Financial Stability Institute

FSI Award 2006 Winning Paper

Stability of a "through-the-cycle" rating system during a financial crisis

Verónica Vallés,
Central Bank of Argentina

September 2006

BANK FOR INTERNATIONAL SETTLEMENTS

The views expressed in this paper are those of the author and not necessarily the views of the Financial Stability Institute or the Bank for International Settlements.

Copies of publications are available from:

Financial Stability Institute
Bank for International Settlements
CH-4002 Basel, Switzerland

E-mail: fsi@bis.org

Tel: +41 61 280 9989
Fax: +41 61 280 9100 and +41 61 280 8100

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

ISSN 1684-7180

Foreword

The Financial Stability Institute is once again pleased to publish the winning FSI Award paper. This award, announced every two years at the time of the International Conference of Banking Supervisors, was established to encourage thought and research on issues relevant to banking supervisors. This year, 13 papers were received on a variety of topics written by supervisors from around the world.

A jury of highly qualified individuals chose the winning paper. The group was chaired by Mr Malcolm Knight, General Manager of the Bank for International Settlements. It also included: Mr Ryozo Himino, Secretary General of the Basel Committee on Banking Supervision; Mrs Ruth de Krivoy, former President of the Central Bank of Venezuela; Mr Myron Kwast, Associate Director of the Division of Research & Statistics, US Federal Reserve Board, and chair of the Basel Committee's Research Task Force; and Mrs Joan De Silva, Director of the Bank Supervision Department at the Central Bank of Sri Lanka.

The jury members and the FSI are pleased to announce that Ms Verónica Vallés of the Central Bank of Argentina has been selected as the winner of the 2006 FSI Award. Ms Vallés' paper looks into the construction of a "through-the-cycle" rating system to assess credit risk in a developing country that has faced a major economic crisis. The author highlights the problems discovered during her research and offers various solutions. She concludes that economic instability in a country does not preclude its banks from using IRB systems under Basel II.

Congratulations to Ms Vallés and the other supervisors who submitted their work for consideration. Their interest in analysing and potentially improving supervisory methods bodes well for the future.

Josef Tošovský
Chairman
Financial Stability Institute
September 2006

Contents

Introduction[1]

The aim of this paper is to research the construction of a "through-the-cycle" (TTC) rating system to assess credit risk in a developing country that has faced a major economic crisis. The country analysed was involved in a severe macroeconomic crisis that spread throughout the financial system.

Constructing a rating system involves estimating a credit scoring model and using the estimated scores to construct risk categories. Furthermore, a TTC rating system uses specific and dynamic information on obligors to assess the credit quality of borrowers. It remains relatively stable in business cycles as stressed scenarios have been considered. But in the developing country analysed, the macroeconomic crisis influenced obligor payment behaviour and therefore affected the rating system. Some problems were encountered in building a stable TTC rating system. The unstable conditions of this particular developing country pose difficulties for constructing rating systems that would be shared by other emerging economies.

The scoring model was constructed estimating a panel data model using public credit registry information on the country's corporate debtors. The database used for the panel regression followed the same debtors in the financial system for a five-year period. The scores obtained from the panel regression were the inputs of the rating system that grouped obligors in different risk categories. Achieving stable risk categories was quite problematic, as the obligors' performance in crisis years was completely different from the one observed in stable ones.

---

[1] The views expressed herein are solely those of the author and do not necessarily represent official policies, statements or views of the Central Bank of Argentina. I would like to thank Cristina Pailhé for her helpful comments and Matías Gutiérrez for his useful advice.

As a matter of fact, the annual frequency of defaults increased from 13% to 30% during the year of the crisis.

Traditionally, credit institutions decided whether or not to grant credit to a particular individual based on human judgment and historical experience about the default risk. However, sophisticated statistical credit scoring models were recently developed to aid the credit granting decision. They are used to estimate the probability of default (PD) using predictor variables taking into account the characteristics and financial situation of applicants. The decision to accept or reject candidates can be taken after comparing estimated PDs with a suitable threshold. These models are also used in the "ongoing" process of the loan to estimate its likelihood of default.

In June 2004, the Basel Committee on Banking Supervision published the *International convergence of capital measurement and capital standards: a revised framework* (Basel II). One of its main objectives is to promote the adoption of stronger risk management practices by the banking industry. An important innovation of the Revised Framework is the possibility of using internal rating systems as inputs for capital calculations after they have met minimum requirements set out in the document.

The Revised Framework considers that human judgment should be used in the decision to grant loans but highlights the necessity of establishing a formal methodology to rate obligors and to estimate the associated PDs. Thus, it describes methodologies for banks to construct their internal ratings-based (IRB) systems. Banks may use IRB systems to calculate regulatory capital requirements but also as the basis for internal risk measures; so they will use these risk measures for pricing, managing portfolio exposures and establishing reserves. It is important that IRB systems should accurately discriminate between bad and good obligors, those that have higher and lower PD. The accuracy of the estimated PDs and the structure of the rating system would influence capital requirements. This is the reason for focusing on the estimation of a credit scoring model and the construction of a rating system in this document.

The risk measures used to calculate capital requirements are the probability of default (PD), loss-given-default (LGD), exposure at default (EAD) and effective maturity (M). There are two IRB approaches, foundation and advanced. Under both approaches, banks have to provide their own estimates of PD subject to minimum requirements. The Revised Framework specifies that all banks using IRB approaches must estimate a PD for each risk category of the rating system distinguishing between corporate, sovereign and bank exposures.

The Revised Framework highlights that estimated PDs must be a long-run average of one-year PDs for borrowers in each category of the rating system. The recent document published by the Basel Committee on Banking Supervision[2] describes different types of rating systems. The rating system can be calculated with information of one period (one year) as a "point-in-time" (PIT) rating system or, in line with the Revised Framework, it can be calculated with information of a longer period, that is, a "through-the-cycle" (TTC) rating system. The latter rating system would consider long-run estimations of the PDs. The Revised Framework explicitly points out that the length of the underlying historical observation period used to calculate PDs must be at least five years.

The present exercise is an empirical research that constructs a rating system for corporate obligors in the entire financial system registered at the Public Credit Registry of Borrowers (PCRB). We are aware that this exercise is different from an IRB system constructed by a bank that has access to detailed information on borrowers. However, the intention is to construct a broader rating system with information on obligors in a crisis to highlight the problems that would be confronted in the validation process of such an extremely stressed situation.

[2] BCBS (2005).

The paper proceeds as follows. Section 1 briefly describes differences between a PIT and a TTC rating system. Section 2 describes the database used in the empirical estimation and presents the macroeconomic variables to be considered. Section 3 calculates the panel model and the score of the rating system that is validated in Section 4. Section 5 presents the rating system for two years that have a similar percentage of default rates and Section 6 highlights the importance of having a database with the financial history of obligors. Finally, the conclusions show problems and suggestions discovered during the research.

1. "Point-in-time" (PIT) and "through-the-cycle" (TTC) rating systems

The IRB approach requires reporting an individual score for each obligor and an individual estimated PD.[3] These are the inputs for constructing "risk buckets" or "risk categories". Obligors that share the same credit quality must be assigned to the same risk bucket. After grouping obligors in risk buckets, a pooled PD of the bucket must be calculated considering that it has to represent the risk of obligors within the group. This is basically a rating system. One important task is to establish the limit scores of risk buckets. The risk buckets' delimitation could be based on a statistical model, on experts' judgment or on both. The present model is going to be constructed based on scores from a panel model exclusively.

The Revised Framework establishes that a borrower's score must represent the bank's assessment of its ability and willingness to comply with the contract terms despite adverse economic conditions. This means that the bank should not just

---

[3] In the case of a logit or probit model, the score is directly related to the PD of each obligor.

rely on present estimations of the PD but should also calculate PDs in stress scenarios with bad economic conditions. The range of economic conditions considered should be consistent with current conditions and those that are likely to occur over a business cycle in a particular country, region or activity. The PDs that incorporate stress scenarios of the business cycle are named “stressed PDs” and the PDs for a definite period of time are the “unstressed PDs”. The unstressed PDs will change with economic conditions while stressed PDs will be relatively stable in economic cycles. The main idea is that stressed PDs are “cyclically neutral” - they move as obligors’ particular conditions change but they do not respond to changes in overall economic conditions.

A rating system that remains relatively constant through different business conditions is a “through-the-cycle” (TTC) rating system whilst a rating system that changes period by period is a “point-in-time” (PIT) rating system. Obligors in the same risk category of a PIT rating system would share similar unstressed PDs, and obligors in a risk category of a TTC rating system would share similar stressed PDs. Thus, the characteristics of PDs associated with each risk category are determined by the underlying rating system and the type of information used.

The information needed to forecast the defaults can be aggregate information, which typically includes macroeconomic variables such as GDP growth rates, exchange rates and interest rates, and specific obligor information that includes characteristics of and relevant financial information on obligors. A TTC score should take into consideration specific obligor characteristics plus macroeconomic conditions, but a PIT score would be based mainly on current information on obligors.

2. Data

The information used in the present document was taken from a Public Credit Registry of Borrowers (PCRB) of a developing country. This database contains information on loans, claims stemming from financial intermediation, leases and other claims, as well as contingent claims (guarantees, agreed overdrafts in checking accounts and other preagreed lines, etc). For each obligor with each institution, the database records the tax identification number, name, whether the obligor is an individual or corporation/institution, outstanding claims in each credit line, interest rate, maturity, guarantees and provisions. The information for some variables, such as interest rate and maturity, shows low quality.

At the origination of a credit, all debtors must be rated into five grades considering the likelihood of honouring the contractual terms of the claim, on the basis of an individual assessment of the future financial situation. A provisioning percentage is established for each rating grade. From category 1 to category 5, there is an increase in the perceived risk. Each month, the bank must inform the PCRB of a rating grade for each borrower considering the evolution of credit conditions.

The PCRB has a variable that makes it possible to identify the type of obligor (corporate or retail). In particular, this document constructs a rating system for corporate debtors, those with the largest debt amounts (currently, one of the regulatory requirements for being classified as a corporate obligor is to have a debt greater than USD 166,666).

Corporate obligors must be rerated based on information regarding their financial situation, balance sheet information, economic sector prospects, arrears, etc. A relevant grade of this rating system is grade 3, the grade for corporate obligors past due more than 90 days, in conjunction with the other criteria for assessing their default risk.

The Revised Framework (paragraph 452) defines *default* as a situation in which either one or both of the following events have taken place: (1) the bank considers that the obligor is unlikely to pay its credit obligations to the banking group in full, without recourse by the bank to actions such as realising security (if held); and (2) the obligor is past due more than 90 days on any material credit obligation to the banking group.

Taking into consideration the Basel II criteria for classifying obligors as *defaults* and the available information from the PCRB, the default borrowers will be those rated in grades 3 to 5 of the standard rating system of the PCRB.

The database used to model the panel was built taking corporate obligors from the private non-financial[4] sector in December of the last available year (2003) and following back those debtors for four years. To construct the default variable, the grade in December of the following year was considered for each obligor (eg for debtors registered in December 2003 the variable computed their grade in December 2004, and for obligors registered in December 2002 it computed their grade in December 2003). This variable accounts for the default or non-default situation of obligors.[5]

The database has a changing number of observations year to year. There are obligors registered in December of a particular year that are not registered in December of another year. Obligors can be entering or leaving the PCRB during a five-year period. The model works at an *obligor-bank* level. Therefore, each observation corresponds to an obligor of a certain bank or financial institution, ie one obligor might obtain different rating grades from two banks. The sample has 10,094 group observations (*obligor-bank* observations) and its total size is 28,114 observations for the five years. Thus, there is an average of 2.8 observations per group.

The explained variable will be a class variable taking value 1 for obligors classified in rating grade 3 or worse after one year

---

4 The PCRB registers obligors from the public sector, financial private sector, non-financial private sector and non-residents. This document considers the non-financial private sector exclusively, as the credit behaviour and rating grade of other sectors could be influenced by other factors.

5 The commercial debtors selected were those classified in rating grade 1 and 2 that are not in default.

(eg the explained variable of *obligors-bank* registered in December 2003 would take value 1 or 0 depending on their rating classification in December 2004).

The percentage of default obligors fluctuates during the period considered as a result of the aforementioned economic crisis. Table 1 presents the annual value of macroeconomic variables and the frequency of default for each year. This a priori analysis is a suggestion that the PD increases during downturns in the economic cycle (GDP decrease and high unemployment) and is lower in more stable years. The difficulty of modelling a rating system with such fluctuating default rates can be inferred.

Graph 1 displays the frequency of defaults and the growth rates by year, whose relationship is similar to that of unstressed PDs and GDP growth. As expected, there is a negative relationship between these variables. It is evident that 2002 was the worst year of the economic crisis, although this process started in 2001.

Graph 1

Relationship between defaults (right) and GDP growth rates (left), 2000-04

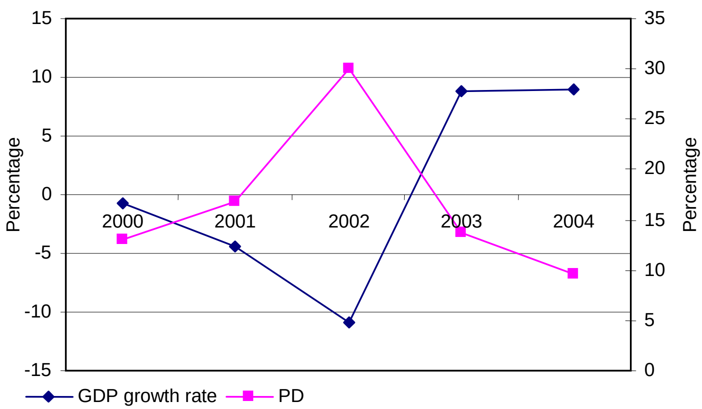


Table 1

Default rates and macroeconomic variables, 2000–04

Period	Number of observations		Percentage of defaults	GDP growth rate	Inflation rate	Unemployment rate
2000	Debtors	6,995				
	Default debtors	912	13.0	–0.8	–0.7	14.7
2001	Debtors	5,885				
	Default debtors	992	16.9	–4.4	–1.4	18.3
2002	Debtors	5,733				
	Default debtors	1,719	30	–10.9	35.1	17.8
2003	Debtors	4,369				
	Default debtors	597	13.7	8.8	3.5	12.6
2004	Debtors	5,132				
	Default debtors	491	9.6	9	6	13

Note: The percentage of default is the ratio of default debtors in December of each year; the GDP growth rate is calculated between December of the previous year and December of the cited year; the inflation rate is the average monthly rate in the cited year; and the unemployment rate corresponds to October of the cited year.

Source: Ministry of economics of the analysed country; PCRB of the analysed country.

3. Panel model of five years

Theory of the panel model

The objective of this paper is to develop a methodology for constructing a TTC rating system with data for a developing country. During the period considered, a major economic crisis affected the country, which would create problems.

There is a great amount of research comparing different models and methodologies for constructing a rating system and the conclusion is that results are similar. In the present document, a probit model is used to estimate the scores of the rating system. The advantage of this model is the direct interpretation of the explanatory variables influencing the PD.

The model used here is a panel model, specifically a random effect probit model. The model is calculated via maximum likelihood, and basically the probability is assumed as:

$$Pr(y_{it} \neq 0 / x_{it}) = \Phi(x_{it}\beta + v_i)$$

for i=1,….n obligors, t=1,….,n_i , v_i are the iid $N(0,\sigma_V^2)$, and Φ is the standard normal cumulative distribution function.

Underlying this model, the variance component model is defined as:

$$y_{it} \neq 0 \Leftrightarrow x_{it}\beta + v_i + \varepsilon_{it} > 0$$

where ε_{it} are iid Gaussian distributed with mean zero and variance 1, independently of v_i.

Empirical results

The explanatory variables can be divided into three groups: (1) the ones that are obligors' financial variables in different years with dimension "*it*"; (2) variables that are characteristics of the obligors that do not change in time, dimension "*i*"; and (3) macroeconomic variables for incorporating the business cycle - these variables change in time but are the same for all the obligors, dimension "*t*".

The variables belonging to group 1 consider the financial situation of debtors. Using the limited information obtained from the PCRB, the explanatory variables account for the rating grade of debtors (present rating grade, rating grade six months before the relevant period, and the worst rating grade registered in the financial system), outstanding debt of the obligor with a bank and with the system, guarantees, number of creditor banks, obligations in default compared to total obligations, and the number of credit lines in default in different periods (see Annex 1 for the definition of explanatory variables and their mean value for default and non-default obligors).

The type of financial institution where the credit is registered is the only variable not changing in time but changing within the group of *obligors-bank* (group 2). The last group of explanatory variables - macroeconomic variables - was presented before (GDP growth, unemployment and inflation). In this research, historical data were used to consider the business cycle but a real implementation of the model would demand predictions for these variables. Clearly, the intention of the model is to forecast the PDs considering future macroeconomic conditions, as is constantly indicated in the Revised Framework.

The results of the regression are presented in Table 2. The signs of the explanatory variables' coefficients are the expected ones.[6] Present, past and worst rating grade variables (constructed with PCRB data) have negative and significant coefficients. High-quality loans with fewer days past due have less probability of default.

---

[6] The random effect panel model estimates the probability that the explained variable has a positive outcome, not an outcome equal to 1. Although the estimate is not strictly a PD, the interpretation of the coefficient is the same.

Table 2

Coefficients of the panel model (default is the explained variable)

Explanatory variables	Coefficient	P>|z|
Intercept	–1.024***	0.004
Previous_rating 0	–0.667***	0.000
Previous_rating1	–0.460***	0.000
Previous_rating 2	–0.489***	0.000
Previous_rating 3, 4 and 5		
Worst_rating 1	–1.265***	0.000
Worst_rating 2	–0.372***	0.000
Worst_rating 3, 4 and 5		
Rating 1	–0.636***	0.000
Rating 2		
Banks 1	–0.758***	0.000
Banks 3	–0.846***	0.000
Banks 5	–0.632***	0.000
Banks 7	–0.381***	0.000
Banks 9	–0.052	0.412
Banks 10		
Institution_type: National public banks	–0.794***	0.000
Institution_type: Local banks with foreign capital	–0.119	0.370
Institution_type: Private cooperative banks	–0.555***	0.001
Institution_type: National private banks	–0.756***	0.000
Institution_type: Public local banks	–0.406***	0.004

Table 2 (cont)

Coefficients of the panel model
(default is the explained variable)

Explanatory variables	Coefficient	P>|z|
Institution_type: Bank branches of foreign inst	–0.347**	0.012
Institution_type: Other financial institutions		
Guarantee	0.225***	0.000
Ldebt_bank	0.059***	0.000
Ldebt_system	–0.233***	0.000
Default_percentage	0.740***	0.000
Ndefault_lines	0.214***	0.000
Ndefault_previous_lines	–0.067***	0.024
GDP_growth	–2.318***	0.000
GDP_growth*Ldebt_bank	0.680***	0.000
GDP_growth*Ldebt_system	–0.985***	0.000
Unemployment_1	23.630***	0.000
Inflation rate	–0.772***	0.002
Insig2u	0.581	
sigma_u	1.337	
rho	0.641	
Likelihood-ratio test of rho=0: chibar2(01) = 691.41 Prob >= chibar2 = 0.000		

Note: ***, ** and * denote confidence levels of 99%, 95% and 90% respectively. The “missing values” correspond to the base dummy variable in each case.

Obligors that have loans from many banks have a higher PD after one year. The obligors' PD changes according to the creditor institution type (eg foreign capital banks, domestic banks, branches of foreign financial institutions and national private banks). Obligors with national public banks are the riskiest ones.

A greater percentage of the loan guarantee (mortgages on residential property and pledges) means a higher PD. Aspects such as "relationship lending" can explain the sign of this coefficient;[7] an alternative hypothesis is that banks may be asking for guarantees when they perceive an increasing risk with the loan.

The larger the amount of outstanding debt with creditor institutions, the riskier the loan, indicating that larger loans have a higher PD.[8] Conversely, as the credit in the financial system increases, the obligors' default is less probable.

The coefficients of the percentage of system debt in default and the number of credit lines in default are positive and significant, which reveals more probability of default with the increase of these variables.

The macroeconomic variables are significant with the expected signs of coefficients. In the years of higher GDP growth rates, obligors have less probability of default. In addition, the positive relationship between the amount of outstanding debt and PD as well as the negative relationship between the amount of system debt and the PD is reinforced in years of higher GDP growth (interaction effects). The unemployment rate increases the PD with a lag of one year, indicating a lag of debtors' response to higher unemployment rates. The coefficient of the inflation rate is negative. There

[7] Berger and Udell (1995); Berglof and Von Thadden (1994).

[8] Barron and Staten (2000).

were some years, such as 2004, with a relatively higher inflation rate but a lower percentage of defaults (Table 1). Nevertheless, the coefficient of the inflation rate has a small value compared to other macroeconomic variables.[9]

The output table shows the rho coefficient, which represents the proportion of the total variance component.[10] As this coefficient is different from zero, the panel estimator is different from the pooled estimator (separate probit model for each year). This result is reinforced by the likelihood test that compares the pooled estimator with the panel one, and the null hypothesis that they are equal can be rejected at a 99% level. Thus, it can be concluded that the panel model works better than a probit model for each period.

Moreover, the random effect model is estimated using a quadrature. To evaluate the stability of the quadrature approximation, the coefficients of the fitted model were compared to other quadrature points and they did not present a significant relative change (see Annex 2).

As presented in the theory, the probit random effect model estimates the probability that the independent variable takes a value different from zero, which is the probability of a positive outcome (not strictly the value one). The estimated probability

---

9 Different models were tested with different macroeconomic variables such as activity, employment, exchange rates and interest rates. As the economic crisis during the period considered altered macroeconomic conditions in general, some of the variables presented a high correlation and cannot be considered together due to multicolineality. The model selection has been done comparing the goodness of fit and the validation measures.

10 The rho coefficient is $\rho = \frac{\sigma_V^2}{\sigma_V^2 + 1}$ which is the proportion of the total variance contributed by the panel-level variance component. When rho is zero, the panel-level variance component is not important and the panel estimator is not different from the pooled estimator.

may diverge from the proportion of the observed defaults in the group.

Thus, the score from the model is the output considered to construct a rating system but the estimated probability from the model was not considered. The negative argument of the normal cumulative distribution presented before, $-x_{it}\beta + v_i$, represents the score of the obligor in a rating system based on a probit model. There is an inverse relationship between this score and the probability of a positive outcome.

4. Risk categories and validation of the model

Discriminatory power

The first step in the validation process is to analyse the discriminatory power of the rating system. The principle of a rating system is to assign a score to each obligor that summarises the default risk information, so it would be desirable that "the better the score, the lower the proportion of default debtors observed". A rating system would discriminate better in the case where the distribution of defaults and non-defaults in the rating grade differs the most. It can be interpreted as the ability of the rating system to distinguish bad from good debtors.

BCBS (2005) describes the commonly used validation measures. Among the statistics to measure discriminatory power, the most popular ones are the accuracy ratio (AR) and its correspondent graphical tool the cumulative accuracy profile (CAP), and the receiver operating characteristics (ROC) curve and its summarised statistic, the ROC measure.

The CAP is also known as the "power curve" or the "Lorenz curve". It is a statistical tool that plots the cumulative frequency of debtors (x axis) and the cumulative frequency of defaulting debtors (y axis) when they have been ordered by their score (from the riskiest to the safest obligors). It is desirable that the rating system should accumulate more default debtors at the beginning as these points correspond to lower scores. Thus, the perfect CAP curve accumulates all the default debtors first. The random model does not have any discriminatory power, so any fraction of debtors with low scores will contain the same fraction of default debtors (diagonal).

A more concave CAP curve shows a better discriminatory power as the CAP curve is closer to the perfect model. The CAP curve plotted in panel A of Graph 2 considers the score of the whole sample of *obligors-bank* from the fitted model and panel B plots the CAP curve in each year. Panel B shows noticeable differences between annual rating systems. The CAP curve in 2000 (*obligors-bank* from December 1999 that can default in December 2000) and that in 2001 are relatively close - these years have similar default rates.

The differences between annual CAP curves are the first signal that the rating system does not work as a TTC rating system with stable scores in the five years. The behaviour of obligors and default rates in the years considered were so different that macroeconomic variables included in the panel model are not enough to account for the economic crisis effect. The obligors' scores and the rating system in crisis years are markedly different.

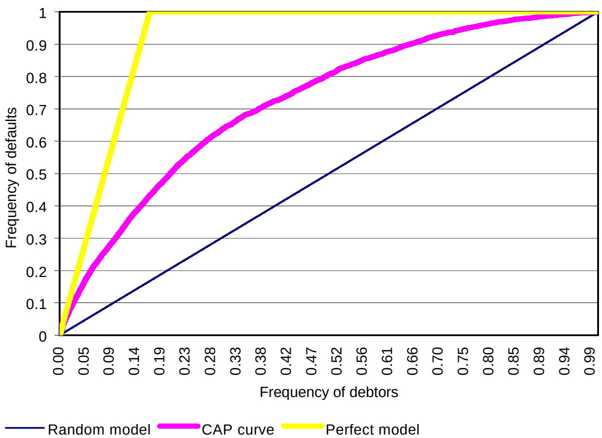
The measure that summarises the CAP curve is the AR. This statistic is calculated as the ratio of the area between the CAP curve and the diagonal (random model) and the area between the CAP curve and the perfect model. The AR increases with the discriminatory power of the rating system, and its maximum value is 100%.

The AR of the estimated panel model is about 54% (Table 3) but the annual AR varies year to year. The rating system for 2002 presents the worst discriminatory power. The most stable years, such as 2000 and 2004, have a greater AR.

Graph 2

CAP curve considering five-year scores and annual CAP curve

Panel A



Panel B

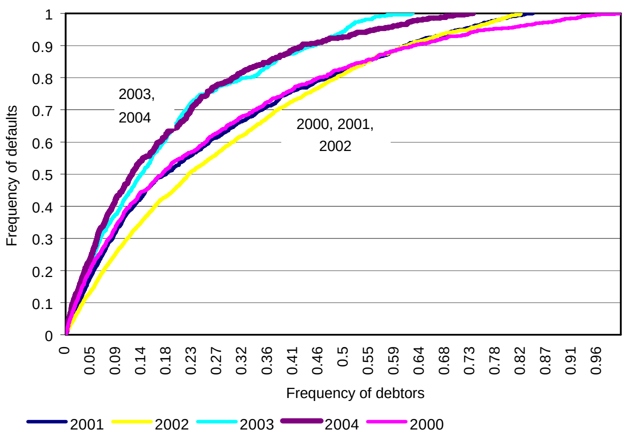


Table 3

Measures of discriminatory power

Commercial debtors	Accuracy ratio (AR)	Receiver operating characteristic (ROC)	Pietra index
Total	**53.7%**	**76.8%**	**0.143**
2000	56.2%	78.1%	0.148
2001	47.4%	73.7%	0.126
2002	45.7%	72.9%	0.119
2003	49.6%	75%	0.148
2004	57.1%	78.5%	0.160
Max (greater discriminatory power)	100%	100%	>
Min (less discriminatory power)	0%	50%	<

The second graphical tool used to evaluate discriminatory power is the ROC curve. This curve relates the hit rate (number of defaults correctly predicted) and the false alarm rate (number of non-defaults incorrectly classified). As with the CAP curve, the more concave this curve the better the discriminatory power of the rating system as the curve would be closer to the coordinates (1, 0) of hit and false alarm rates. The statistic that summarises the ROC curve is the ROC index presented in Table 3. This index considers the area under the ROC curve, and it is basically a linear combination of the AR; therefore the relative order of annual measures is the same as for the AR.[11]

[11] The AR can be calculated as a linear combination from the area under the ROC curve, area A. AR=2A–1. See BCBS (2005).

There is an additional discriminatory measure, the Pietra index, which considers the maximum area of a triangle inscribed between the ROC curve and the diagonal of the united square. The relative discriminatory power is the same as the other measures.

Calibration

The next step in validating a rating system is to check the calibration. This is a different task from testing the discriminatory power as the accuracy of the estimated PDs should be reviewed. Whereas a rating system with a satisfactory discriminatory power has different distributions of defaults and non-default obligors in rating grades, the calibration considers the distance between estimated PDs and realised default rates.

The calibration of the rating system should be tested after establishing the limits of the risk categories. After assigning a score to each *obligor-bank* related to their default risk, a rating system should group obligors into “risk categories” or “risk buckets”. Then, under an IRB approach, a pooled PD should be assigned to each risk category. Calibration should be performed on these PDs.

There are some minimum requirements that a rating system structure should meet taking into consideration the Revised Framework (paragraphs 403-408). Among the standards for corporate, sovereign and bank exposure, a meaningful distribution of exposures across grades with no excessive concentrations is required. It is mentioned that a rating system must have a minimum of seven rating grades for non-default obligors and one grade for defaults. Additionally, the survey of bank practices in relation to the IRB approach[12] of the Basel Committee on Banking Supervision indicates that the average

[12] BCBS (2000).

number of grades reported by banks as covering non-impaired corporate loans was about 10, and the average number of reported problem grades was about three.

Considering the requirements of the Revised Framework and bank practices, the rating system constructed with the scores of the panel model has 13 rating categories. The pooled PD assigned to each risk category in the rating system is the frequency of defaults. The last five categories are for default debtors as they have a pooled PD greater than 16% and the average default rate is 16.7% (Table 4). It has to be mentioned that, in developed countries, default rates are lower than observed in developing countries.[13] The higher default rates observed, even in stable periods, resulted in a greater number of default categories.

The methodology used to construct "risk buckets" is a K-means cluster analysis. Cluster analysis is often used in marketing research as a classification tool to identify discrete categories (taxonomy). The primary use of this methodology in marketing has been for market segmentation, but it also could be used to analyse buyers' behaviour and to develop potential new products, among other uses.[14]

The idea behind cluster analysis is very simple, that is, to identify groups or clusters of individuals using multiple variables. There are different cluster analysis methods, and in particular the K-means method breaks the observations into distinct non-overlapping groups. This non-hierarchical method begins with the partition of observations into a specific number of clusters, and observations are reassigned to clusters by moving them to the cluster whose centroid is closest to the case. Reassignment continues until every case is assigned to

---

13 As a reference, BCBS (2003) indicates that the group of largest banks studied (group 1) had on average 72% of corporate exposures with a PD below 0.8% and that 3% of the corporate exposures were in default.

14 Punj and Stewart (1983).

the cluster with the nearest centroid. This procedure implicitly means minimising the variance within each group. Hence it seems to be an accurate method to establish risk categories.

The required criterion was that each category should group obligors with similar scores and that the rating system should have approximately an equal number of obligors in each bucket to avoid exposure concentrations. Thus, the K-means method for clustering obligors in risk categories was performed on the scores, cumulative frequencies of debtors and cumulative frequencies of defaults from the observation of the five years. The intention was to construct risk categories for a TTC rating system.[15]

BCBS (2005) presents different calibration tests. The Chi-squared test (Hosmer-Lemeshow) is a test to check the estimated PDs for all the risk categories. This test is subject to the assumption of independent default events, and the null hypothesis is that the estimated PDs are the true default probabilities.

The calibration process performed in this case would consider whether the frequency of defaults distributed in the risk categories, pooled PDs, remains relatively stable along the five years (*backtesting*). The panel model considering macroeconomic variables has been constructed to have a TTC rating system. But whatever the limits of the risk buckets, the rating system is not stable during the periods analysed. The calibration of the rating system with annual frequencies of default is not possible because of the great differences among default rates as a consequence of the economic crisis. The results from annual comparisons are presented in Table 5. For the most unstable years, 2002 and 2003, the rating system does not even meet the rule of decreasing PDs from safer to riskier buckets.

---

[15] There were different risk categories established with different criteria, but the selected rating system minimised the Chi-squared test (calibration test).

Table 4

Risk categories of the rating system

Risk categories	Number of obligors	Number of default obligors	Max category score	Min category score	Pooled PD (freq defaults)
1	352	6	5.66	3.35	**0.017**
2	1,311	28	3.35	2.87	**0.021**
3	2,236	64	2.87	2.56	**0.029**
4	2,702	127	2.56	2.30	**0.047**
5	2,961	225	2.30	2.06	**0.076**
6	2,835	264	2.06	1.81	**0.093**
7	2,773	357	1.81	1.56	**0.129**
8	2,574	350	1.56	1.28	**0.136**
9	2,302	373	1.28	0.98	**0.162**
10	2,255	549	0.98	0.67	**0.243**
11	2,011	621	0.67	0.33	**0.309**
12	1,573	569	0.33	–0.09	**0.362**
13	2,229	1,178	–0.09	–3.29	**0.528**

Table 5

Calibration of the five-year rating system

Risk categories	Pooled PD (freq defaults)	Annual PDs				
		2000	2001	2002	2003	2004
1	**0.017**	0.000	0		0.074	0.016
2	**0.021**	0.026	0.018		0	0.021
3	**0.029**	0.030	0.048		0.006	0.026
4	**0.047**	0.048	0.060	1	0.031	0.037
5	**0.076**	0.080	0.086	0.294	0.070	0.050
6	**0.093**	0.117	0.097	0.048	0.062	0.064
7	**0.129**	0.154	0.131	0.134	0.075	0.142
8	**0.136**	0.204	0.172	0.111	0.067	0.103
9	**0.162**	0.218	0.239	0.166	0.063	0.122
10	**0.243**	0.371	0.289	0.245	0.123	0.189
11	**0.309**	0.424	0.421	0.300	0.214	0.236
12	**0.362**	0.579	0.500	0.378	0.224	0.333
13	**0.528**	0.722	0.601	0.603	0.364	0.429
Average default rate	**0.168**	0.130	0.169	0.300	0.137	0.096
Chi-squared		118.67	76.43	64.98	215.47	36.20
Prob of Chi-squared		0.000	0.000	0.000	0.000	0.138

The Chi-squared test performed to compare the pooled PDs from the rating system with the total number of debtors and default debtors each year rejects the null hypothesis that pooled PDs are the true ones.

As a conclusion, it is not possible to construct a rating system with *obligors-bank* considering a five-year period with a deep economic crisis. The change in the default frequency year by year and the exceptional increase in the years of crisis make it impossible to calibrate such a rating system.

5. A rating system for two similar years

The five-year database considers diverse years of default rates as a consequence of a major economic crisis. However, the methodology developed in this document enables the construction of a calibrated rating system with years of closer frequencies of defaults.

By taking obligors from December 1999 and 2000 and considering their score that measures the risk of default one year later, a calibrated rating system can be constructed. Table 6 presents the risk buckets of a rating system developed for these years with 12 categories, the last four being for default obligors (14.7% is the total average rate of default). This rating system is calibrated year by year, as can be inferred from the simple comparison of annual PDs with the pooled PDs of the rating system. The value of the Chi-squared statistic is small enough to not reject the null hypothesis that pooled PDs of the rating system are the true ones when they are compared with the annual information of the number of debtors and the number of default debtors.

The AR calculated with obligors ordered by their scores in 2000 and 2001 is 52.5%, the ROC measure for the same sample is 76.2%, and the Pietra index is 0.14. These are reasonable measures of discriminatory power.

Thus, it is possible to construct a calibrated rating system with a satisfactory discriminatory power using the methodology developed in this document. Year-by-year calibration of the rating system should be performed in years of relatively similar default rates, and this rating system would pass the backtesting.

Table 6

Calibration of the two-year rating system

Risk categories	Number of obligors	Number of default obligors	Max category score	Min category score	Pooled PD (freq defaults)	Annual PD	
						2000	**2001**
1	166	1	4.01	3.17	**0.01**	0.01	0.00
2	651	18	3.16	2.82	**0.03**	0.03	0.01
3	1,046	36	2.82	2.57	**0.03**	0.03	0.05
4	1,250	64	2.57	2.38	**0.05**	0.04	0.07
5	1,327	86	2.38	2.20	**0.06**	0.07	0.06
6	1,372	123	2.20	2.04	**0.09**	0.08	0.10
7	1,340	142	2.04	1.88	**0.11**	0.12	0.09
8	1,295	165	1.88	1.70	**0.13**	0.14	0.12

Table 6 (cont)

Calibration of the two-year rating system

Risk categories	Number of obligors	Number of default obligors	Max category score	Min category score	Pooled PD (freq defaults)	Annual PD	
						2000	**2001**
9	1,161	170	1.70	1.51	**0.15**	0.15	0.15
10	939	180	1.51	1.27	**0.19**	0.23	0.17
11	1,268	366	1.27	0.61	**0.29**	0.30	0.27
12	1,065	553	0.60	–1.71	**0.52**	0.55	0.50
Average default rate					**0.148**	0.130	0.169
Chi-squared						8.51	8.26
Prob of Chi-squared						0.99986	0.99989

6. Is it worth collecting historical information?

The database used in this document includes the credit behaviour of obligors during a deep economic crisis. Thus, the construction of a stable rating system for the five years was not possible. However, the financial information on obligors in the crisis period is important for assessing obligors' risk. The supervisory authority should encourage banks to save this information, although they could argue that this information does not reflect obligors' current conduct and, therefore, cannot be used to construct a rating system.

This argument is false because the history of obligors could be used to construct a more accurate PIT rating system. The information about obligors' behaviour in different years, including years of severe economic conditions, is a useful tool for forecasting their likelihood of default.

The PIT model for estimating the PD with explanatory variables with five years of information is more powerful than a PIT model with information for one year. A database with the history of debtors makes it possible to construct a great number of explanatory variables, such as the ones related to age of credit lines. Additionally, with five years of information, it is possible to construct the same explanatory variables for a longer period, such as the average amount of debt in five years or the worst rating within five years.

An example is presented in Annex 3, where a PIT model with information for five years is compared with a PIT model with information for one year. Both are probit models that estimate the probability that obligors registered with the PCRB in December 2003 default in December 2004. The goodness of fit of the five-year model is greater, the Pseudo R^2 is 0.72 compared with 0.32 for the one-year model, and the discriminatory power is better too, the AR of the five-year model being about 93% compared to an AR of 66% for the one-year model. From a broad comparison of the expected value of the PD (considering estimated PDs of the probit model) and the realised frequency of default, a satisfactory calibration of the model can be inferred.

Even though it was not the intention to develop a PIT model to estimate the PDs, the exercise of comparing a PIT model with a different quantity of information was important in order to highlight the necessity of collecting the financial history of obligors. The scores and PDs estimated from a longer historical database can be used to construct more accurate rating systems. In the present case, with a database that includes an important economic crisis, the estimated scores that consider such a stressed situation were a superior proxy of the obligors' default risk.

Conclusions

This paper has developed a rating system for corporate debtors in a developing country for a period of time that includes a major economic crisis. The model used for estimating obligors' scores was a panel model. It includes macroeconomic variables as explanatory variables. The discriminatory power of the model differs from year to year.

The limits of risk categories were established using a cluster analysis method controlling for the variables of debtors' score, cumulative frequency of debtors and cumulative frequency of defaults. Regardless of score limits, it was not possible to construct a stable rating system with such distinct annual default rates. Some of the annual rating systems did not meet the basic requirement of descending PDs. Consequently, rating system calibration from year to year was not possible as long as the crisis spread among obligors and boosted default rates in crisis years.

It was not feasible to construct a rating system when the period includes years of crisis. This kind of problem could be faced by many emerging economies. Developing countries have experienced many macroeconomic crises during the

1990s with severe consequences for their financial systems. It could be difficult to have a long database with financial information stable enough to construct a TTC rating system.

The Revised Framework (Basel II) demands long-run estimations of PDs compatible with a TTC rating system. Although this is a feasible task in developed countries with relatively stable business cycles, it could be a difficult one in emerging economies with periodic severe economic problems.

The instability of developing countries is a specific aspect, among many others, that would have to be considered in the implementation of Basel II. It is important to analyse the difficulties that financial institutions and supervisory authorities would face in economies such as the one presented here. The strict application of the Revised Framework might not be practical. Therefore the supervisory authority would have to rethink suitable requirements to implement Basel II.

Difficulties do not have to discourage the implementation of Basel II in emerging economies. Limitations are going to be a challenge to supervisory authorities and financial institutions, which should make efforts to adapt the requirements to the specific conditions of the country. The main objective of Basel II is to perform a more accurate management of risk in the banking industry, and this is a task that can certainly be accomplished.

The main problem of the five-year rating system presented in this paper was the annual heterogeneity of the database. We can formulate some alternatives such as the one presented in this paper. A rating system can be designed with information from stable years and close frequencies of defaults. This rating system of course would not be a TTC rating system with stressed PDs, being closer to a PIT rating system instead. Another alternative is to construct a rating system with the information from the entire period but weighting each year differently. We could try to estimate average PDs assigning different weights to crisis PDs, and then the result would lie in between a TTC rating system and an average annual PIT rating system. It would be better to have a longer database and enlarge the influence of stable periods to perform this alternative.[16]

In emerging economies with financial information affected by macroeconomic crisis, it is not sufficient to collect historical data. Considerable changes in economic conditions influence the variables used to assess risk. There are other aspects to be considered, such as the quality of the data and the methodology and adjustments required to incorporate crisis data. Human judgment can play an important role in assessing risk in these cases. Thus, considering human judgment to incorporate the stressed scenarios, a TTC rating system could be constructed.

Financial institutions would have to make an effort to construct stable rating systems in emerging economies and supervisory authorities would have to establish rules to validate them. But the instability of an economy should not limit the use of IRB systems; it is just another challenge on the way to Basel II implementation.

[16] See BCBS (2005). It is described that some banks have proposed reporting PDs derived making moving averages of current and lagged unstressed PDs. The result would be "smoothed" PDs that consider aggregate information and obligor-specific information but place a greater weight on information that does not change over time.

Annex 1:
Explanatory variables

Default: Class variable that identifies the defaulting debtors. This variable assumes value 0 if the debtor is rated in grades 1 or 2 - PCRB grades - and it takes value 1 if the debtor is in grades 3, 4 or 5 in December of the next year (eg for an obligor from December 2003, this variable considers its rating in December 2004).

Rating: Class variable that records the rating grade of the debtor (at *obligor-bank* level). It assumes values 1 or 2 if the obligor is rated 1 or 2 - PCRB grades - respectively, in December of each year.

Previous_rating: Class variable that considers the rating grade at *obligor-bank* level. This variable takes values from 0 to 3 representing the rating grade of the debtor in June of each year - PCRB grades. It takes value 0 if the obligor was not registered in June of each year and value 3 for obligors classified in grade 3 or higher.

Worst_rating: Class variable that assumes values from 1 to 3 representing the worst (highest) *obligor-bank* rating grade in each year - PCRB grades. A value of 3 identifies debtors rated in grades 3 to 5.

Banks: Class variable that considers the number of creditor financial institutions of the obligor. This variable assumes 1 if one institution has claims with the obligor, 3 if that is true for two or three institutions, 5 for five institutions, 7 for six or seven institutions, 9 for eight or nine institutions and 10 for 10 or more institutions.

Institution_type: Class variable that classifies financial institutions according to capital origin. This variable is broken down into seven excluding dummies: national public banks; local banks with foreign capital; private cooperative banks;

national private banks; public local banks; branches of foreign financial institutions; and other financial institutions.

Ldebt_bank: Logarithm of the outstanding amount of the claims of a specific financial institution with a certain obligor in December of each year.

Ldebt_system: Logarithm of the outstanding amount of overall claims of the banking system with a certain obligor in December of each year.

Guarantee: Percentage of the claims that are guaranteed in December of each year.

Default_percentage: Proportion of the claims of the banking system with a certain obligor that is reported in default in December of each year.

Ndefault_lines: Number of corporate credit lines of the obligor classified as in default in the banking system in December of each year.

Ndefault_previous_lines: Number of credit lines of the obligor classified as in default in the banking system as of June in each year.

GDP_growth: GDP growth rate during each year (eg for obligors in December 2003 that may be default obligors in December 2004, the growth rate during 2004 is considered).

Inflation rate: Inflation rate for each year.

Unemployment_1: Unemployment rate of the previous year; this variable is lagged one period.

Table

Taxonomy of explanatory variables of *obligors-bank*

Variables		Number of observations		Percentage	
		Non-default	Default	Non-default	Default
Previous_rating	0	2,506	404	10.7%	8.6%
	1	19,607	3,607	83.8%	76.6%
	2	1,069	550	4.6%	11.7%
	3	121	51	0.5%	1.1%
	4	75	51	0.3%	1.1%
	5	25	48	0.1%	1%
Worst_rating	1	18,390	2,176	78.6%	46.2%
	2	2,348	1,016	10%	21.6%
	3	1,022	472	4.4%	10.0%
	4	819	526	3.5%	11.2%
	5	824	521	3.5%	11.1%
Rating	1	21,910	3,677	93.6%	78.1%
	2	1,493	1,034	6.4%	21.9%
Banks	1	3,548	537	15.2%	11.4%
	3	6,792	1,076	29.0%	22.8%
		4,689	937		
	5			20.0%	19.9%
	7	2,866	669	12.2%	14.2%
	9	1,871	533	8.0%	11.3%
	10	3,637	959	15.5%	20.4%

Table (cont)

Taxonomy of explanatory variables of *obligors-bank*

Variables	Number of observations		Percentage	
	Non-default	**Default**	**Non-default**	**Default**
Institution_type:				
National public banks	778	133	3.3%	2.8%
Local banks with foreign capital	7,458	127	31.9%	2.7%
Private cooperative banks	1,081	1,862	4.6%	39.5%
National private banks	8,081	164	34.5%	3.5%
Public local banks	2,086	1,116	8.9%	23.7%
Financial institutions	3,484	573	14.9%	12.2%
Other financial institutions	435	736	1.9%	15.6%
Means of continuous variables				
Guarantee	0.23	0.28		
Ldebt_bank	5.85	5.61	\$346,130	\$274,480[1]
Ldebt_system	7.22	7.00	\$1,361,350	\$1,093,120[1]
Default_percentage	0.07	0.17		
Ndefault_lines	0.04	0.14		
Ndefault_previous_lines	0.11	0.17		

[1] Amount of debt in the bank and in the system expressed in USD.

Annex 2:
Test of quadrature stability

The random effect model is estimated using a quadrature (M-point Gauss-Hermite quadrature). For the cases where this quadrature approximation is not accurate, the change in the number of points from the quadrature would change the coefficients and this approximation should not be used. An estimation of the model for two more quadrature points was performed and the results were stable as relative differences of the significant coefficients were less than 2.3%.

Table

Relative difference with quadrature of 12 points (fitted model)

Variables	Comparison quadrature 8 points	Comparison quadrature 16 points
Log likelihood	0.003%	–0.002%
Previous_rating 0	–0.23%	–0.03%
Previous_rating 1	–0.67%	–0.08%
Previous_rating 2	–0.96%	–0.05%
Worst_rating 1	–0.23%	0.16%
Worst_rating 2	0.07%	–0.07%
Rating 1	–0.36%	0.11%
Bank 1	1.44%	0.03%
Bank 3	1.08%	–0.03%
Bank 5	1.19%	0.02%
Bank 7	1.54%	0%
Bank 9	9.13%	–0.24%

Table (cont)

Relative difference with quadrature of 12 points (fitted model)

Variables	Comparison quadrature 8 points	Comparison quadrature 16 points
Institution_type: National public banks	–0.55%	0.67%
Institution_type: Local banks with foreign capital	–7.79%	2.10%
Institution_type: Private cooperative banks	–1.73%	0.68%
Institution_type: National private banks	0.05%	0.42%
Institution_type: Public local banks	–1.77%	0.25%
Institution_type: Bank branches of foreign inst	–2.30%	0.68%
Guarantee	2.13%	0.04%
Ldebt_bank	1.66%	–0.21%
Ldebt_system	0.79%	–0.01%
Default_percentage	0.33%	0.12%
Ndefault_lines	–0.22%	0.17%
Ndefault_previous_lines	–3.55%	0.09%
GDP_growth	0.50%	–0.41%
GDP_growth*Ldebt_bank	–0.32%	0.12%
GDP_growth*Ldebt_system	–0.14%	0.19%
Unemployment_1	0.40%	0.28%
Inflation rate	1.74%	0.18%
Constant	1.24%	0.73%
Insig2u	0.54%	0.91%

Annex 3

A probit model was used to estimate the probability that corporate obligors in December 2003 default in December 2004. The model with history has explanatory variables constructed with the information on obligors over five years. The explanatory variables are the same as those used in the previous panel model but they considered the information over five years (eg Worst_rating considered the worse rating grade of obligors over five years, and Guarantee considered the average percentage of guarantee debt over five years). The new explanatory variables of the history model are Never_default, which takes value 1 if the obligor has never defaulted a corporate credit line during the five-year period or zero otherwise, and Nyears_loan, which considers the age of the credit line.

Table

Probit regression with five-year and one-year information

Explanatory variables	Model with history	P>|z|	One-year model	P>|z|
Intercept	0.309	0.415	–0.462	0.282
Previous_rating (five-year average)	0.201	0.131		
Previous_rating 0			–1.763***	0.000
Previous_rating 1			–1.614***	0.000
Previous_rating 2			–0.976***	0.000
Previous_rating 3, 4 and 5				
Worst_rating 1	–0.272	0.143	–0.577***	0.000
Worst_rating 2	–0.407**	0.044	–0.074	0.407
Worst_rating 3, 4 and 5				
Rating 1	–0.203	0.131	0.29**	0.014
Rating 2				
Never_default	–1.75***	0.000		
Banks (five-year average)	0.132***	0.000	0.067***	0.002
Institution_type: National public banks	0.251**	0.038	1.219***	0.003
Institution_type: Local banks with foreign capital	0.969***	0.000	1.472***	0.000
Institution_type: Private cooperative banks	0.04	0.872	0.522	0.224
Institution_type: National private banks	0.455**	0.038	0.796**	0.043

Table (cont)

Probit regression with five-year and one-year information

Explanatory variables	Model with history	P>|z|	One-year model	P>|z|
Institution_type: Public local banks	–0.377	0.138	1.34***	0.001
Institution_type: Bank branches of foreign inst	0.761***	0.001	1.04***	0.008
Institution_type: Other financial institutions				
Guarantee	0.301**	0.025	0.158**	0.029
Ldebt_bank	0.095**	0.027	0.04*	0.063
Ldebt_system	–0.205***	0.000	–0.18***	0.000
Default_percentage	8.962***	0.000	0.35***	0.007
Nyears_loan	–0.412***	0.000		
N_lines	–0.085*	0.057	–0.065**	0.011
Plines_default	8.962***	0.000		
Pseudo R^2	0.7211		0.3186	
Mean estimated PD	9.53%		9.46%	
% of sample defaulters	10.47%		10.51%	
AR	93.18%		66.10%	
ROC measure	96.59%		83.05%	
Pietra index	0.296		0.181	

Note: ***, ** and * denote confidence levels of 99%, 95% and 90% respectively. The “missing values” correspond to the base dummy variable in each case.

Bibliography

Barron, J M and M Staten (2000): *The value of comprehensive credit reports: lesson for the US experience - summary*, World Bank, mimeo.

Basel Committee on Banking Supervision (2000): *Range of practice in banks' internal rating systems*, discussion paper, January.

——— (2003): *Quantitative Impact Study 3: overview of global results*, May.

——— (2004): *International convergence of capital measurement and capital standards: a revised framework*, June.

——— (2004): *An explanatory note on the Basel II IRB risk weight functions*, October.

——— (2005): "Studies on the validation of internal rating systems", *Working Paper*, 14, May.

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

Berglof, E and Von Thadden (1994): "Short term versus long term interest: capital structure with multiple investors", *Quarterly Journal of Economics*, 109, pp 1055-84.

Gordy, M (1998): *A comparative anatomy of credit risk models*, Board of Governors of the Federal Reserve System, December.

Grunert, J, L Norden and M Weber (2005): "The role of non-financial factors in internal credit ratings", *Journal of Banking and Finance*, 29, pp 509-31.

Moody's Investors Service (2000): *Benchmarking quantitative default risk models: a validation methodology*, March.

Punj, G and D Stewart (1983): "Cluster analysis in marketing research: review and suggestions for application", *Journal of Marketing Research*, 20 (2), May, pp 134-48.

Stata Corporation (2003): *Cross-sectional time series*, Reference Manual, release 8, Stata Press Publication.

Wooldridge, J M (2002): *Economic analysis of cross section and panel data*, MIT Press, Cambridge.