 473-6.

Grossman, S J and Z Zhou (1996): "Equilibrium analysis of portfolio insurance", *Journal of Finance* 51, 1379-403.

Kahneman, D H and A Tversky (1979): "Prospect theory: an analysis of decision under risk", *Econometrica* 47, 263-90.

Karatzas, I, J Lehoczky and S Shreve (1987): "Optimal portfolio and consumption decisions for a small investor on a finite horizon", *SIAM Journal on Control and Optimization* 25, 1157-86.

——— (1990): "Existence and uniqueness of multi-agent equilibrium in a stochastic, dynamic consumption/investment model", *Mathematics of Operations Research* 15, 80-128.

Karatzas, I and S Shreve (1998): *Methods of mathematical finance,* Springer Verlag, New York.

Leland, H (1998): "Agency costs, risk management, and capital structure", *Journal of Finance* 53, 1213-44.

Lucas, R E Jr (1978): "Asset pricing in an exchange economy", *Econometrica* 46, 1429-45.

Merton, R (1969): "Lifetime portfolio selection under uncertainty: the continuous-time case", *Review of Economics and Statistics* 51, 247-57.

Modigliani, F and M H Miller (1958): "The cost of capital, corporation finance, and the theory of investment: Reply", *American Economic Review* 49, 655-69.

Pliska, S R (1986): "A Stochastic Calculus model of continuous trading: optimal portfolios", *Mathematics of Operations Research* 11, 371-82.

Stiglitz, J E (1969a): "Theory of innovation: discussion", *American Economic Review* 59, 46-9.

——— (1969b): "A re-examination of the Modigliani-Miller theorem", *American Economic Review* 59, 784-93.

——— (1974): "On the irrelevance of corporate financial policy", *American Economic Review* 64, 851-66.

Tversky, A and D H Kahneman (1992): "Advances in prospect theory: cumulative representation of uncertainty", *Journal of Risk and Uncertainty* 5, 297-323.

Annexes

Annex 1
Conference programme

Opening address

Andrew D Crockett (BIS).

Session 1: Banking stability

Chair: Kazuo Ueda (Bank of Japan).

Franklin Allen and Douglas Gale:
Financial fragility.

Elena Carletti, Philipp Hartmann and Giancarlo Spagnolo:
Bank mergers, competition and financial stability.

Mariassunta Giannetti:
On the causes of overlending: are guarantees on deposits the culprit?

Luncheon address

Yutaka Yamaguchi (Bank of Japan).

Session 2: Market contagion

Chair: Francesco Papadia (ECB).

Discussant: Takatoshi Ito (Hitotsubashi University).

Mardi Dungey, Renée Fry, Brenda González-Hermosillo and Vance Martin:
International contagion effects from the Russian crisis and the LTCM near collapse.

Graciela Kaminsky and Carmen Reinhart:
The centre and the periphery: the globalisation of financial turmoil.

Marco Cipriani and Antonio Guarino:
Herd behaviour and contagion in financial markets.

Session 3: Liquidity I

Chair: Peter Praet (National Bank of Belgium).

Discussant: Harry Stordel (Credit Suisse Group).

Ben Cohen and Hyun Shin:
Positive feedback trading under stress: evidence from the US Treasury market.

Matthew Pritsker:
Large investors: implications for equilibrium asset returns, shock absorption and liquidity.

David Tien:
Hedging demand and foreign exchange risk premia.

Session 4: Liquidity II

Chair: José Viñals (Bank of Spain).

Discussant: Frank Roncey (BNP Paribas SA).

Jón Daníelsson and Richard Payne:
Measuring and explaining liquidity on an electronic limit order book: evidence from Reuters D2000-2.

Paul Harrison:
Market liquidity in times of stress in the corporate bond market.

Jim Wong and Laurence Fung:
Liquidity of the Hong Kong stock market since the Asian crisis.

Session 5: Risk Measurement

Chair: Christine Cumming (Federal Reserve Bank of New York).

Discussant: Paul Shotton (JP Morgan Chase).

Torben Andersen, Tim Bollerslev, Francis Diebold and Paul Labys:
Modelling and forecasting realised volatility.

Yasuhiro Yamai and Toshinao Yoshiba:
Comparative analyses of expected shortfall and VaR under market stress.

André Lucas, Pieter Klaassen, Peter Spreij and Stefan Straetmans:
Tail behaviour of credit loss distributions for general latent factor models.

Luncheon address

Tommaso Padoa-Schioppa (ECB).

Session 6: Market behaviour and monitoring

Chair: William White (BIS).

Discussant: Timothy Wilson (Morgan Stanley).

Martin Blåvarg and Patrick Nimander:
The Riksbank's approach to systemic risk by monitoring counterparty exposures in the interbank market.

Reint Gropp, Jukka Vesala and Giuseppe Vulpes:
Equity and bond market signals as leading indicators of bank fragility.

Arjan Berkelaar, Phornchanok Cumperayot and Roy Kouwenberg:
The effect of VaR-based risk management on asset prices and the volatility smile.

Annex 2
List of conference participants

Name/affiliation

Franklin Allen,
University of Pennsylvania

Terry Allen,
Financial Services Authority

Naohiko Baba,
Bank of Japan

Jeremy Barson,
Bank for International Settlements

Ric Battellino,
Reserve Bank of Australia

Martin Blåvarg,
Sveriges Riksbank

Raymond Bo,
Hong Kong Monetary Authority

Claudio Borio,
Bank for International Settlements

Alex Bowen,
Bank of England

Wolfgang Bühler,
University of Mannheim

Inês Cabral,
European Central Bank

Elena Carletti,
University of Mannheim

Maria Caspar,
Bank for International Settlements

Marco Cipriani,
New York University

Benjamin Cohen,
Bank for International Settlements

Andrew D Crockett,
Bank for International Settlements

Christine Cumming,
Federal Reserve Bank of New York

Phornchanok Cumperayot,
Erasmus University Rotterdam

Jón Daníelsson,
London School of Economics

Sally Davies,
Federal Reserve Board

Dietrich Domanski,
Bank for International Settlements

Name/affiliation

Darrell Duffie,
Stanford University

Mardi Dungey,
Australian National University

John Eatwell,
Queen's College, Cambridge

Ingo Fender,
Bank for International Settlements

Renato Filosa,
Bank for International Settlements

Allen Frankel,
Bank for International Settlements

Stefan Gerlach,
Hong Kong Monetary Authority

Mariassunta Giannetti,
Stockholm School of Economics

Michael Gibson,
Federal Reserve Board

Norbert Goffinet,
Central Bank of Luxembourg

Brenda González-Hermosillo,
International Monetary Fund

Charles Goodhart,
The London School of Economics

Stéphane Guéné,
European Central Bank

Antonio Guarino,
New York University

Jacob Gyntelberg,
Bank for International Settlements

Paul Harrison,
Federal Reserve Board

Philipp Hartmann,
European Central Bank

Marc Henrard,
Bank for International Settlements

Glenn Hoggarth,
Bank of England

Akira Ieda,
Bank of Japan

Tetsuya Inoue,
Bank of Japan

Takatoshi Ito,
Hitotsubashi University

Marianne Johnson,
Bank of Canada

Eui Jung Chang,
Central Bank of Brazil

Graciela L Kaminsky,
George Washington University

George G Kaufman,
Loyola University Chicago

Nobuhiro Kiyotaki,
London School of Economics

Roy Kouwenberg,
AEGON Asset Management

Myron L Kwast,
Federal Reserve Board

Paul Labys,
Charles River Associates, Inc

Jae-Ouk Lee,
The Bank of Korea

Lim Phang Hong,
Monetary Authority of Singapore

Jose Lopez,
Federal Reserve Bank of San Francisco

André Lucas,
Vrije Universiteit

Robert McCauley,
Bank for International Settlements

Jean-Pierre Matt,
Bank for International Settlements

Anne McKiernan,
Bank of Ireland

Janet Mitchell,
National Bank of Belgium

Philippe Molitor,
European Central Bank

Dewet Moser,
Swiss National Bank

Naruki Mori,
Bank of Japan

Takahiro Nagata,
Japanese Financial Services Agency

Hiroshi Nakaso,
Bank of Japan

Yoshinori Nakata,
Bank for International Settlements

Uwe Neumann,
Bank for International Settlements

Stefan Niessner,
Deutsche Bundesbank

Patrick Nimander,
Sveriges Riksbank

Werner Nimmerrichter,
Deutsche Bundesbank

Tommaso Padoa-Schioppa,
European Central Bank

Francesco Papadia,
European Central Bank

Sissy Papagiannidou,
Bank of Greece

Peter Praet,
National Bank of Belgium

Matthew G Pritsker,
Federal Reserve Board

Alexane Pundek Rocha,
Central Bank of Brazil

Frank Roncey,
BNP Paribas

Robert Wayne St Clair,
Monetary Authority of Singapore

Max Schieler,
UBS AG

Tokiko Shimizu,
Bank of Japan

Hyun Song Shin,
London School of Economics

Paul Shotton,
JP Morgan Chase & Co

Kenneth J Singleton,
Stanford University

Harry Stordel,
Credit Suisse Group

David Tien,
UC Berkeley

Josef Tosovský,
Bank for International Settlements

Kostas Tsatsaronis,
Bank for International Settlements

Philip Turner,
Bank for International Settlements

Kazuo Ueda,
Bank of Japan

Jens Ulrich,
Bank for International Settlements

L M T van Velden,
Netherlands Bank

Jukka Vesala,
European Central Bank

José Viñals,
Bank of Spain

Yong-Gie Wang,
The Bank of Korea

Beatrice Weder,
University of Mainz

William R White,
Bank for International Settlements

Timothy S Wilson,
Morgan Stanley

Jim Wong,
Hong Kong Monetary Authority

Yutaka Yamaguchi,
Bank of Japan

Yasuhiro Yamai,
Bank of Japan

Toshinao Yoshiba,
Bank of Japan