Working Paper 2006 | 4

Credit Scoring models with truncated samples and their validation

Verónica Balzarotti / Matías Gutiérrez Girault
Verónica Vallés
BCRA

May, 2006

ie | BCRA

Investigaciones Económicas
Banco Central
de la República Argentina

Banco Central de la República Argentina
ie | Investigaciones Económicas

May, 2006
ISSN 1850-3977
Electronic Edition

Reconquista 266, C1003ABF
C.A. de Buenos Aires, Argentina
Tel: (5411) 4348-3719/21
Fax: (5411) 4000-1257
Email: investig@bcra.gov.ar
Pag.Web: www.bcra.gov.ar

Credit Scoring models with truncated samples and their Validation

By Verónica Balzarotti, Matías Gutiérrez Girault and Verónica Vallés

I. Introduction

The main object of this paper is to develop a credit scoring methodology for Argentine bank commercial obligors based on information available in the Public Credit Registry of the Central Bank of Argentina (*Central de Deudores*, "CD") as a referential tool to assess credit risk in local banks. Previous experience in this field has shown promising results[1]; in this paper, we focus on two innovative aspects: firstly, the potential bias introduced by the fact that a considerable number of obligors are removed from the database for no traceable reason, and secondly, the application of validation techniques to the resulting models as proposed by the document recently published by the BCBS[2].

The credit bureau is a rich source of information to assess credit risk. However, it imposes some limitations. In particular, the database comprises those obligors who are already bank clients, while data on potential clients are not recorded. Thus, any credit scoring system developed using the CD database is subject to the effects of this truncation. The resulting bias is well-known in the literature as "selection bias"[3]. In the context of this study, however, such a bias is not of great concern, as long as we are aiming at the estimation of credit risk on bank portfolios from a regulatory point of view, in contrast to the case in which the objective was to develop a system to make loan-granting decisions. There is another limitation in the database of particular interest to us in this paper, which is the fact that a group of obligors regularly disappears from the CD for no traceable reason. In consequence, it is impossible to know with certainty if obligors have been removed because they have settled all their debts or because the bank has ceased actions to collect. That is, two opposed reasons may be behind the removal of an obligor from the database, besides the possibility of a straight data flaw. If the amount of removals is considerable and the removal process is not random, there may be a bias in the estimation model. We will try to prove if this bias exists and its impact in the Argentine case. Accordingly we will formulate a proposal to introduce some changes to the Argentine database, while we explore some methodologies to correct the bias.

Additionally, there had been a certain lack of consensus in the literature as to the validation methodologies to be applied to credit scoring models and rating systems, until this gap was to some extent filled by the publication of the BCBS working paper *Studies on the Validation of Internal Rating Systems*[4]. We apply some of the methodologies proposed in that paper to our models to examine discriminatory power and validation.

A wide range of statistical models have been developed to aid the credit granting decision. These models can be used not only for this purpose but also to estimate obligors' probabilities of default. These "PDs" are demanded by the Revised Framework of International Convergence of Capital Measurement and Capital Standards (Basel II) to calculate capital requirements within the framework of Internal Ratings Based (IRB) approaches. Consequently, credit scoring models and their quality have received much attention lately.

---

[1] Balzarotti, V., C. Castro and A. Powell (2004) and Balzarotti, V., C. Castro and A. Powell (2002).
[2] Basle Committee on Banking Supervision, "Studies on the Validation of Internal Rating Systems", Working Paper N° 14, May 2005.
[3] Crook, J. (2002); Banasik, J. and Crook, J (2004); Hand, D. J. and Henley, W. E. (1997); Verstraeten, G. and Van den Poel, D. (2004) and Greene, W. (1992).
[4] see footnote #2.

In this document we are not trying to construct a credit scoring model as a tool to be used in the administration of credit risk within a bank, nor to discuss pros and cons of different models. Instead, we apply a probit model, which has proved to work well in previous studies[5], and prefer to focus on the innovations of this paper, namely the correction of the abovementioned bias and an exploration of validation techniques.

In what follows, section II describes the CD and states the definition of default used in the estimation of the probit model, while section III describes the theoretical model, explains the difference between truncation, censoring, selection bias and the problem of removed-obligors. Section IV describes the different estimated models that attempt to correct this last problem and section V presents their validation (discriminatory power and calibration). Finally, section VI contains the conclusions.

II. Argentine Public Credit Registry: "Central de Deudores"

The BCRA has developed a database named "Central de Deudores"(CD) in which consumer and corporate debts with financial institutions, credit card issuers and financial trusts are recorded if the outstanding claim is above Arg.$50 (this amount was equivalent to USD50 before the devaluation in 2002 and is now equivalent to approx. USD17). Any debtor data is freely accessible on the BCRA web page by submitting the obligor's tax identification number. Institutions send the information to the BCRA monthly.

Records in the CD cover loans, claims stemming from financial intermediation, leasing and other claims, as well as contingent claims (guarantees, agreed overdrafts in current accounts and other pre-agreed lines, etc.). Positive and negative information is recorded. For each obligor in 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 (as recognized by the BCRA), rating grade and provisions. There is no demographic data nor information on arrears. The information for some variables show low quality, such as interest rate and maturity.

The creation of this database pursued a number of objectives. These can be briefly summarized as: (i) to foster credit accessibility, (ii) to strengthen supervision, (iii) to promote competition, (iv) to strengthen willingness to pay and (v) to supply useful information for economic and financial risk research, macroeconomic policy making and banking regulation.

For risk classification purposes, the portfolio of claims is divided into two groups: "Consumer and Housing Loans Portfolio" covers loans aimed at consumption financing, home mortgages and, at the bank's option, claims to fund commercial activities of less than Arg.$500.000[6]. The "Commercial Portfolio" includes the rest.

The present layout of the CD dates back to 1997, when the BCRA merged in a single database two former, partial databases. Today, it contains information on 5.36 million individuals and 96,500 corporate obligors, which together account for 7.4 million claims in the banking sector. As of the end of 2005, and considering private sector obligors only, the consumer portfolio (including SMEs) amounted to $35.8 billion (with 7.5 million obligors) and the commercial portfolio to $43.6 billion (19,600 obligors).

[5] Balzarotti, V., C. Castro y A. Powell (2004) y Balzarotti, V., C. Castro y A. Powell (2002).
[6] This threshold was increased from Arg. $200,000 in March 2005.

At the origination of the credit, all debtors must be rated into five grades (1 being the best) according to the BCRA's set of rules, which fundamentally consider the probability of honoring 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 and there is a 50% reduction of provisions on secured debt. Obligors whose debts are totally covered by the best type of collateral (such as cash, term deposits, etc.) are not rated and are reported to the CD in the best grade.

Commercial obligors' ratings must be re-assessed according to their projected financial situation, insolvency legal status and, when the claim has been restructured, according to the progressive cancellation of the outstanding principal[7]. The regulation enumerates a set of criteria to analyze the financial situation, including timely balance-sheet information, economic sector prospects, arrears, etc. On the other hand, consumer obligors are re-rated only according to arrears and insolvency legal status. Rules set by the Central Bank establish that arrears of 90 days or longer force a "3" grade rating for consumer obligors, while the same 90 day-arrear *suggests* such a rating in the case of a commercial obligor (as the other criteria must also be considered). Another paragraph in the regulation makes a bank change its rating if there is a difference greater than one grade between its rating and those assigned by at least two other institutions whose claims account for at least 40% of the claims of the obligor in the banking sector.

Definition of Default

As indicated, there are five grades under the standardized rating system[8]. In deciding which grades mean "default" we have mainly considered the "consensus" definition set out in Basel II. In that framework, default takes place when either or both of the following events have taken place:

- 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 realizing security.
- The obligor is past due more than 90 days on any material credit obligation to the banking group[9].

Thus, we classify credits in grades 3 to 5 as defaults. Again, grade 3 means that an obligor in the consumer portfolio is past due more than 90 days, and the same is true for commercial credits, except for the fact that in this case the criterion is only indicative and could be overridden by other criteria (although this is not usual).

Removed obligors

On a regular basis groups of debtors cease to be recorded in the CD. For example, 16 % of the initial sample of commercial credits used for the present study (see characteristics of the sample below) was not present in the CD one year later, and the same happens with 24% of the obligors

[7] These rules were repeatedly changed after the crisis in 2002 and this is one of the reasons to choose 2000 as the base year.

[8] A rating category number "6" includes a small number of non-performing credits of liquidated banks. We decided not to consider these loans in the analysis.

[9] For retail and Public Sector Entity (PSE) obligations, and as long as the national supervisor considers it appropriate due to local conditions, 180 past due days can be used instead of 90 days.

in the consumer credit segment (Table 1). Moreover, the removals take place from all the rating grades, as shown in Annex 1[10]. Even for monthly horizons, there are removals taking place from all the rating grades.

Table 1: Missing values in commercial and consumer credits

	Default variable	Frequency	Percentage
Commercial	Non Default (0)	12,632	76%
	Default (1)	1,352	8%
	Missing (.)	2,668	16%
	Total	16,652	100%
Consumer	Non Default (0)	3,352,717	68%
	Default (1)	364,956	7%
	Missing (.)	1,188,773	24%
	Total	4,906,446	100%
	Total sample	4,923,098	
	Total population	6,153,873	

The reason for these removals is not recorded, therefore these missing debtors cannot be classified as default or non-default at the end-point[11]. A priori, there are two opposed reasons for a removal:

1) Total settlement of the claims. These debtors should be classified as non-defaults
2) The bank writing off the claims. These obligors should be classified as defaults.

BCRA regulations state that arrears of 365 days or more are indicative of a "loss", i.e. "grade 5" rating for a commercial loan, and must be fully provisioned (except when there is collateral or guarantees, in which case the levels of provisioning are lower). Loans classified as "loss" and totally provisioned must be transferred to off-balance sheet accounts after being seven months in that situation. Loans with guarantees lose the benefit of reduced provisioning after remaining two years in categories 4 or 5 (in general). Thus, the existence of guarantees may delay the final migration to off-balance sheet accounts. A claim can remain in these accounts as long as the institution continues to carry on with collection efforts.

Claims in off-balance sheet accounts must also be reported to the CD. Therefore, if a commercial loan originally rated "1" or "2" (indicative maximum arrears of 89 days) does not appear at all in the CD in one year's time, and assuming it is due to the writing off of a default, it would imply that the bank has applied very strict writing off procedures, much stricter than those demanded by the BCRA. This is not very likely, although not impossible, considering that the amount of any commercial loan is relatively important for local banks. In this line, it can be observed that some obligors classified as "1" or "2" in December 1999 are classified as "5" in December 2000 in our sample (see Table 2).

[10] The scoring model works with obligors that are rated 1 or 2 at the starting point, however, the fact that some obligors are removed from all the rating grades may be informative and that is why it is pointed at.
[11] There is a separate database where write-offs should be recorded, but the information available has low quality.

Table 2: Rating of Commercial credits December 1999-2000

Rating in December 1999	Rating in December 2000						
	0	1	2	3	4	5	Total
1	2,533	11,785	449	282	572	143	15764
	16%	75%	3%	2%	4%	1%	
2	135	137	261	104	179	72	888
	15%	15%	29%	12%	20%	8%	

A third possible reason for removals is the sale of the claims to another institution or to a fund. The claim will continue to be furnished to the CD unless the buyer is a fund that is not obliged to inform (financial fiduciary funds are the most common type of buyer and they must report). Even if the buyer is not a financial fiduciary fund and the selling bank, or another bank, provides collecting services, then this institution must report the claims to the CD. However, as our analysis is at an obligor-bank level, a change of holder will be taken as a removal. Only at a systemic level could transfers be inferred. Additionally, there is an operational reason that may lead to a temporary loss of transferred claims from the database, which is originated in the fact that new funds can report only when they have completed certain authorization procedures, which take a couple of months. If this delay coincides with the end-point observation, some claims may be missing for this reason.

A fourth, final reason for the disappearance of an obligor from the CD is a direct data error, which would be random.

While four reasons can therefore be found for a removal, it can be argued that repayments, sales or mistakes are more likely to be behind them. In order to gather more clues, the set of disappearing obligors can be described and compared to obligors who remain in the database. It can be seen (Annex 1) that both groups are highly similar. There is only a clear discrepancy in the percentage of foreign-incorporated obligors, which is higher in the set of disappearing obligors. This also points in the direction of repayment, sales or errors, rather than defaults and total ceasing of collection efforts.

Another useful hint may come from the analysis of outstanding debt in the system of the obligors who are removed from one bank but still have credit (commercial or consumer) in the financial system (remember the data enters the model at obligor-bank level). Both amount and rating can be enlightening and are presented in Table 3. In general these obligors' total credit in the system has decreased due to the disappearing of some credits with certain banks.

Table 3:Change of outstanding debt in the system, December 1999 to December 2000, and worst rating in December 2000 for disappearing debtors

Debt change in the system	Frequency	%	Average change rate	Average max rating
Increase	537	20%	118%	1.30
Decrease	1,531	57%	-52%	1.68
Disappear	600	22%		
Total	2,668	100%		

III. The Credit Scoring Model and Calculation of PDs

There are several established statistical methods to construct a credit scoring. The most popular methodologies are discriminant analysis, neural networks, decision trees and traditional statistical methods such as Logit and Probit regressions. There is a considerable amount of research comparing the performance of these credit scoring models and the main conclusion is that results are relatively similar. The technique used in the present document is a Probit regression with a binary explained variable for the default of the obligor. We have selected this model because of its simplicity and because the influence of the explanatory variables on the probability of default is quite easily obtainable[12]. It has also proved to work well in previous studies and has the advantage, together with logit models, that rating scores can be easily translated into the borrower's probability of default (PD).

In the Probit model the binary explained variable is the situation of default and can be represented as:

$$D_i = \begin{cases} 1 \ if \ \ X_i^{'}\beta + \varepsilon_i > 0 \\ 0 \ if \ \ X_i^{'}\beta + \varepsilon \le 0 \end{cases}$$

Where $X_i{}'$ is the matrix of explanatory variables of each obligor *i* and ε_i is the regression error. The $-X_i{}'\beta_i$ of these estimations is used as the scores of the debtors, a greater score means less probability of default. The PD can be calculated from the Normal Cumulative Distribution function for Probit models together with the score. Therefore, the probability of default can be estimated as:

$$P(D_i = 1 / X_i) = \Phi(X_i^{'} \beta)$$

We have worked with private, non-financial debtors[13] and have distinguished between commercial and consumer loans[14]. The estimation of the consumer loan model deserves further research since it would be more accurate to make a distinction according to consumer credit type. This would be more in line with Basel II, as IRB banks must group retail debtors according to their characteristics (such as type of credit, days past due, etc.) and estimate pooled PDs shared by all credits within each segment.

Basel II has also established that the pooled PD of each grade in a rating system must be a long run average of yearly default rates within each grade, or of individual PDs of the borrowers assigned to each grade. In this document we have restricted our analysis to the estimation of unstressed annual PDs, that are an input in the construction of *Point-in-Time* (PIT) rating systems, and that constitute a starting point for further research that considers stressed PDs and Through-the-Cycle (TTC) rating systems.

---

[12] Among the requirements of Basel II, the understanding and use of internal rating model as a methodology to grant credits is highlighted. Banks should use simple and comprehensive models to estimate PDs, so people working in the commercial department can understand how these models work and how explanatory variables influence PDs.

[13] Including domestic and foreign obligors.

[14] This distinction could not be obtained directly from the database for the months in the study since the category variable indicating commercial or consumption credit is not available. Thus, it was constructed applying the limit of Arg$200.000 outstanding debt, which, in the regulation, is the greatest commercial loan that can be treated as a consumption loan, at the option of the bank.

The horizon of the model is one year, in order to obtain annual PDs, and we have chosen to work with the period December 1999 - December 2000. The estimated PDs represent the probability that a debtor rated as non-default (grades “1” or “2”) in December 1999 could be rated as default (grades “3”, “4” or “5”) in December 2000. Though there was a recession going on in that year, it was still previous to the very serious crisis of the banking sector that accelerated in late 2001. While we are of the view that the levels of risk arising from the model may probably be above those arising from a long run average, still the conclusions drawn on topics such as the existence of the bias due to removals and the application of validation tools are of great interest and relevance.

The model works at a *debtor-bank level*, i.e. one obligor might obtain different ratings in two banks (subject to the restriction of maximum difference between material debts, stated above). While the CD contains the whole population of debtors, a sample was taken to test the performance of the credit scoring model in-sample and out-of-sample. A sequential sampling method was used to take 75% of the observations controlling for the main variables (rating grades of obligors in December 2000 and outstanding debt in each institution). A Probit model was estimated with correction of outliers considering *dfbetas* or *dcook* of observations.

The sample of commercial obligors used in the estimations has 16,652 debtor-bank observations, although the number of debtors consolidating across banks is 10,343 (the same firm may be obligor in more than one bank; on average each debtor has credits with 1.6 banks). Banks have on average 152 commercial claims, but these are not equally distributed among them: out of 109 institutions, 81 have less than 100 commercial claims each.

The explanatory variables used in the models have been constructed with the limited information of the CD, whose variables have an uneven quality. Therefore we have used those which show relatively good quality. These variables mainly refer to the rating grade of the debtor in different periods, the worst grade recorded, outstanding debt of the obligor with a bank and with the system, guarantees and obligations in default compared to total obligations (for further details see Annex 2).

Tentative treatment of removals

The treatment of removals (borrowers that disappear from the sample between December 1999 and December 2000) will determine different versions of the scoring model.

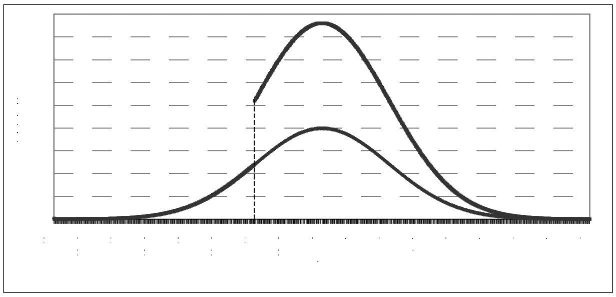
Before looking into the models, it is worth giving some detail about the difference between truncation, censoring, “selection bias” and the problem of disappearing obligors. This is done in an attached box.

The difference between truncation, censoring, selection bias and the removed-obligors problem

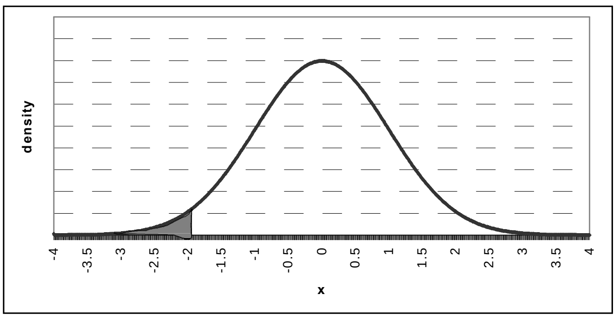
The effect of truncation occurs when sample data is drawn from a subset of a larger population of interest. Censoring is essentially a defect in the sample data. When the dependent variable is censored, values in a certain range are reported as a single value. This introduces a distortion into statistical results. The sample selection problem is a form of truncation, but arises when the truncation process is not random.

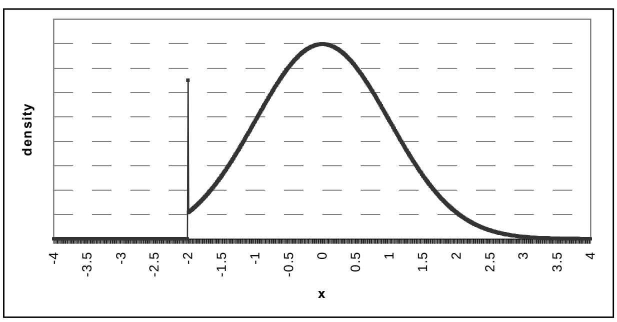
The following graphs illustrate truncation and censoring problems for a Normal distribution.

Truncated Normal Distribution (at –1)



Complete and Censored Normal Distributions (at –2)

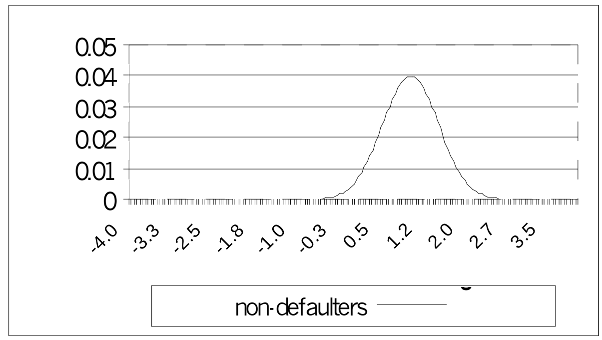




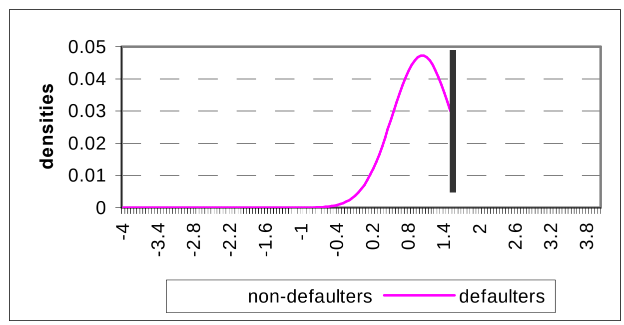
Whether the effect of truncation on the coefficients is of interest depends on the intended inferences of the study. If the analysis is to be confined to the subpopulation, then obtained biased coefficients are of interest. If the study is intended to extend to the entire population, then the unbiased coefficients are actually of interest.

When there is sample selection, sampling is non-random. Recent literature has studied the effects of this drawback on the properties of conventional estimators. In the case of banks assessing credit risk, sample selection is a common problem. Banks usually keep records only for applicants that get credit and not for rejections. In those cases, as opposed to the bias we study in this paper, the truncation is one-sided and there are a number of proposals in the literature to correct its effects. Sample pre-selection can be represented by a condition of acceptance of an application equal to the score being smaller than a threshold. Although this is a simplification of the criteria for granting credit applied by a bank, it is still quite intuitive. Thus, the distribution of defaults and non-defaults will be observable for the censored scores, as illustrated by the next graphs. The first graph shows complete densities of defaults and non-defaults, according to scores, and the second, the densities of defaults and non-defaults with sample selection, where the right tails of the densities are not observed (the areas under both truncated curves add up to 1).

Complete densities of default and non-default obligors



Densities of default and non-default obligors under sample selection



The potential bias of the coefficients arises when the selection mechanism proxies are omitted in the PD equation. Some corrections have been studied in the literature and applied in practice.

In the case the present study, a kind of truncation arises from the fact that some obligors who are present in the CD in Dec. 1999 are removed by a non random process in Dec. 2000. Not considering these group may also lead to biased estimations.

The bias due to obligor removals, however, has no aprioristic side or is both sided. Removals may be due to total cancellation and sales or, at the other extreme, to total writing-off and removing from balance-sheet, as well as to data errors.

Not only does the existence (or non-existence) of removed obligors affect model estimation, it also influences the assessment of the performance of rating models.

IV. The 5 models

Model 1 : *"Incomplete Sample – No Correction Model".* A first approach is to proceed with the estimation of the model restricting the database to the group of obligors that are present both in December 1999 and in December 2000, ie., with no attempt to correct the eventual bias. This is going to be the benchmark model and it is also the approach used in previous research in Argentina and other Central Banks.

Model 2: *"Incomplete sample - Heckman type correction"*. The same truncated sample as Model 1 is considered but fits a Maximum-Likelihood Probit model with sample selection. Thus the method used to correct selection bias in Model 2 is similar to the Heckman method, but with a binary explained variable (default), as follows:

Let X_t be the matrix with the explanatory variables (financial information about obligors). Let Y_t be a dummy variable that indicates whether a specific obligor remains in the system or not (this is the variable that considers the selection process). This variable assumes value 1 if a debtor is present at the end-point and 0 otherwise. D_t is a dummy variable that indicates default. Hence, D_t is observable only in cases where Y_t =1, ie, when the obligor remains in the CD, but it is unknown for those that are removed (Y_t =0). The matrix of financial information, X_t combines the probability of remaining in the CD and the probability of defaulting. The chain determining events is:

$$X_i \text{ -> } Y_i\ (X_i) \text{ -> } D_i(Y_i, X_i)$$

If Y_t was established at random, the second step would not add additional information and there would be no need to consider the selection process. In that case Model 1 would be suitable to estimate PDs. But if Y_t contains additional information, the selection process determining permanence in the CD should be considered in order to avoid a bias. Model 2 assumes that Y_t adds information and therefore tries to model the selection process.

Models 3 to 5 : Another approach that is going to be explored is the re-incorporation of missing obligors, assuming their behavior (default or not) at the end point. This methodology resembles that of Hand and Henley[15]. To proceed in this way, it is necessary to decide how a default or non-default tag will be assigned to obligors in order to reincorporate them to the sample. The default classification rule for removed obligors defines models 3 to 5.

Model 3: *"Global Reincorporation as Cancellations"* Following the a priori rationalization that has been put forward above, there are many arguments supporting that most of removed obligors are in fact settlements or sales (or may be errors). From the point of view of risk, sales represent realizations of claims and are similar to getting the credits paid back. Thus, one extreme assumption will be that all disappearing obligors have indeed paid off. Thus, under this rule all of them will be reincorporated with a non-default tag. This seems to be an extreme approach, but in the right direction.

[15] Hand D.J. and Henley. W.E. (1993) and Hand D.J. and Henley. W.E. (1994)

Model 4: *"Reincorporation according to average score of defaulters".* Observing that the features of the set of removed obligors are in fact very similar to remaining obligors (which is later on confirmed by the scores estimated for both groups), an exercise can be done in which, first, a scoring is estimated using the truncated sample of obligors that are present in both periods, December 1999 and December 2000 (similar to Model 1). Then removed obligors get a score using the parameters obtained with that model. After that, two rules can be thought of to discriminate defaulters and non-defaulters, using these estimated scores: The first rule is to assign a default tag to those cases whose scores are lower than the average score of defaulting obligors of the remaining obligors set. Finally, after classifying removed obligors they are integrated into the sample and PDs are estimated again using a Probit model, this time with the increased sample.

This approach is similar to that of Hand and Henley. They analyze the selection problem in the decision process of credit granting. In their case they cannot follow people in the 'rejected' region as they were not granted a credit. With a similar process they re-include the rejected people in the sample, although they find that this appears to have little additional information.

The second rule would be to assign a default tag to a percentage of the removed obligors who get the lowest scores with Model 1, so that this percentage is equal to the default rate obtained from the remaining obligors set. In practice, this rule is almost exactly the same as the previous rule (i.e., according to average score of defaulters) and estimated parameters are also almost identical. For this reason, it is not presented as another model.

Model 5: *"Unspotted obligors have repaid, the rest have been written off".* Removed obligors are assigned a non-default tag only if they have a "clean" rating track in the bank and in the system. This means that only obligors with a grade greater or equal to "3" in the maximum rating of the system or in the previous rating (June 1999) are considered defaults, the remaining obligors are considered non default. Other variables such as outstanding debt, guarantees, etc., are not considered in order to discriminate assumed defaulters. The model is run on the increased database.

Table 4: Statistics of the 5 models

Model	Default or not	Frequency	Default rate
Models 1 and 2	Non-default	12,626	10.69%
	Default	1,350	
Model 3	Non-default	15,294	8.83%
	Default	1,350	
Model 4	Non-default	15,079	10.38%
	Default	1,565	
Model 5	Non-default	15,090	10.0%
	Default	1,554	

Results of the models

Table 5 shows the parameters estimated by the five models. There are differences in the estimated values of the parameters and even coefficients of different sign. Coefficients in Models 1 and 3 to 5 are the most similar and results of Model 2 are markedly different. In general, coefficients have the expected signs. Annex 2 shows a priori univariate relationships between explanatory variables and default rates.

The coefficients for rating grades have the expected signs: lower rating grades decrease the probability of default. There is a positive relationship between the number of creditor banks that an obligor has and the probability of default, although obligors with one creditor institution are riskier than those with two.

Table 5: Default is the explained variable

Explanatory Variables	Model 1	Model 2	Model 3	Model 4	Model 5
Intercept	**1.76*** **	0.35	**2.36*** **	**2.15*** **	**2.1*** **
Previous_Rating 0	**-0.44** **	-0.04	**-0.8*** **	**-0.45*** **	**-0.57*** **
Previous_Rating1	-0.29*	-0.19	**-0.6*** **	-0.23	**-0.42*** **
Previous_Rating 2	-0.26	-0.13	**-0.57*** **	-0.2	**-0.42** **
Previous_Rating 3, 4 y 5					
Worst_rating 1	**-0.49*** **	**-0.25*** **	**-0.8*** **	**-0.59*** **	**-0.93*** **
Worst_rating 2	0.14	0.13	**-0.18** **	0.07	**-0.31*** **
Worst_rating 3, 4 y 5					
Rating 1	**-0.66*** **	**-0.63*** **	**-0.56*** **	**-0.83*** **	**-0.59*** **
Rating 2					
Banks 1	**-0.56*** **	**0.14* **	**-0.74*** **	**-0.65*** **	**-0.61*** **
Banks 3	**-0.73*** **	-0.11	**-0.81*** **	**-0.8*** **	**-0.72*** **
Banks 5	**-0.62*** **	**-0.2*** **	**-0.64*** **	**-0.66*** **	**-0.58*** **
Banks 7	**-0.49*** **	**-0.16*** **	**-0.49*** **	**-0.52*** **	**-0.47*** **
Banks 9	-0.1	0.01	-0.12*	-0.1*	-0.1
Banks 10					
Institution_type: National public banks	**-0.29** **	**-0.23** **	**-0.22*** **	**-0.37*** **	**-0.27** **
Institution_type: Local banks with foreign capital	**-0.38*** **	**0.24*** **	**-0.48*** **	**-0.46*** **	**-0.38*** **
Institution_type: Private cooperative banks	-0.2	-0.21**	-0.15	-0.23*	-0.14
Institution_type: National private banks	**-0.47*** **	**-0.24*** **	**-0.45*** **	**-0.51*** **	-0.43
Institution_type: Public local banks	-0.09	-0.09	-0.04	-0.14	-0.04
Institution_type: Banks branches of foreign fi	**-0.34*** **	-0.02	**-0.32*** **	**-0.43*** **	**-0.32*** **
Institution_type: Other financial institutions					
Guarantee	**0.16*** **	-0.04	**0.18*** **	**0.2*** **	**0.19*** **
Ldebt_bank	-0.05*	**-0.11*** **	-0.02	-0.05*	**-0.05* **
Ldebt_system	**-0.12*** **	**0.08*** **	**-0.16*** **	**-0.14*** **	**-0.11*** **
Default_percentage	**0.51** **	**0.61*** **	**0.48** **	**0.63*** **	**0.47** **
Ndefault_lines	**0.28*** **	**0.27*** **	0.02	**0.28*** **	**0.23*** **
Ndefault_previos_lines	**0.43*** **	**0.37*** **	**0.12** **	**0.47*** **	**0.4*** **
Pseudo R^2	0.21	0.072	0.194	0.2684	0.2562
Media estimated PD	10.64%	22.86%	8.79%	10.31%	10.25%
% of sample defaulters	10.69%	10.69%	8.83%	10.38%	10.30%

Notes: ***, ** and * are used for 99%, 95% and 90% respectively confidence levels.
Missing values or dummy variables are the base variables of the models

There is a difference in the risk of debtors of different groups of banks, classified by capital ownership (e.g. foreign capital banks, domestic banks, branches of foreign financial institutions and national private banks): domestic private banks get the most favorable coefficient.

The percentage of credit that is backed by collateral or guarantees (home mortgages and pledges) has a positive and significant coefficient in two of the models, indicating that more backing leads to higher default probabilities. Aspects such as relationship lending may be present[16] or banks

[16] see Berger A.N and G.F. Udell (1995) and Berglof E. and Gollier, C. (1997)

may be asking for guarantees to counterbalance increased perceived risk in order to increase recovery amounts, while the existence of guarantees may be lowering the default behavior only partially[17]. The relationship a priori between the frequency of defaulters and guarantees is weak (see Annex 2)[18].

The estimated probability of default decreases with higher amounts of outstanding debt with a bank as well as with the amount of debt in the system. On the contrary, estimated risk increases for obligors showing higher percentage of their total debt in default or more defaulted lines.

The Goodness of Fit is measured with the Pseudo R^2 statistics (Lagrange Multiplier) adjusted by the number of explanatory variables. Model 4 shows the best fit.

V. Validation: discriminatory power and calibration

BCBS's publication *"Studies on the Validation of Internal Rating Systems" (2005)* suggests some statistical tools to validate rating systems. In the present section we show the results of applying the most widely used techniques described in that document to evaluate different aspects of our estimated scoring models.

As indicated, the data provided by a CD to estimate a rating system is necessarily deficient, as banks have access to information of borrower characteristics that are not recorded in the CD. Nevertheless, it is important to test the merits of such models for methodological interests as well as to assess if this kind of model could be developed by supervisors and used as a sort of risk benchmark or to calibrate prudential regulation.

Rating systems and score functions may be seen as classification tools that provide indicators of the future situation of the obligor. There are two aspects to be evaluated about them: discriminatory power and calibration.

The set of obligors that gets better grades by a *good* rating system will show low frequencies of default and the group that is allocated to worse grades will show higher frequencies of default. Thus, a rating system has more discriminatory power the more discrimination or greater difference is found between the distribution of scores for non-default and default obligors (ability to discriminate good obligors from bad ones).

We present some statistical measures of ***discriminatory power***. Absolute measures of discriminatory power of a rating system have a limited meaning. Rather, they are used to compare among rating systems and to get an idea of the adjustment of the model, for example, to detect statistical "noise".

The measures used in the present document, namely the Accuracy Ratio, Receiver Operating Characteristics and Pietra Index are those frequently used by banks to evaluate the discriminatory power of rating systems.

Checking the ***calibration*** of a rating system is different from checking its discriminatory power. To assess calibration, obligors must be separated into rating grades considering their scores. A

[17] A similar result was obtained by Schechtman et al (2004) when they estimated PDs for corporate debtors in Brazil and borrower's guarantees resulted statistically insignificant.
[18] In the case of smaller consumer credits the relationship between the secured portion of the credit and the frequency of default is clearly negative.

system with a correct calibration would show similar forecast PD and default rate for obligors belonging to the same rating grade.

In practice, estimated PDs will differ from observed default rates. These deviations may be due to random factors or they may occur systematically. The second case indicates that the model needs improvement. In the present document the Binomial test and Chi-squared test are considered to evaluate the quality of PDs estimated from the rating system.

We will choose a model according to discriminatory power and then test calibration for that model only.

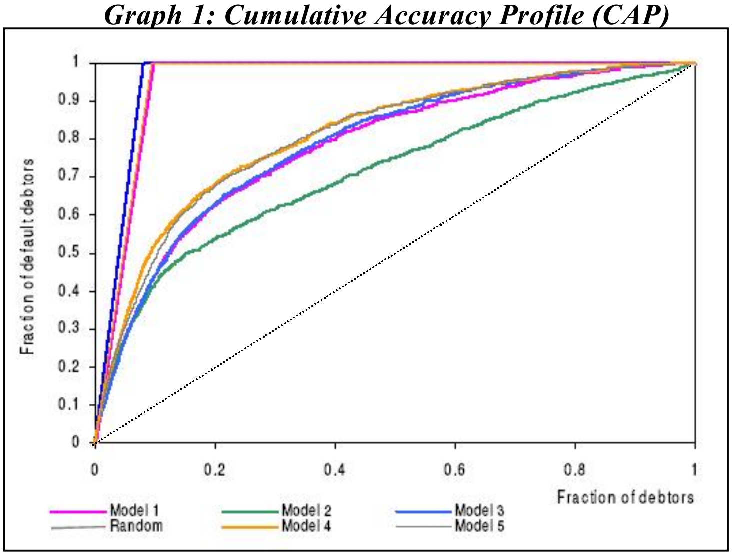
Discriminatory Power

- **The Cumulative Accuracy Profile (CAP) and the Accuracy Ratio (AR):** The Cumulative Accuracy Profile is also known as the power curve or the Lorenz curve. It shows the cumulative percent of observed default obligors attributed to a ranking of observations ordered by their scores. Graphically, the CAP curve is determined by plotting the cumulative percentage of all borrowers on the horizontal axis, from riskiest to safest according to their score, and the corresponding cumulative percentage of defaulters on the vertical axis.

Ideally, the rating process would give all defaulters the lowest scores, then, the CAP curve would rise linearly at the beginning before becoming horizontal. Thus, the steeper the CAP curve at the beginning, the more accurate the rating process.

The other extreme would be a purely random model, which would not have any discriminatory power. The expected CAP curve would, in this case, be the diagonal, as a fraction X of debtors will contain X percent of defaulters. The more concave the CAP, the better the discriminatory power of the rating model, as a more concave curve would be closer to the ideal model.

Graph 1: Cumulative Accuracy Profile (CAP)

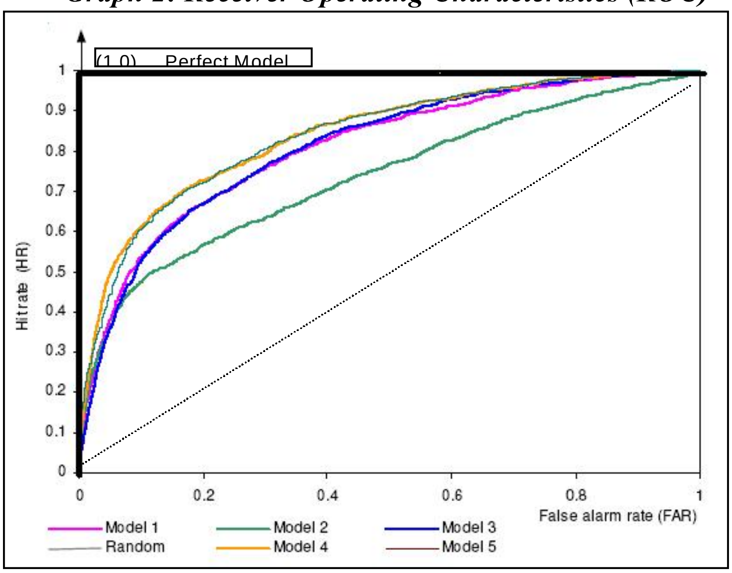


The summary index of the CAP is the Accuracy Ratio (AR), which is based on the Gini coefficient of the CAP. In this ratio the numerator is the area between the CAP curve and the diagonal (the random model) and the denominator is the area between the ideal model and the diagonal (see Annex 2). A rating system is more accurate the closer the AR is to one.

Comparing the rating models estimated in the present document, the CAP curve of the rating system derived from Model 4 is the most concave and has a higher AR. We will make some qualifications on these results below.

- **Receiver Operating Characteristics (ROC):** One of the features of a good rating system is that it has as high a "hit rate" as possible (correct classification of a borrower as a potential defaulter) and at the same time as low a "false alarm rate" as possible (incorrect classification of a creditworthy borrower as a potential defaulter). The ROC curve is a concept related to these two rates and also related to the CAP curve. To construct the ROC curve, the hit and false alarm rates are calculated for every score, taking each level of score as a cut off value for granting credit (see Annex 2). A rating system performance is better the steeper the ROC curve and the closer to the point (0;1). The ROC curves for the models are presented in Graph 2. Using this criterion, the rating system derived from Model 4 has the best performance. We will make some qualifications on this result below.

Graph 2: Receiver Operating Characteristics (ROC)



The area under the ROC curve is measured by the ROC index. The value of this index ranges from 0.5 for a random model (the ROC curve is the diagonal) to 1 for an ideal model.

- **Pietra Index.** The Pietra index considers the greatest triangle area that can be drawn between the ROC curve and the diagonal (Annex 2).

These metrics for the discriminatory power of the estimated models are presented in Table 6. The rating system estimated by Model 4 shows the best performance considering these three indices. In particular, it is clearly better off than Model 2, which also tries to correct the bias arising from removed obligors. We make some qualifications about these results next.

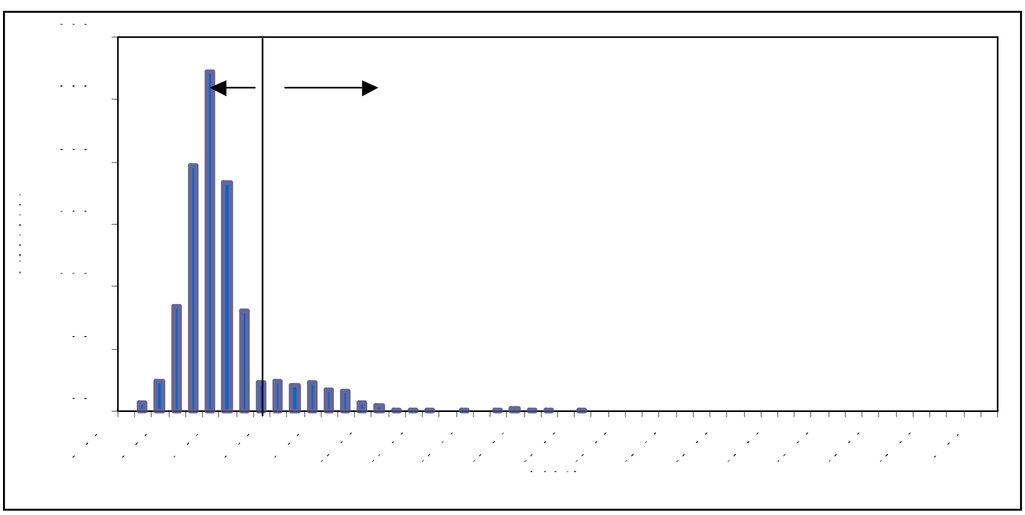
Table 6: Statistical measures of discriminatory power considering all the obligors

Commercial Debtors	*Accuracy Ratio (AR)*	*Receiver Operating Characteristic (ROC)*	*Pietra Index*
Model 1	62.3%	81.2%	0.168
Model 2	47.4%	73.7%	0.136
Model 3	62.7%	81.4%	0.169
Model 4	**68.8%**	**84.4%**	**0.190**
Model 5	68.1%	84.0%	0.188
Max (greater discriminatory power)	100%	100%	>
Min (less discriminatory power)	0%	50%	<
References			
Studies on the Validation of IRB (Basel Committee)	50-80%		
Benchmarking Quantitave Default Risk Models: A validation methodology (Moody´s)	50-75%		

Qualification on the results for Model 4

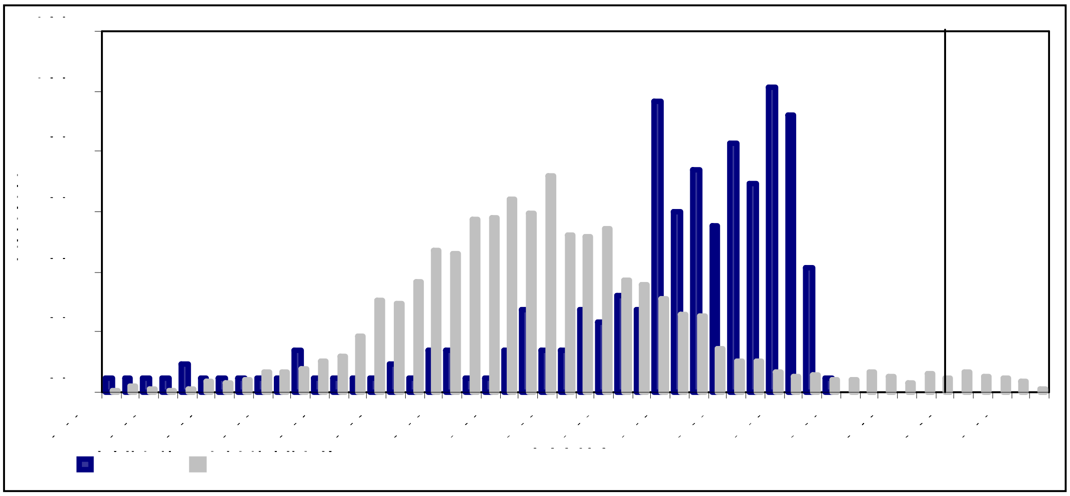
To estimate Models 4 to 6 we have reincorporated removed obligors, assigning a “default” or “non-default” classification to them, following different rules. In doing so, we are actually affecting performance indicators. For Model 4, performance indicators are improved as a result of assigning default tags to those obligors that get the lowest scores with Model 1. Graph 3 shows the division made with Model 1 (obligors that get a default tag are those with Model 1 scores lower than the average score of defaulters that remain in the CD).

Graph 3: Rule of assignation of default and non-default classifications to removed obligors according to scores using Model 1



As estimates for Model 4 are different from those for Model 1, after running Model 4 the distribution of scores for defaulters and non-defaulters in the group of removed obligors also changes. This is shown in Graph 4.

Graph 4: Distributions of Defaults and Non-Defaults in the group of removed obligors, according to their scores using Model 4



The separation between both distributions is not exact now, as a consequence of the difference between Model 1 and Model 4. It can also be seen that re-incorporated obligors tend to get higher scores (indicating less risk) in Model 4 than in Model 1. This, we argue, is a result of the rule being too conservative. As we have explained earlier, there are reasons to believe that removed obligors are mostly repayments, sales or errors, rather than losses. In other words, we believe that the rule assumes a higher default rate than the real one for removed obligors, and thus such rule means a higher probability of assigning a non-default tag to a defaulter than the contrary (assigning a default tag to a non-defaulter), which is acceptable from a supervisory point of view although it is not so good for the sake of the model

As described, participating and disappearing obligors have similar characteristics (Annex 1). Hence re-incorporated obligors in Model 4 will tend to be ranked almost in the same order as in Model 1. Then the technique implies incorporating a sub-set of observations that, by construction, have a close-to-perfect accuracy ratio, even after estimating Model 4. Thus, we are most probably increasing the accuracy ratio and other discriminatory power statistics for the whole sample. This is not clear at all with other rule assumptions.

Trying to take account of this effect is a challenge. In the traditional case of selection bias it can be shown that (i) the bias is considerable when testing discriminatory power, and (ii) the bias might be positive or negative, i.e., there is no simple rule to take account for this bias[19]. In our case, when the bias has no side or is both sided, it would be even more difficult to find such a correcting rule, especially taking into account the low quality of explanatory variables to model the selection process .

Thus, we have calculated "hard" discriminatory power statistics for the different models using only the set of observations of obligors that are present at both points in time. If these statistics improve, most probably the whole power of the model has improved. If the hard statistics do not show any improvement, the difference between the estimated parameters should be measured in order to evaluate if it is worth re-incorporating obligors that disappear.

[19] Kraft, Kroisandt and Müller (2004)

Table 7: Statistical measures of discriminatory power considering obligors present in Dec-99 and Dec-00.

Commercial Debtors	*Accuracy Ratio (AR)*	*Operating Characteristic*	*Pietra Index*
Model 1	62.3%	81.2%	0.168
Model 2	47.4%	73.7%	0.136
Model 3	61.7%	80.8%	0.167
Model 4	**62.2%**	**81.1%**	0.168
Model 5	62.0%	81.0%	0.167
Max (greater discriminatory power)	100.0%	100.0%	>
Min (less discriminatory power)	0.0%	50.0%	<
References			
Studies on the Validation of IRB (Basel Committee)	50-80%		
Benchmarking Quantitave Default Risk Models: A validation methodology (Moody´s)	50-75%		

Again, discriminatory power statistics show that Model 2 looks like a poor solution to take account for the bias due to removals, as values deteriorate markedly. Among the other models, the "hard" statistics (calculated using only obligors who remain but with the models estimated with the increased samples) show that Model 4 is able to keep the same values of Model 1, while the rest show a slight deterioration. As expected, "soft" statistics (measured on the increased sample) are clearly better for all Models but Models 2 and 3, always in comparison to Model 1.

Therefore, Model 4 is chosen to classify the obligors into grades and test its calibration as well as to conduct out-of-sample tests.

If the hard statistics did not show any improvement in any of the models, the difference between the estimated parameters should be measured in order to evaluate if it is worth re-incorporating obligors that disappear. In the case of our selected Model 4, even when the "hard" version of discriminatory power statistics are very similar to those of Model 1, still the parameters show that the characteristics of the obligor receive different consideration under the two models. For example, an obligor in a foreign bank branch whose previous rating was 3 (six months ago), but is rated 1 at the starting point and is also rated 1 in other banks (two other banks have claims with her), has no guarantees, the outstanding debt with the bank is $500,000 and $700,000 overall in the system, will get an estimated 0.3% probability of default according to Model 1 and a third of that, 0.1%, according to Model 4. On the other hand, let us consider an obligor in a cooperative bank, whose rating 6 months ago was 2 with the bank, as it is at the starting point, while the worst rating is 4 in the system, has eight other creditor banks and one line of credit has been reported as defaulted in the system (at the starting point and 6 months earlier). There are guarantees covering 100% of the claim, outstanding debt with the bank is $ 50 million and with the system $ 75 million. This obligor will get an estimated PD of 23% with Model 1, while it will be of 30% with Model 4 (30% increase).

Calibration

The rating system estimated by Model 4 has been selected due to its econometric fit and its discriminatory power. Calibration should be checked to have a complete validation of the model. A rating system is calibrated when its forecast PDs are similar to realized defaults rates for each rating grade. A borrower grade is defined as an assessment of borrower risk on the basis of rating criteria from which estimated PDs are derived.

Basel II establishes minimum requirements for the design of a rating system. Among the requirements for the rating structure of corporate, sovereign and bank exposures it is indicated that banks must distribute their exposures across grades with no excessive concentrations. In order to meet this requirement it must have a minimum of eight grades: seven for non default debtors and one for default.

Hence, we have established eight rating categories based on the results of the scoring Model 4. As this model uses a Probit statistical method, PDs associated to each borrower are directly calculated. At this stage, we do not consider different types of credit, so the PDs represent borrower's risk[20]. We have constructed rating grades with the results for commercial debtors of the whole financial system, not with borrowers of a particular bank.

Thus, commercial debtors with similar probabilities of default (or scores) were grouped into eight rating grades. The technique to determine the ranges for each grade was mainly based on graphical comparison of moving average realized default rates and estimated PDs.

Calibration tests

The document *Studies on the Validation of Internal Rating Systems* (2005)[21] puts forward different tests to evaluate the calibration of a rating system. The most widely used, the Binomial and Hosmer-Lemeshow (H-L) tests, assume that default events are independent. While the Binomial test examines each rating category separately, the H-L test considers the quality of all rating categories at the same time (Annex 3).

The mentioned document indicates that the Binomial test is the most powerful test at a fixed level for testing the adjustment of PDs. Also, it is pointed out that independence of default events is a powerful assumption and that, empirically, some small correlations are usually observed. For the present study we have decided to take a conservative approach and apply the Binomial test. By not considering that default events are correlated we are increasing the probability of unjustifiably rejecting the null hypothesis (Ho= estimated PDs are correct). So if the Binomial test indicates that rating grades are correct, this is a robust result.

In the Binomial test, the observed number of defaulters in each rating grade, N_D, must be compared to the k^* statistic related to a certain confidence level. If $N_D < k^*$ the null hypothesis of correct PDs cannot be rejected at that confidence level.

Table 8 shows that the rating categories constructed with estimated PDs are correct according to the Binomial Test.

Assuming again independence of default events, the Hosmer-Lemeshow statistic is calculated. The result is a non rejection of the null hypothesis that the average PDs of the rating grades are the true probabilities.[22]

[20] This dimension is going to be considered in future research.
[21] BCBS Publications No. 14, May 2005.
[22] The null hypothesis is not rejected at a 95% confidence level, using the Hosmer-Lemeshow statistical with a value of 15.36 (Pr(χ^2_9) = 0.0815).

Thus, the calibration of rating grades constructed from the rating system of Model 4 seems to be acceptable, considering each category individually and all of them simultaneously.

Table 8: Calibration measures of Model 4

Rating Categories		*Debtors*		*Defaults*				*Calibration Result*
N°	*Range of PD (%)*	*N*	*(%)*	N_D	N_D/N (%)	*Estimated media PD (%)*	*k* (99% confidence level)*	
1	**0-1.5**	1686	10.13	**10**	0.59	1.01	**27**	Correct
2	**1.5-2.7**	3101	18.63	**55**	1.77	2.12	**85**	Correct
3	**2.7-3.75**	2618	15.73	**75**	2.86	3.19	**105**	Correct
4	**3.75-4.75**	1815	10.90	**64**	3.53	4.24	**97**	Correct
5	**4.75-5.6**	1254	7.53	**78**	6.22	5.16	**83**	Correct
6	**5.6-6.31**	859	5.16	**64**	7.45	5.94	**67**	Correct
7	**6.31-18.3**	3241	19.47	**322**	9.94	9.47	**346**	Correct
8	**18.3-100**	2070	12.44	**897**	43.33	42.96	**942**	Correct

Performance of Model 4, out-of-sample test

To validate the use of a statistical model in a rating process, it is necessary to check its performance including "out-of-time" and "out-of-sample" performance tests. Thus, the number of exposures in the sample used to estimate risk components must be sufficient to provide confidence and robustness to the estimates while it is at the same time desirable to keep a number of observations out of the sample to carry out these tests[23].

To estimate PDs a sequential sample of the 75% of bank-debtors was taken from the CD. The out-of-sample test would be performed on the rest. The present out-of-sample test consists in calculating PDs using the estimated coefficients from Model 4 and the explanatory variables of out-of-sample bank-debtors. Then, the rating system of the out-of-sample debtors is constructed.

Table 9 presents the results of discriminatory power tests of the rating system on the out-of sample exercise. In this case out-of-sample debtors have a lower performance in terms of discriminatory power compared to in-sample debtors (AR is about 62% for out-of-sample debtors while it was 68% for in-sample debtors). However, this performance is still acceptable taking into account established practice and literature. Out-of-sample calibration is shown in Table 10, where it can be seen that risk grades perform well also for out-of-sample debtors.

Table 9: Out-of-sample tests of discriminatory power

Commercial Debtors	*Accuracy Ratio (AR)*	*Receiver Operating Characteristic (ROC)*	*Pietra Index*
Model 4	62.9%	81.4%	0.172
References			
Studies on the Validation of IRB (Basel Committee)	50-80%		
Benchmarking Quantitave Default Risk Models: A validation methodology (Moody´s)	50-75%		

[23] Basel 2 indicates this requirement as well, see International Convergence of Capital Measurement and Capital Standards, Basel Committee (2004), paragraph 251.

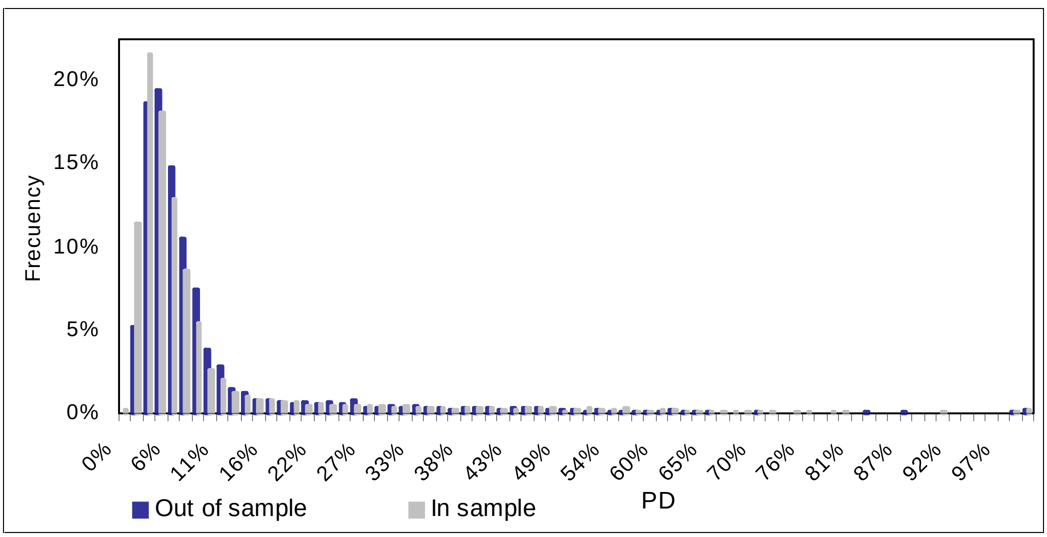
Table 10: Out-of-sample calibration statistics for Model 4

Rating Categories		***Debtors***		***Defaults***				***Calibration Result***
N°	***Range of PD (%)***	***N***	***(%)***	N_D	N_D/N (%)	***Estimated media PD (%)***	***k* (99% confidence level)***	
1	**0-1.5**	230	4.13	**1**	0.43	1.09	**6**	Correct
2	**1.5-2.7**	777	13.94	**16**	2.06	2.14	**26**	Correct
3	**2.7-3.75**	926	16.62	**29**	3.13	3.21	**42**	Correct
4	**3.75-4.75**	719	12.90	**18**	2.50	4.19	**43**	Correct
5	**4.75-5.5**	477	8.56	**19**	3.98	5.13	**36**	Correct
6	**5.5-6.32**	413	7.41	**29**	7.02	5.91	**36**	Correct
7	**6.32-18.3**	1368	24.55	**124**	9.06	9.33	**153**	Correct
8	**18.3-100**	662	11.88	**270**	40.79	39.14	**288**	Correct

Finally, Graph 5 shows the distribution of PDs from in sample and out-of sample observations. PD distributions are similar though the distribution of in-sample debtors with lower PDs is slightly more concentrated on lower values (to the left).

In conclusion, the rating model has an acceptable performance considering out-of-sample tests.

Graph 5: Comparison of out-of-sample and in-sample PD distributions



VI. Conclusions

The conclusions arising from this study concern three subject matters: (i) the uses of a model like the one developed here, which is of interest mainly to regulators and supervisors; (ii) technical conclusions concerning the problem of disappearing obligors or similar difficulties with databases, which are of interest to regulators and the industry as well and (iii) a practical application of validation methodologies.

The results of the model are very good in spite of the limited selection of explanatory variables, which refer basically to outstanding amount of debt and its distribution (in the institution and in the system), guarantees and rating grades (at present, in other institutions or in the near past). We have not used information on activity sector, financial ratios or other balance sheet information, nor other variables in the CD which unfortunately show low quality. Clearly, a bank could resort to a much broader set of data about its debtors; a model like the one developed in the study is not a archetype to be used by the banks.

However, modeling credit risk on such a limited set of CD information seems to be a powerful tool, especially for regulators and supervisors, according to the performance statistics obtained in this study. We can foresee different uses:

(i) To evaluate regulation and make informed decisions about it. For example, the level of mandatory provisions could be compared to expected default rates (which can be obtained from PDs, and recovery rates), the overall level of capital requirement for credit risk could be evaluated by developing a portfolio model of credit risk that uses the resulting PDs. It can also be used, in countries like Argentina where there is a standardized rating system set up by Central Bank regulation, to assess changes to that system. In our case, for instance, it can be concluded that the first rating grade should be broken down into smaller, more precise grades, to avoid the concentration of claims and to split up the wide range of risks that are assigned to the same bucket.

(ii) To use the results as a benchmark to compare bank models.

(iii) To make adjustments to credit bureaus, in particular, to avoid any loopholes referring to the behavior of the individuals.

In relation to this last topic, we have analyzed the apparently innocuous decision of taking out of the sample those debtors that do not remain in the database when the reasons for the removals are not known and cannot be modeled. We have shown that this may introduce a bias that cannot be ignored and is difficult to correct without distorting performing statistics. Thus, it is of primary importance to ensure that credit risk databases rule out any possible loopholes.

This last point is of interest to the industry, to supervisors and to researchers, since it is frequent to find models in which a group of obligors have been taken out of the sample because their information shows some drawback such as incompleteness, poor quality, or any other difficulty. This could lead to an undesired bias or result in “cherry picking” the best obligors in order to improve the validation measures of credit scoring models used by banks. Thus, this practice should be avoided for the sake of the precision of the model and should not be accepted by supervisors.

Additionally, this study applies validation techniques as proposed by the BCBS regarding calibration and discriminatory power. The performance statistics are very good in spite of the limitations of the explanatory variables. The understanding of validation tools, such as those presented in this document, is necessary for supervisors that would evaluate bank credit scoring models.

Annex 1: Table of Characteristics of participant and removed obligors

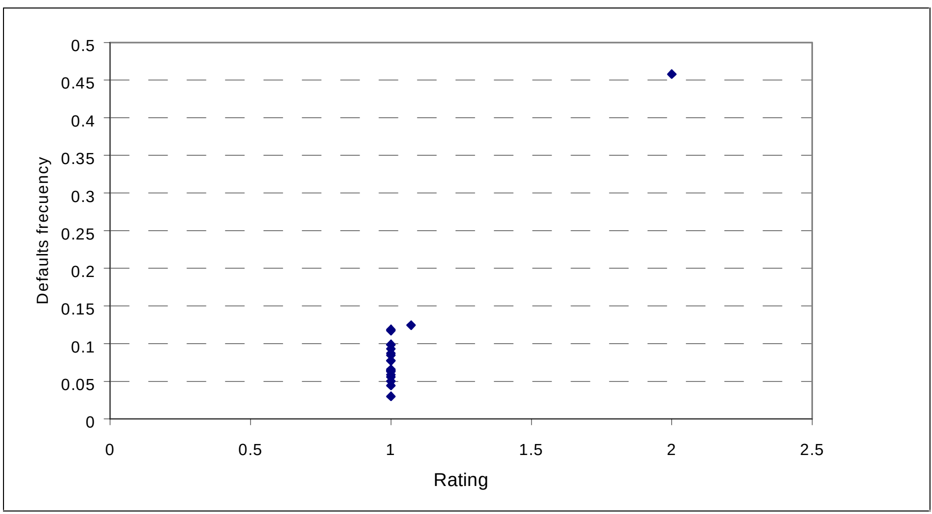
Number of obs/percentage	Participants	Non Participants
	Previous_Rating	
0	1,652	590
	11.8%	22.1%
1	11,706	1,990
	83.7%	74.6%
2	502	81
	3.6%	3.0%
3	58	5
	0.4%	0.2%
4	30	1
	0.2%	0.04%
5	36	1
	0.3%	0.04%
	Rating	
1	13,231	2,533
	94.6%	94.9%
2	753	135
	5.4%	5.1%
	Worse_rating	
1	11,624	2,260
	83.1%	84.7%
2	1,377	209
	9.8%	7.8%
3	414	83
	3.0%	3.1%
4	338	74
	2.4%	2.8%
5	231	42
	1.7%	1.6%
	Banks	
1	1,652	579
	11.8%	21.7%
3	3,903	541
	27.9%	20.3%
5	2,996	401
	21.4%	15%
7	1,837	315
	13.1%	11.8%
9	1,298	265
	9.3%	9.9%
10	2,298	567
	16.4%	21.3%
	Sector	
Private Sector-	13,979	2,391
national residents	99.96%	89.6%
Private Sector-	5	277
foreign residents	0.04%	10.4%
Total	13,984	2,668
	100	100
Medias	***Participants***	***Non Participants***
colateral	33%	19%
ldeuban	6.39	6.65
ldeusis	7.48	7.97
porcdef	1.3%	1.7%
nlindef	7.6%	10.6%
nlinprevdef	5.6%	5.6%

Annex 2: Explanatory variables and a priori univariate relationships with default rates

The definitions of the explanatory variables are given below, in some cases together with scatter plot graphs that illustrate the relationship between default frequency. These are a priori univariate relationships, i.e., there is no control for other variables. To draw each graph, the whole set of observations are first ordered by the x-axis variable. Then, observations are divided into 20 quartiles and the average value of the x-axis variable as well as the default rate is calculated, and plotted, for each quartile. In other words, each point in the graph accounts for 5% of observations. In this way, the graphs show not only the a priori relationship between the variables and the default rate but also the distribution of the observations in different values of the x variable.

Default: Class variable to identify the defaulting debtors. This variable assumes value 0 if the debtor is rated in grades 1 or 2 and it takes value 1 if the debtor is in grades 3, 4 or 5 in December 2000.

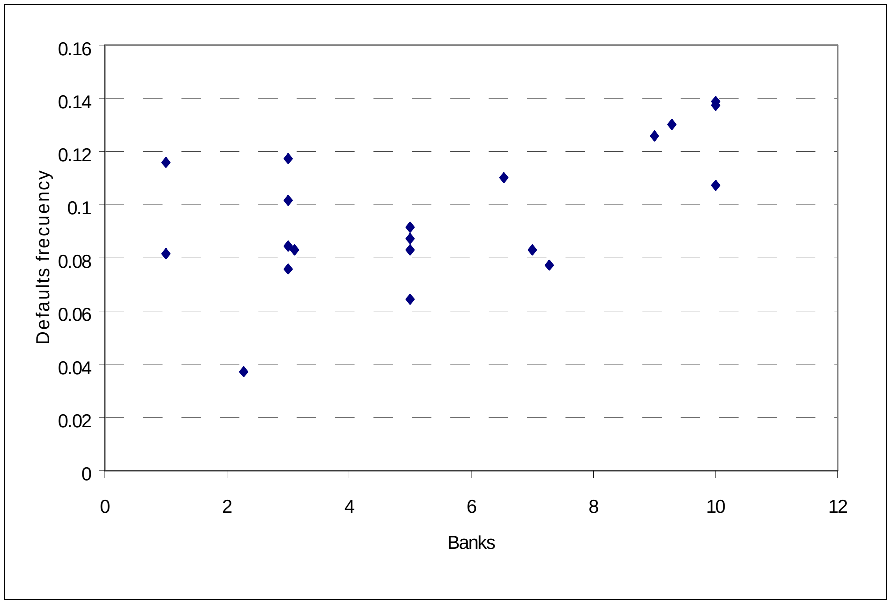
Rating: Class variable that records the rating grade of the debtor (at obligor-institution level). It assumes values 1 or 2 if the obligor is rated 1 or 2, respectively, in December 1999.



Previous_Rating: Class variable that considers the rating grade at obligor-bank level. This variable takes values from 0 to 5 representing the rating grade of the debtor in June of 1999. It takes value 0 if the obligor was not registered in June 1999.

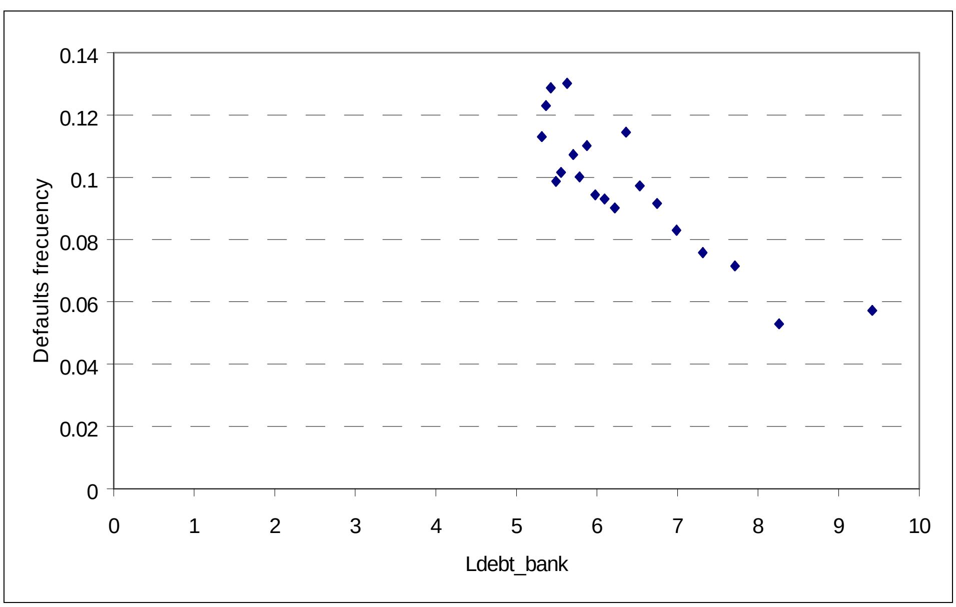
Worst_rating: Class variable that assumes values from 1 to 3 representing the worst (highest) obligor-bank rating grade in December 1999. A value of 3 identifies debtors rated in grades 3 to 5.

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 ten or more institutions.



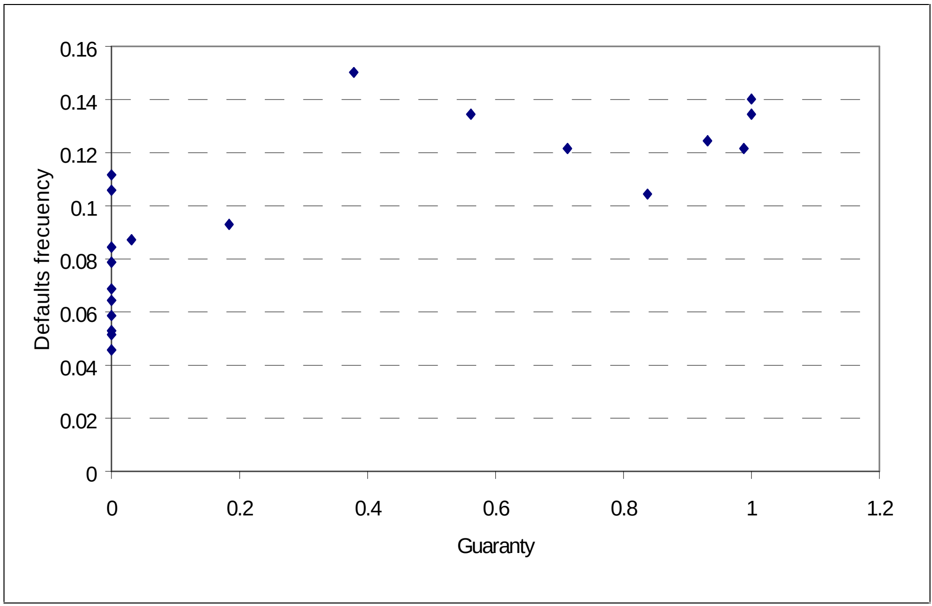
Institution_type: Class variable that classifies financial institutions according to capital origin. This variable is broken down into 7 excluding dummies: national public banks, domestic banks with foreign capital, private cooperative banks, national private banks, public local banks, branches of foreign financial institutions and other financial institutions.

Ldebt_bank: the logarithm of the outstanding amount of the claims of a specific financial institution with certain obligor as of December 1999.



Ldebt_system: the logarithm of the outstanding amount of overall claims of the banking system with certain obligor as of December 1999.

Guarantee: Percentage of the claims that are guaranteed as of December 1999.



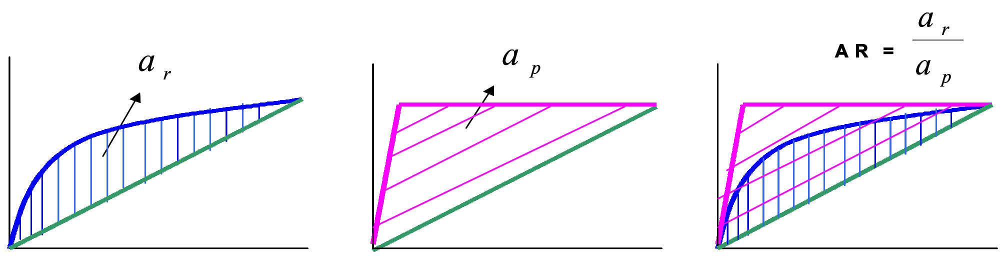
Default_percentage: proportion of the claims of the banking system with certain obligor that is reported in default as of December 1999.

Ndefault_lines: Number of credit lines of the obligor classified as in default in the banking system as of December 1999.

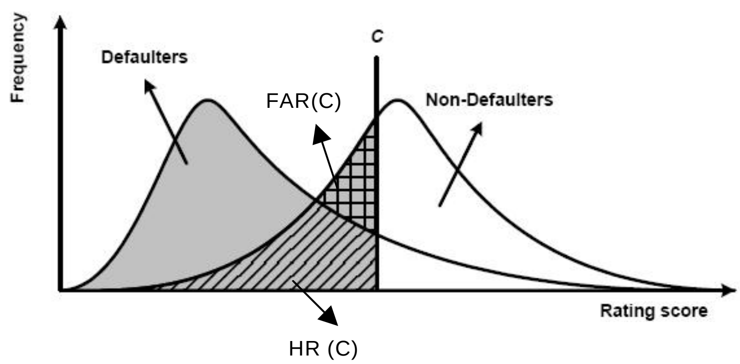
Ndefault_previous_lines: Number of credit lines of the obligor classified as in default in the banking system as of June 1999.

Annex 3: Statistical measures of discrimination and calibration

Accuracy Ratio (AR): It is defined as the ratio of the area a_r between the Cap of the rating model being validated and the CAP of the random model, and the area a_p between the CAP of perfect rating model and the CAP of the random model. This index takes values between 0 to 1, the better that rating model the closer to 1.



ROC Index: The construction of ROC curve considers the distributions of rating scores for defaulting and non defaulting debtors. Distributions showed in the following graph should be as separate as possible:



Considering a cut-off score C, to classify debtors as potential default if their scores is less than C and non potential default if their score is more than C. Given C, the hit rate and false alarm rate that are measuring corrected and wrong predictions of defaulters.

The hit rated, HR (C), and the false alarm rate, FARC (C), are the following:

$$HR(C) = \frac{H(C)}{N_D} \qquad FARC(C) = \frac{F(C)}{N_{ND}}$$

Where, H (C) is the number of defaulters predicted correctly from the cut-off C and N_D is the total number of sample defaulters. F(C) represents the number of false alarms, that is number of non defaulters classify incorrectly as defaulters by using C as a cut-off score and N_{ND} are the total number of non defaulters debtors.

The ROC curve is a scatter plot line constructed computing the hit rate and the false alarm rate for the cut-off values contains in the range of the rating score estimated.

Pietra Index: Geometrically, this index can be estimated as the maximum area of a triangle inscribed between the ROC curve and the diagonal. Equivalently, the Pietra Index can be

$$Pietra\ \ Index = \sqrt{2}/4 \max_c \left| HR(C) - FARC(C) \right|$$

calculated as the maximum distance between the ROC curve and diagonal. In the case of concave curve it can be calculated as:

Binomial Test: This is a test that considers the calibration of each rating category established by the banks from their rating system. This test is constructed with the assumption that default event are independent so Binomial distribution is considered.

The null hypothesis (H_0) is that PDs of the rating system are correct. For a confidence level q (e.g. 99%), the null hypothesis is rejected if the number of default debtors in a rating category N_D is greater than the critical value $k*$.

$$k* = \Phi^{-1}(q)\sqrt{nPD(1-PD)} + nPD$$

Hence, Binomial test is simple but it is founded on the assumption of independent default events. Empirically, it is known that defaults are correlated with small coefficient of correlation so the statistic changes if that correlation is considered.

The presence of correlation means that large deviations of estimated PD and default rates are probable. Therefore this makes it impossible even for large number of debtors to get estimated PDs closer to observed default rate. The law of large numbers does not work because of the existence of correlation.

However, as critical values of PD tests that incorporate correlation tend to be greater than critical values of binomial test (with independence assumption), an application of the latter would be conservative in presence of correlation. Binomial test leads to earlier rejection of H_0 in a context of correlation. The true size of type I error (reject H_0 when it is true) will be higher than the nominal level that the test indicated. Therefore testing the calibration of each rating category with Binomial test, as is done in the present document, implied being conservative in the sense that if there is a small correlation on default events we could be rejecting H_0 when is true, saying that estimated PDs are not correct when they really are.

Chi-squared test (or Hosmer-Lemeshow)

This test assumes also independence of default events, but it test all rating categories simultaneously opposite to Binomial test that is a test for each rating category.

If the forecasted probabilities of default where: $p_0, \ldots, p_k$ for k rating categories, n_i the number of obligors in the rating i and θ_i the number of default obligors in rating i, the statistic of the test is defined as:

$$T_k = \sum_{i=0}^{k} \frac{(n_i p_i - \theta_i)^2}{n_i p_i (1 - p_i)} \sim \chi^2_{k+1}$$

When $n_i \to \infty$, by the central limit theorem, the distribution of the T_k will converge to a χ^2_{k+1} distribution if all the p_i were the true default probabilities.

Bibliography

Basel Committee on Banking Supervision, *"Studies on the Validation of Internal Rating Systems"*, Working Paper N° 14, may 2005.

Basel Committee on Banking Supervision, *"International Convergence of Capital Measurement and Capital Standards. A Revised Framework*", June 2004.

Basel Committee on Banking Supervision, *"Studies on the Validation of Internal Rating Systems*", Working Paper N° 14, may 2005.

Balzarotti V., Castro C. and Powel A., "*Capital Requirements in Emerging countries: Calibrating Base II using Historical Argentine Credit Bureau Data and Credit Risk*+", 2004, Business School Working Papers, Universidad Torcuato Di Tella.

Balzarotti V., Castro C. and Powell A., "*Reforming capital Requirements in Emerging Countries*", 2002 , Business School Working Papers, XIX, Universidad Torcuato Di Tella.

Barron, J.M. and Staten M., "The Value of Comprehensive Credit Reports: Lesson for the U.S. Experience Summary", Word Bank, Mimeo, 2000.

Berger A.N and G.F. Udell, *"Relationship Lending and Lines of Credit in Small Firm Finance"*, Journal of Business, 68(3), July 1995, pages 351-81.

Berglof E. and Von Thadden, "*Short Term Versus Long Term Interest: Capital Structure with Multiple Investor"s,* Quarterly Journal of Economics, 109, 1994, pages 1055-1084.

Crook, John, "*Adverse Selection And Search In The Bank Credit Card Market",* Credit Research Centre, University of Edinburgh, June 2002.

Dubin, J. and Rivers, D., *"Selection Bias in Linear Regression Logit and Probit Models"*, Sociological Methods and Research, Vol. 18, 2&3, November 1989/Febrery 1990.

Hand D.J. and Henley W.E., *"Statistical methods in Consumer Scoring: A Review, Journal of the Royal Statistical Society. Series A (Statistics in Society)"*, Vol. 160, N°3 , 1997, 523-541.

Kiefer N. M. and Larson E. C., *"Specification and Informational Issues in Credit Scoring"*, Office Comptroller of the Currency, E&PA Working Paper 2004-5, December 2004.

Moody´s Inverstors Service, *"Benchmarking Quantitative Default Risk Models: A Validation Methodology"*, march 2000.

Schechtman, R., Salomão García, V., Mikio Koyama, S. and Cronemberger Parente, G., *"Credit Risk Measurement and the Regulation of Bank Capital and Provision Requirements in Brazil-A Corporate Analysis*", Working Paper Series 91 Banco Central Do Brasil, December 2004.

Verstraeten, G. and Van den Poel, D., "The Impact of Sample Bias on Consumer Credit Scoring: Performance and Profitability", Working Paper Universiteit Gent, march 2004.