The Importance of Borrowers' History on Credit Behavior: The Mexican Experience

José L. Negrin[♦]

BANCO DE MÉXICO

2004

Abstract

Credit sharing information mechanisms represent the institutional answer to the asymmetric information problems inherent to credit markets. It is generally accepted that sharing information is beneficial for the participant institutions, however, there are few studies that have measured the impact of past behavioral information on risk analysis. Applying a Probit model to the micro level database gathered by the Mexican Public Registry of Credit Information we find that historical variables, like previous defaults and previous missing payments are highly significant in explaining the probability of default. In particular, having defaulted a loan in the past, increases current loan's default probability in 30 percentage points. We also find that the longer the borrower has been in the market and the larger the loan, the less likely it is that the current loan will be defaulted on. Additionally, we measure the effects of macroeconomic fluctuations over individual loans' probability of default; we find that inflation significantly increases it while economic growth reduces it. Our results imply that effort should be exerted to develop more complete databases on individuals' past behavior. This is particularly relevant in the Latin American context were the credit sharing industry is not very developed.

JEL classification: G23; G14; D82.

[♦] Economic Researcher, Economic Studies Division, Banco de México. Ave. 5 de Mayo # 6, 3er. Piso. Col Centro, 06059, México DF. Tel. 5227-8606. e-mail jlnegrin@banxico.org.mx

I thank Alfonso Álvarez and Regina Weber for their technical help so as Rodrigo García Verdú, Javier Márquez, Pascual O´Dogherty, Joyce Sadka, Alejandro Werner and the participants in Banco de Mexico's internal seminars for useful comments. I also thank Enrique Bátiz for his help with the estimation. All responsibility remains mine.

1. Introduction

It is well known that credit markets suffer from asymmetric information problems. When a loan is granted, the lender faces adverse selection since it does not know the borrower's intrinsic risk level. Once the loan is provided, the lender faces moral hazard given that it cannot perfectly monitor borrower's efforts to repay the loan. Given these information constraints, lenders make loans decisions based on applicants' average characteristics (Rothschild and Stiglitz, 1976). This gives place to a worsening of payment likelihood and to higher interest rates. Higher rates increase adverse selection because only high risk borrowers will be willing to take them (Stiglitz and Weiss, 1981).

Coordination between lenders by sharing credit information alleviates asymmetric information problems. This is the function of sharing mechanisms, whether they are Public Registries of Credit Information (PRCI) or private credit bureaus. These institutions collect, organize and consolidate information on borrowers' past behavior; such information emerges from the interaction of PRCI associated banks with their respective clients. Sharing mechanisms use this information to build credit reports on individuals, which they provide to lenders. Sharing information has two positive effects on information problems. On the one hand, it mitigates adverse selection by allowing lenders to identify potential clients' characteristics. On the other, it alleviates moral hazard by providing incentives to borrowers to fulfill their credit obligations[1]. These positive effects explain that the industry of credit reporting has expanded rapidly in many countries, including Mexico.

Lenders' coordination is based on the principal that a borrower's past credit behavior is a good indication of its current and future actions. This seems obvious; however, there is little empirical evidence that this is the case. The main goal of this paper is to measure to what extent a borrower's past behavior affects its default likelihood on current loans. Although me cannot separate the moral hazard and adverse selection effects of information sharing, we quantify the impact of past borrower behavior on her current credit performance .

[1] Padilla y Pagano (1997) call this the *disciplinary effect* of information sharing.

For this study we use the individual-loan level database from the Mexican PRCI. Membership in the PRCI is compulsory for all banks in the Mexican financial system; hence the database covers all loans provided by Mexican financial institutions. The database allows us to design an array of historical variables that describe borrowers' past behavior, such as past defaults, past delays, past loan restructures, length of borrowers' participation in the credit market and borrowers' consolidated debt (within the banking system). Although the database does not include balance sheet information- usually considered as the most important determinants of credit risk- it allows us to construct proxy variables for such information. Hence, we analyze the relevance of historical variables on the probability of default (PD) controlling for balance sheet information. Our estimation uses a dichotomous dependent variable (Probit) model.

We find that information on borrowers past behavior is crucial in PD determination. In particular, having previous defaults increases PD in 30 percentage points. The relative importance of this number can be properly seen by considering that the average estimated PD is 3.97%. Similarly, previous delays and past loan restructures significantly increase PD. On the other hand, length of participation in the credit market reduces PD, as consolidated debt and current loan amount do. Interestingly, collateral has little significance in PD determination. This may be due to the legal limitations on the recovery of collateral associated with nonperforming loans. Finally, we find that at the microeconomic level, inflation significantly increases PD while economic growth reduces it. This finding is relevant because it stresses that a healthy credit system needs of a stable macroeconomic environment.

Our results fit in the literature started by Chandler and Parker (1989) and extended by Barron and Staten (2002) and Kallberg and Udell (2002). Chandler and Park use a credit scoring model to test the effect on credit ratings of applications information versus credit bureau information. They find that credit bureau information predicts borrowers' credit risk better. Barron and Staten (2002) also use credit scoring and a simulation mechanism to analyze the influence of positive and negative information on availability of credit to households. They conclude that more information mitigates credit constraints. Kallberg and Udell (2002) introduce an index that summarizes borrowers' credit history in a logistic model to show that voluntary information sharing is useful in determining default.

The contributions of this study are the following. Firstly, the quantification of the effect of past behavior variables in the context of an emerging economy has important policy implications for the performance of credit institutions. It is very likely that one of the reasons for the limited importance of credit in Mexico has been the lack of information. Secondly, the design of several variables that describe borrowers' past behavior allows us to add precise effects to the findings of Kallberg and Udell and Barron and Staten. Thirdly, our estimations are based on information from a PRCI of compulsory participation, as opposed to the voluntary interchanges of Kallberg and Udell (2002). We show that even non-voluntary sharing is useful to determine default[2]. Finally, the measurement of the effect of macroeconomic fluctuations on default probability at the microeconomic level is a contribution to the literature.

This paper is structured as follows. The second section reviews the related literature. The third presents a conceptual description of the database and variable design and of the Probit model. In the forth section we present the data description. The fifth section includes the estimation results. The sixth section discusses some policy issues and concludes.

2. Related literature

Asymmetric information in credit markets increase PD and negatively affects the granting of credit. The institutional answer to these problems has been the formation of mechanisms by means of which banks share information on borrowers' past credit behavior. Sharing mechanisms, whether they are private credit bureaus or PRCI, gather, organize and consolidate information on borrowers provided by associated banks[3]. The bureau then forms credit reports and sends the consolidated information back to the banks upon request[4]. There have been three strands of literature related to information sharing. The first one is theoretical, the second is empirical at the macro level and third one is also empirical but at the micro level. This study fits in the latter strand.

---

[2] It could be thought that forcing credit institutions to participate would reduce the quality of the information reported.

[3] Although our research uses information from a public registry, the analysis is valid for private mechanisms. We sometimes refer indistinctly to the mechanism as PRCI or as bureau.

[4] See Negrin (2001) for a detailed description of the way sharing mechanisms work.

Despite the benefits involved in sharing credit information, not all countries have such sharing institutions[5]. Furthermore, the origin and development of these institutions around the world has not been unique. Sometimes the sharing mechanism has a private, spontaneous, and voluntary origin, like many credit bureaus. In other cases, financial authorities start a PRCI; these institutions usually have coverage that is restricted to financial institutions, whose participation in the mechanism is compulsory[6]. In these cases, private bureaus have often entered markets not covered by the PRCI. In yet other cases, financial authorities have started PRCI even though there already were private bureaus in the market (Miller, 2000).

In terms of the willingness of lenders to share information, Pagano and Jappelli (1993) find that sharing is more likely to occur the larger the population, the greater the level of mobility and the more heterogeneous the individuals are[7]. A related aspect that affects sharing decision is provided by Jappelli (1997) and Padilla and Jappelli (1997). They claim that when there is less competition among lenders, information sharing is more feasible since participants have not fear of loosing their good clients[8].

The information that is shared could be negative (overdue payments, defaults and frauds) or it could also include positive information (for instance, good payment behavior in the past, current debt that is being paid and payment patterns). The type of information shared has effects both on borrowers' and lenders' behavior. Lenders are more willing to share negative information than positive information. Sharing negative information helps lenders to identify high risk borrowers. However, sharing positive information may reduce the rents that lenders can extract from the good clients that they have already identified.[9]

---

[5] See Jappelli and Pagano (1999), Miller (2000) and Negrin (2001).

[6] In the voluntary participation case, Klein (1992) studies the decision to subscribe to the mechanism. A different approach is taken by Pagano and Jappelli (1993); they consider that once a significant proportion of firms has joined the mechanism, all the others will soon follow.

[7] Klein (1992) states that credit bureaus in "the Great Society" play the role of gossip in smaller communities; hence, sharing information mechanisms only emerge in large enough societies.

[8] Even though sharing may intensify competition between lenders, Padilla and Pagano (1997) find that such harsher competition may have a desirable effect for lenders, in addition to the enhancement of borrowers discipline. They claim that, given that competition disallows informational rents, borrowers perform better because they perceive that the lender is not appropriating all the benefits of their effort to repay their loans.

[9] Padilla and Pagano (1999) claim that sharing negative and positive information may have an undesirable effect on borrowers behavior. Revealing just negative information imposes discipline on debtors; in this case, information on the payments missed is the only signal of bad risks that lenders receive. When positive information is shared, the lending decision is based not just on negative information, but also on other

Consequently, some countries' financial authorities, like Australia, have restricted the interchange to negative information.

Apart from these theoretical analyses, there is a strand of literature oriented to measure the effects of information sharing on the amounts of credit at the economy wide level. Miller (2000) and Jappelli and Pagano (1999) perform international comparisons of the mechanisms to share information. Jappelli and Pagano (1999) perform a cross country study to find that sharing information mechanisms have positive effects on aggregate credit provision. A similar approach, although based on data on firms listed in the stock market of the corresponding country, is followed by Galindo and Miller (2001) to show that sharing mechanisms are useful to mitigate credit constraints.

The third literature strand studies the effect that the information collected by the credit bureaus has on current loans performance at the microeconomic level. Our research fits in this strand. Chandler and Parker (1989) use a credit scoring model to test the effect on credit ratings of applications information (age, employment information, income, dependents, other debts, banks relationships, etc.) versus credit bureau information (number of reports on debtors, debtor rating, operations, credit line use, debtor number of accounts, etc.). They find that the latter predicts borrowers' credit risk better. Barron and Staten (2002) also use credit scoring and a simulation mechanism to analyze the influence of positive and negative information on households' credit availability. They conclude that the more information, the lesser the credit constraints; hence, it is useful to include positive information. Kallberg and Udell (2002) introduce an index that summarizes borrowers' credit history in a logistic model to show that voluntary information sharing is useful in determining default.

This article is related to the latter literature strand in that it uses microeconomic information to determine risk. However, our particular design of the variables that describe past behavior allows us to measure the impact of these variables on current performance, adding variety and precision to the literature (see Barron and Staten (2002) and Chandler and Park (1989)). The use of information from a compulsory participation sharing mechanism, as opposed to the voluntary sharing of Kallberg and Udell (2002), makes the results

---

elements (positive information); hence, the incentives that borrowers receive to never miss payments deteriorate.

particularly relevant for those countries that have a public registry of credit information. Finally, the measurement of the effect of economy wide fluctuations on default at loan level is an extension to the existing literature.

3. Methodology

3.1 The database

The database comes from the Mexican PRCI (Senicreb), founded in 1964 and still run by the Central Bank today. The PRCI was originally conceived as an aid to the supervision of bank compliance with directed credit regulation, by monitoring the allocation of credit to different economic sectors, and by providing aggregate statistics on the financial system. The provision of credit reports to banks was a secondary service. Moreover, the measurement of credit risk was not contemplated originally. Senicreb collects monthly data from banks and other financial institutions for which it is compulsory to provide this information[10]. All loans over 200,000 pesos are registered individually.[11] The database includes loan identification (borrower's tax number, name, address), economic sector, State, currency denomination, loan amount, loan type, loan situation (performing or not), consolidated debt (only within the financial system) and loan rating. From its creation until 1995, Senicreb was practically the only credit information institution in Mexico[12]; however, with the financial liberalization and the consequent credit expansion of the early nineties, a private credit bureau has entered the market, displacing Senicreb as a provider of reports. Nevertheless, Senicreb has continued to gather information for regulatory purposes.

The original database includes all the loans provided by Mexican banks from January 1997 to November 2002 (71 months). However, for this analysis we restricted the data to loans provided by commercial banks to the private sector; that is, we exclude loans provided by development (public) banks and loans received by the different levels of government (from any source). It is important to notice that our analysis is performed at the loan level rather

[10] Other non-bank lending institutions are not part of Senicreb. To assess the importance of Senicreb's database in terms of the fraction of all credit information that it contains, it is enough to notice that for 1997, banks provided around 2/3 of all loans to the private sector. Even though that proportion has fallen in recent years, banks still provided over 40% of total loans to the private sector in 2003.

[11] This threshold leaves out consumption loans, such as credit cards. Loans below that limit are reported in aggregated form, but were excluded from the sample. It is interesting that some banks report loans smaller than 200,000 pesos individually; if they were reported individually they were included in the sample.

[12] Previous to 1995, there were several private attempts to share credit information, but they failed (See Negrin, 2001).

than at the borrower level because we are interested in individual loan performance. Each borrower, whatever its history, may have several loans and may not perform equally in all of them. Our database includes both loans to firms and to individuals or households. As mention below, the only loans to households that we consider are mortgages. However, mortgages represent around half of the loans included in the final sample.[13]

We determine the PD over a one-year horizon; that is, we follow the performance of each new loan for 12 months in order to determine whether or not it was defaulted on.[14] Hence, it is necessary to identify first the new loans provided at any specific point in time, and then track their payment history through the year in order to register if any of them defaulted. A new loan is one for which no information was reported in the 3 previous months. If there were loans that at the beginning of the period of study were already reporting, they were not considered new and were not included in the sample.

Another filter to identify new loans was provided by the "loan rating" information. The database includes a rating for a number of the loans because the regulation requires banks to rate a certain proportion of their portfolios. The rating system is set by the National Banking Commission (CNBV).[15] Rates range from A to E (five levels), where A is the highest rating and E corresponds to non-performing loans; the only criteria for rate deterioration is the increase in the number of missing payments. Given that we are only including new loans, it would be incongruous to keep loans that already register missed payments. Hence, we eliminated from the database all the loans that had a rating different than A[16]. We also kept in the data base the loans that were not rated.

Once the loan was identified as new, we needed to follow its performance for the next 12 months in order to see if it had defaulted or not. Although our database does not include a default definition, it identifies the months in which a borrower missed a payment. We adopted the following default criterion: a loan was consider defaulted if it had missed 3

[13] It is feasible that loans to firms and to households have a different set of determinants; hence, it is likely that historical variables affect them differently. We test this hypothesis in one of our exercises.
[14] The Basle Committee on Banking Supervision (1999) indicates that this is a common bank practice. This method is also followed in the new capital proposal (see Basel Committee on Banking Supervision, 2001b).
[15] Banks could use internal ratings if they comply with certain regulations. See CNBV circulars 1480, 1460 and 1499.
[16] Only around 40% of the loans identified as new were rated.

consecutive payments (90 days) within a year.[17] This criterion was adopted by the Basle Committee on Banking Supervision (2001a) for the application of the internal rating approach. Also, Barron and Staten (2002) use the same criterion although their estimation of PD is over a two year horizon. This system of measuring performance means that the last batch of new loans in our sample was identified in November of 2001.

Similarly, to determine borrowers' past credit behavior we need to build their history. Hence, we need to follow the history of all borrowers, looking back from the moment when they are identified as receiving a new loan. We consider at least one year of history for each borrower holding a new loan. Consequently, the first batch of new loans that we identified was provided in January of 1998. We then constructed historic variables by following the borrower's past performance in previous loans. The variables included are: borrowers' past defaults, borrowers' past payment delays, borrowers' past restructures and borrowers' past renewals[18]. From our database, we also design a variable that indicates borrower's length of participation in the credit market and a variable that shows borrower's consolidated debt[19].

At this point it is useful to summarize the filters we used to form our database. Firstly, we only included loans provided by commercial banks to private firms or individuals. Secondly, we only included loans that we identified as *new*; that is, al loans that were already running at our starting point (January 1997) were excluded. Additionally, all loans that passed our criteria of new, but had a rating different than A were excluded. Thirdly, the loans included needed to have at least 12 months of life; loans with shorter maturities were excluded. Fourthly, the loans that in their 12 months of life, interrupted their reports for more than two months, were excluded from the database.

Table 3.1 presents a list of the variables included in the database, their characteristics and the expected effect on default probability. The list of economic activities is reported in Appendix 2. Bank names are excluded for secrecy reasons. The loan currency

[17] It is important to mention that some loans in our original database presented a continuity problem. That is, some banks failed to report the performance of a particular loans during a certain month of set of months, and resume reporting them at a later period. The loans that exhibited this lack of continuity for two or more periods within a year were excluded from the sample.

[18] It is interesting to notice that by design, some loans may have a longer history than others, depending on the moment when they were identified as new. We do not consider that his introduces any bias to our estimation because it affects the same performing and non-performing loans.

[19] Each time a loan is identified as new, we check if the corresponding borrower had previous market participations. If the borrower did, then we count the time from the first participation to the time of the current loan.

denomination could be pesos, dollars or Units of Investment (UDIS) pegged to the CPI. Although all historic variables are straight forward, some other variables included in the table need further explanation.

Table 3.1

Variable	Type	Description	Expected effect
Dependent Variable: Default.	Dummy	D=1 means default, 0 otherwise	
Debtor location	Dummy	1 per each of 32 Estates	Uncertain
Debtor economic activity	Dummy	1 dummy per each of 22 economic activities (see Apendix2)	Uncertain
Bank that provides loan	Dummy	1 dummy per each of 36 banks	Uncertain
Currency denomination	Dummy	3 dummies: national, foreign, and UDIS (indexed to inflation)	Uncertain
Collateral	Dummy	D=1 if there is collateral, D=0 otherwise	(-)
Initial loan amount	Continuous	From 0 to 25,000,000 pesos	Uncertain
Interest rate (loan specific)	Continuous	Ratio of interest paid to loan amount. Reported for 54% of the loans	(+)
Consolidated debt	Continuous	Aggregated debt of the same borrower with the whole banking system at the beginning of the current loan	Uncertain
First participation in financial market	Discrete	Number of months since the borrower had a loan for the first time to the starting date of the current loan	(+)
Previous defaults	Dummy	D=1if there were previous defaults, D=0 otherwise	(-)
Previous missing payments	Dummy	D=1if there were previous missing payments, D=0 otherwise	(-)
Previous restructured loans	Dummy	D=1if there were previous restructures, D=0 otherwise	(-)
Previous renewed loans	Dummy	D=1if there were previous renewals, D=0 otherwise	(+)
Inflation	Continuous	Measured by monthly change in CPI	(+)
Aggregated growth	Continuous	Economy monthly growth rate measured by industrial activity index	(-)
Economy wide interest rate	Continuous	Measured by monthly change TIIE 28	(+)
Month of loan origination	Discrete	From 1 to 47	Uncertain

The variable *collateral* indicates if a loan had guarantees. Although there is no direct information on collateral in our database, there is an indicator on the *type* of loan. This allows us to identify the existence or lack there-off of guarantees. Of the 13 loan types[20] contemplated in the original database, the following were considered to have collateral: loans identified directly as collateralized, loans with a deposit certificate, loans with industrial guarantees, loans for the acquisition of raw materials, loans to buy machinery or directed to factory repair and maintenance, mortgages, housing loans, and other loans with real estate guarantees. Those considered as non-guaranteed loans are: discount loans (the bank gets a written payment commitment), direct loans, simple loans in current accounts and consumer loans.[21] A limitation of this variable is that it does not allow for fractional guarantees: it takes the value of 1 if there is collateral and 0 otherwise. It is important to

[20] Observations where the type of loan was missing were dropped from the sample.
[21] Due to the high loan threshold, consumer loans are excluded.

notice that in the case of Mexico, loan guarantees may be a secondary consideration for loan providers, due to the cumbersome and inefficient process of recovery of collateral associated with non performing loans[22].

The variable *interest rate* needs explanation for conceptual and practical reasons. Conceptually, the interest rate charged on a loan could summarize all information a bank has about a borrower. If after analyzing this information, the bank considers that the likelihood of repayment is low, it will assign the loan a high interest rate. Given that most of the information that banks have about a borrower is balance sheet information (both, in the case of firms and people), it is likely that interest rates would summarize it[23]. Hence, we used it as proxy variable for balance sheet information. This assumes that the interest rate variable does not include information on borrower's history.

Interest rates are not reported directly in the original database. The data only registers the amounts paid as interest on a monthly basis. We obtained interest rate per loan by averaging the monthly ratio of interest amount paid to current debt, over the life of the loan. It is necessary to mention that only 54% of the loans reported interest rates payments. In this case we left the unreported rates as zero because there are many arrangements to pay back a loan, like paying the whole debt –principal plus interests- at maturity or discounting the interests from the loan at the beginning[24].

The variable *initial loan amount* allows us to separate firms according to their size, provided that loan amount is directly proportional to firm size. Firm size could also be used as a proxy variable for some balance sheet information, like sales. We generate two dummy variables to separate large from small firms, based on 500,000 and 750,000 pesos thresholds of initial loan amount, respectively. It is important to mention that we set a maximum initial loan amount of 25 million pesos to get rid of outliers[25].

---

[22] The head of the Bankers Association declared in 2000 that, with the bankruptcy laws in place at the time, it would take from 8 to 10 years to recover collateral in Mexico (Héctor Rangel Domene, declared to the daily "La crónica de hoy", 27/IV/2002). Although such law was revamped in 2001 recovery problems persist.

[23] Since the Altman (1968) paper on bankruptcy prediction, there have been many papers on the relevance of financial variables in default determination. Recent references on the relevance of financial and non-financial variables in internal credit ratings are Grunert, et al (2002) and Hayden (2003).

[24] We also tried assigning the average interest rate to the missing rate observations but it did not alter our estimation significantly.

[25] This meant the exclusion of 2,614 loans. Among them there were 691 loans of over 100 million pesos.

Finally, we included inflation and growth variables in our estimations to control for the macroeconomic environment (in the case of these variables, the inflation or growth rate is the same for all loans current at each moment in time). Inflation was calculated as the monthly change in the consumer price index (CPI) while growth was obtained as the monthly change in the industrial production index. We also used an economy wide interest rate given that many loans did not report a particular interest rate[26]. As with inflation, we took the monthly changes in interest rate as a control variable.

3.2 The Probit model

We use a Probit model to estimate the PD. In these models, the dependent variable is dichotomous, taking the value of "1" when the loan defaulted and "zero" when it remained performing.[27] The PD is obtained by applying the following formula:

$$\hat{P}_i = \int_0^{x_i'\hat{\beta}} \frac{1}{\sqrt{2\Pi}} e^{-\frac{1}{2}(x_i'\hat{\beta})} dx$$

where $\hat{P}_i$ is the estimated PD for loan i, x_i is the vector of explanatory variables for loan i and $\hat{\beta}$ is the vector of estimated coefficients. Omitting the individual loan index for notational simplicity, we can write the so called Probit Index ($x'\hat{\beta}$) as follows:

$$x'\hat{\beta} = \sum_{i=1}^{36} \hat{\alpha}_i D_i^B + \sum_{j=1}^{22} \hat{\mu}_j D_j^E + \sum_{k=1}^{3} \hat{\gamma}_k D_k^C + \sum_{l=1}^{32} \hat{\eta}_l D_l^G + H^{S}{}'\hat{\lambda} + Z'\hat{\delta} \ \ldots (1)$$

where

D_i^B are dummy variables for the i banks;

D_j^E are dummy variables for the j economic sectors;

D_k^C are dummy variables for the k currencies of loan denomination;

---

[26] Notice that this rate is different than the loan specific rate derived from within our database. In this case, the same interest rate was assigned to all the loans provided at the same time. We used the 28 days interbank equilibrium interest rate (TIIE) as the basis of our calculations. This rate is determined by Banco de México based on information provided by credit institutions.

[27] For a full explanation of the Probit model see among many others Maddala, (1983).

D_l^G are dummy variables for the l geographical regions;

H^S is the matrix of historical variables (past defaults, past missing payments, past renewals, past restructures, time of participation in the banking system and consolidated debt);

Z is the set of all other positive variables (initial rating, initial debt amount, collateral, and interest rate). This could be considered positive information (Barron and Staten, 2002).

It is important to point out that the coefficients obtained from the Probit model are not susceptible to direct interpretation and must be transformed in order to obtain the usual marginal effects.[28] The coefficients that we report in the tables of results are already transformed and can be interpreted as marginal effects. Given that the models are non-linear, the coefficients should be interpreted as the change in the probability of default due to a change in the corresponding variable, assuming that all other independent variables are fixed at their mean values. It is also important to mention that all models were corrected for heteroskedasticity.[29] Finally, we reported the goodness of fit estimators appropriate for the dependent dichotomous variable, in the presence of heteroskedasticity.[30]

4. Data description

The sample database contains 90,805 new loans of which 3,588 defaulted on (4.0%). In broad terms, the average initial amount per loan was 1,147,907 pesos[31]; however, if we partition the sample into deciles, the average initial loan amount for the first 9 deciles (84, 143 loans) was 524,812 pesos, while the average loan amount of the last decile (6,662 leans) was 9,017,766 pesos.[32] Although the average consolidated debt per borrower across the banking system was 5,806,207 pesos, the average consolidated debt of the first nine deciles was only 3,318,142 pesos while that of the last decile was over 37,231,200 pesos.

---

[28] See Green, 2000.

[29] Appendix 3 presents the estimated Probit model corrected for heteroskedasticity. It presents also the mechanism to obtain the marginal effects in the presence of heteroskedasticity.

[30] Appendix 4 presents a brief explanation of the goodness of fit measures that we report in the tables.

[31] The exchange rate at the time of the last month included in the sample (November of 2002) was 10.12 pesos per US dollar. For the data that appears in pesos in the rest of the section it is useful to just divide it over 10 to obtain a close approximation of its amount in dollar terms.

[32] Recall that we excluded loans over 25 million pesos.

Of the total loans, almost 60% had some form of collateral. In 6.5% of the loans, the holder had a history of non performing loans in the system, while in 3.6% of the credits the holder had missed payments in previous loans. Only 1.1% and 1.6% had previous restructures and renewals, respectively. With respect to non-performing loans, in 84% of them the holder had had non-performing loans in the past, and in 36.2% the holder had missed payments in previous loans. Half of non-performing loans had collateral (See Table A.1 in Appendix 1).

The average interest rate for the loans that reported it was 2.41% (a 28.92% yearly rate)[33]. The average interest rate for the loans that defaulted was 4.27% (a yearly rate of 51.2%). This may reflect the fact that lenders may have known the credit history of their clients since they seemed to have assigned a higher rate to lenders with bad credit history[34].

With respect to previous participation in the credit market, it is likely that current borrowers that participated before would have a lower PD than new borrowers since they have already enjoyed credit benefits and would not want to lose them.[35] In our sample more than 76.9% of the loans were received by new market participants, while 23.1% of the loans were provided to borrowers with previous market experience. Of the loans that defaulted, 19.7% had had previous market experience (See Appendix 1 Table A.2).

In terms of currency denomination, 66.7% of the total loans were denominated in pesos, 24.6% in "investment units" (UDIS) and 8.7% in dollars. Of the loans that defaulted almost 91.7% was denominated in pesos, while 8.1% were provided in dollars (See Appendix 1 Table A.3).

The database indicates high loan concentration in terms of source, geographical region and economic activity. Only 6 banks provided almost 86% of the loans provided, while 65% of the total number of loans was directed to only 5 States (See Appendix 1, Tables A.4 and A.6). Regarding the borrowers' economic activity, 75.2% of all loans were provided to 5

---

[33] Only 53% of the total loans reported interest rate payments.

[34] In 1998, financial authorities issued a regulation that forced banks to obtain the corresponding report on applicant's past behavior, before providing a loan. The regulation indicates that the bank that does not consult such report, has to provision that credit at the highest level. It is interesting that even after this regulation was issued, the data shows that banks still provided loans to people that had a bad history. Hence, banks must have provisioned these loans much higher than those that had no bad history. It is also interesting that the interest rate reported on loans received by borrowers with bad history was considerably higher than the average rate. See CNBV circulars 1413 (September 30, 1998), 1476 (August 16, 2000) and 1503 (August 14, 2001) for the original regulation and further modifications.

[35] Similarly, borrowers with longer credit histories would have a lower default rate.

activities; in particular, 51.8% of the loans in the sample were mortgages. It is interesting to notice that housing has the lowest default rate (3.1%) whereas agriculture has the highest (8.3%). (See Appendix 1 Table A.5)

To conclude this section, Table 4.1 contains the basic statistics of the variables included in the estimations. These data are relevant because the Probit model is non-linear; hence, the probabilities of default are calculated at the corresponding variable mean value.

Table 4.1

Variable	Mean	Standard Dev.	Max	Min
Dependent Variable: Actual Default	0.0395	0.1948	1	0
Original rating	2.1822	1.3359	3	0
Guarantees and collateral	0.6045	0.4890	1	0
Initial loan amount *	1147.91	2738.79	25000	1
Interest rate**	0.0127	0.0437	1	0
Consolidated debt with banking system*	5806.2	361018	41600000	1
Participation in the credit system	3.3378	8.7250	66	0
Previous missing payments with system	0.0650	0.2465	1	0
Previous restructured loans in the system	0.0108	0.1033	1	0
Previous renewed loans with system	0.0156	0.1240	1	0
Previous missing payments with system	0.0443	0.2483	5	0
Monthly inflation rate	0.0080	0.0063	0.0253	0
Monthly change in interest rate (TIIE 28)	-0.0029	0.1197	0.6307	0
Monthly output growth rate	-0.0022	0.0412	0.1054	0
Month of loan origination	21.549	13.6375	47	4

* Thousands of current pesos.
** Average rate for all loans, including the loans that did not report a rate.

5. Estimations and Results

In this section we present the estimation results. We show first the importance of historical variables on PD determination in a reference model. Then, we study the relevance of historic variables under different scenarios and, finally, we test several secondary hypothesis. Among them, we test the effect of technological change on PD determination, the relevance of the regulation on information usage, the effect of inflation and aggregate growth and whether historical variables affect commercial and personal (mortgages) loans differently. The coefficients reported in all the tables in this section, are the marginal effects of each variable. These coefficients should be interpreted as the change in the probability of default due to a change in the corresponding variable, assuming that all other independent variables remain at their mean values[36].

5.1 Reference Model

The first column of Table 5.1 presents our *reference model* which focuses on the effect of borrowers' past performance in the credit market. The variables included are previous defaults, previous missed payments, previous restructures and renewals, length of participation length in the market, consolidated debt, current loan amount, collateral and *control variables* for bank (that provided the loan), State of borrower's residence, borrowers' economic activity and loans' currency of denomination. For presentation convenience, we only report the marginal effects of some control variables[37].

We see in Model 1 that all historical variables have the expected sign and most of them are significant. Having defaulted on a previous loans is highly significant and increases PD by 30 percentage points (pp). This is by far the highest effect on PD. Having missed payments in the past and having restructured a loan (both significant at 1%) increase PD in 1 and 0.01 pp, respectively. Length of participation in the market (significant at 1%) reduces PD in 0.01 pp while loan renewals are not significant.

Although the expected sign was uncertain, both initial loan amount and consolidated debt significantly reduce the PD. Starting from the mean loan value of 1,147,910 pesos, an increase of 100,000 pesos of loan amount diminishes PD in 1.8 pp. Similarly, from the mean value of consolidated debt of 5,806,200 pesos, an increase of 100,000 pesos reduces PD in 0.07 pp. It is interesting to notice that loans denominated in UDIS and in dollars (both significant at 1% level) meant a reduction of PD in 0.56 and 0.06 pp, respectively. Although the variable collateral is hardly significant (at 10%), it reduces PD in 0.02 pp. Finally, the average PD estimated with this model is 3.97%.

[36] The models were corrected for the presence of heteroskedasticity. See Appendix 3 for an explanation.
[37] The coefficients that we do not report are bank, state and economic activity. Complete estimations are available upon request.

Table 5.1

VARIABLES	MODEL 1		MODEL 2		MODEL 3		MODEL 4		MODEL 5	
Previous Default	0.30621	***	0.31063	***	0.30225	***	0.30983	***	0.30587	***
	(0.04195)		(0.0422)		(0.04234)		(0.04166)		(0.04205)	
Previously Missed Payments	0.01037	***	0.01029	***	0.00987	***	0.01054	***	0.01003	***
	(0.02889)		(0.0289)		(0.02882)		(0.02869)		(0.02860)	
Restructures	0.00098	***	0.00091	***	0.00050	*	0.00102	***	0.00053	*
	(0.07870)		(0.07892)		(0.07835)		(0.07827)		(0.07791)	
Renewals	-0.00023		-0.00029		-0.00023		-0.00020		-0.00021	
	(0.10375)		(0.10376)		(0.10373)		(0.10319)		(0.10315)	
Length of Participation	-0.00010	***	-0.00010	***	-0.00009	***	-0.00010	***	-0.00009	***
	(0.00184)		(0.00185)		(0.00184)		(0.00183)		(0.00183)	
Total Debt	-7.25E-09	***	-7.11E-09	***	-6.80E-09	***	-7.32E-09	***	-6.86E-09	***
	(0.00000)		(0.0000)		(0.00000)		(0.00000)		(0.00000)	
Initial Loan Amount	-1.80E-07	***	-1.79E-07	***	-1.66E-07	***	-1.14E-07	***	-1.04E-07	***
	(0.00001)		(0.0000)		(0.00001)		(0.00001)		(0.00001)	
Collateral	-0.00026	*	-0.00027	*	-0.00027	*	-0.00024		-0.00025	*
	(0.04944)		(0.0495)		(0.04944)		(0.04921)		(0.04921)	
Loan Interest Rate					0.00676	***			0.00673	***
					(0.16786)				(0.16726)	
$500,000 Threshold							-0.00049	***	-0.00046	***
							(0.03875)		(0.03867)	
Dollar Denomination	-0.00062	***	-0.00061	***	-0.00055	***	-0.00056	***	-0.00050	***
	(0.05561)		(0.0556)		(0.05563)		(0.05564)		(0.05566)	
UDIS Denomination	-0.00565	***	-0.00537	***	-0.00519	***	-0.00561	***	-0.00515	***
	(0.17911)		(0.1797)		(0.17838)		(0.17731)		(0.17661)	
Technological Trend			-0.00001	***						
			(0.0011)							

***significant at the 1%, ** significant at the 5%, * significant at the 10%

//Degrees of Freedom

GOODNESS OF FIT

Num. Obs.	90552		90552		90552		90552		90552	
Log-Lik Intercept Only:	-15099.203		-15099.203		-15099.203		-15099.203		-15099.203	
Log-Lik Full Model:	-5554.383		-5545.143		-5478.704		-5539.256		-5464.553	
Count R2:	0.981		0.981		0.981		0.980		0.981	
Adj Count R2:	0.511		0.512		0.516		0.507		0.512	
AIC:	0.125		0.124		0.123		0.124		0.123	
BIC:	-1.021e+06		-1.021e+06		-1.022e+06		-1.021e+06		-1.022e+06	

5.2. Other hypothesis

Model 2 adds a time variable in order to capture technological effects on PD determination. We would have expected that technological improvements in information gathering and processing would have tended to reduce the PD. We find that the time trend is highly significant and has a negative sign but it has a small effect on PD. The relevance of historical variables remains practically unchanged.

Balance sheet information has been always considered the most important determinant of PD. In order to analyze the relevance of historical variables when controlling for balance sheet information, in Models 3, 4 and 5 we introduced the appropriated proxy. Model 3 includes the variable that represents individual loans interest rates, which accounts for all the information that the bank had to use to set the rate. As expected, this variable increases the PD in a significant way. A marginal increase of interest rate above its mean value (of 0.0127) increases the PD in 0.6 pp.[38] In Model 4 the proxy variable is a dummy that represents large and small firms. The size of the firm would be a proxy for sales. We find that the dummy is significant and that increasing the size of the firms reduces the PD in 0.05 pp.[39] Finally, Model 5 includes both proxy variables simultaneously. Both of them remain highly significant and have the right signs. The relevance of the latter exercises is that historical variables remain highly significant and their coefficients are robust even when we control for balance sheet information. Hence, it is likely that models that are based exclusively on balance sheet information and that leave out historical behavior variables may be miss specified.

The richness of our data set allowed us to test a set of other hypothesis that are presented in Table 5.2. To the best of our knowledge, the effect of macroeconomic variables, like inflation and output growth, on PD determination at the loan level had not been measured before. Model 1 indicates that inflation increases PD in 3.69 pp while economic growth reduces PD in 0.31 pp. In Model 2 we introduced an economy wide interest rate[40] to see its effect over PD; we find that it increases PD in 0.18 pp. Inflation usually implies higher interest rates. However, their changes may not coincide in time; hence they may be capturing different effects on PD. In Model 3 we introduced them together. Both, inflation and economy wide interest rate, are significant and increase PD. This results show that the macroeconomic environment affects individual loans' PD in a significant manner; hence, macroeconomic stability and growth are necessary condition to achieve a sound financial system.

[38] As it was explained before, this variable is obtained by averaging out the monthly interest payments of each particular loan. However, only around half of the loans reported monthly payments. In this exercise, the loans that did not reported an interest rate were assigned a zero interest rate. We performed an additional exercise in which we assigned the average interest rate to the loans that had not reported it. This variable was also significant and increased the PD in 0.55 pp.

[39] The dummy variable was based on an initial loan amount threshold value of 500,000 pesos. However, we performed another exercise with another dummy based on a threshold value of 750,000 pesos. The outcome was practically the same.

[40] This variable is not the interest rate charged per loan used before, but an economy wide interest rate. We used the 28 days interbank equilibrium interest rate (TIIE 28).

Table 5.2

VARIABLES			MODEL 1		MODEL 2		MODEL 3		MODEL 4	
Previous Default			0.30637	***	0.30312	***	0.30606	***	0.30322	***
			(0.04271)		(0.04251)		(0.04270)		(0.04238)	
Previously Missed Payments			0.01004	***	0.01020	***	0.01026	***	0.00974	***
			(0.02895)		(0.02904)		(0.02908)		(0.02872)	
Restructures			0.00039		0.00042		0.00036		0.00053	*
			(0.07856)		(0.07876)		(0.07882)		(0.07818)	
Renewals			-0.00028		-0.00020		-0.00024		-0.00023	
			(0.10385)		(0.10408)		(0.10403)		(0.10262)	
Length of Participation			-0.00009	***	-0.00009	***	-0.00009	***	-0.00009	***
			(0.00185)		(0.00185)		(0.00185)		(0.00183)	
Total Debt			-6.58E-09	***	-6.89E-09	***	-6.71E-09	***	-6.66E-09	***
			(0.00000)		(0.00000)		(0.00000)		(0.00000)	
Initial Loan Amount			-1.67E-07	***	-1.72E-07	***	-1.71E-07	***	-1.23E-07	***
			(0.00001)		(0.00001)		(0.00001)		(0.00001)	
Collateral			-0.00027	*	-0.00030	*	-0.00030	**	-0.00026	*
			(0.04957)		(0.04947)		(0.04954)		(0.04909)	
Loan Interest Rate			(0.00685)	***	(0.00688)	***	(0.00690)	***	(0.00665)	***
			(0.16774)		(0.16836)		(0.16819)		(0.16743)	
Dollar Denomination			-0.00054	***	-0.00057	***	-0.00056	***	-0.00056	***
			(0.05558)		(0.05570)		(0.05568)		(0.05517)	
UDIS Denomination			-0.00481	***	-0.00517	***	-0.00488	***	-0.00516	***
			(0.18077)		(0.17927)		(0.18123)		(0.17718)	
Inflation			(0.03689)	***			(0.02870)	***		
			(2.43446)				(2.51922)			
Output Growth			-(0.00310)	***	-(0.00372)	***	-(0.00319)	***		
			(0.33311)		(0.33169)		(0.33564)			
Change in Economy Wide Interest Rate					(0.00179)	***	(0.00156)	***		
					(0.09005)		(0.09200)			
Commercial Loan Dummy									(0.00069)	
									(0.25272)	
Interaction = Initial Amount * Commercial Dummy									(0.00000)	***
									(0.00009)	

***significant at the 1%, ** significant at the 5%, * significant at the 10%

//Degrees of Freedom

GOODNESS OF FIT

Num. Obs.			90552		90552		90552		90552	
Log-Lik Intercept Only:			-15105.538		-15105.538		-15105.538		-15099.203	
Log-Lik Full Model:			-5456.134		-5446.67		-5438.885		-5460.825	
Count R2:			0.981		0.981		0.981		0.981	
Adj Count R2:			0.515		0.516		0.518		0.519	
AIC:			0.122		0.122		0.122		0.123	
BIC:			-1.022e+06		-1.022e+06		-1.022e+06		-1.022e+06	

The last hypothesis that we test, is whether PD behaves differently for commercial and for personal loans. As we have mentioned before, the only personal loans included in the database are mortgages, which represent around 51% of the sample. We separated personal and commercial loans by means of a dummy variable which takes the value of 0 if the loan is commercial and 1 if it is a mortgage. The coefficient turns out to be non significant. Nevertheless, we interacted this dummy variable with loan amount in order to determine if changes in loan amount affect mortgages differently than commercial loans. The coefficient of this variable is highly significant and negative, which means that the drop in PD as loan amount increases is higher for mortgages than for commercial loans[41].

6. Conclusions and Policy Issues

We have shown that information on borrowers' past behavior is essential in the determination of individual loans' PD. In particular, if a borrower had past defaults or past missing payments, her current loan's PD increases dramatically. Having participated in the credit market in the past, as well as loan amount and total debt within the banking system, reduce PD significantly. These results remain significant after controlling for a number of variables, like balance sheet information and technological change. Our data base allowed us to test a number of relevant effects on probability of default. Among them, the impact of macroeconomic stability and growth; we find that while inflation increases loan's PD, economic growth reduces it.

This study has several policy implications. It is clear that institutions to share information are vital for a sound credit system. It is irrelevant the institutional origin –public or private- of the sharing mechanism. Hence, our analysis suggests that greater effort should be invested to conform larger and more complete databases on individuals credit history. Finally, our analysis stresses the relevance of macroeconomic stability for sound credit practices.

[41] The lack of significance of the dummy's coefficient means that the PD distribution of commercial loans and mortgages have the same mean. This may be related to the restriction on the initial amount of loans (25 million pesos) that we imposed. However, the variance of the two distributions differ; mortgages distribution registers a significantly smaller variance than commercial loans.

References

Altman, E.I., 1968, “Financial ratios, discriminant analysis and the prediction of corporate bankruptcy”, Journal of Finance, 23, 589-605.

Barron, John M. and Michael Staten, 2002, “The Value of Comprehensive Credit Reports: Lessons from the U.S. Experience” to appear in Credit Reporting Systems and the International Economy, edited by Margaret J. Miller, MIT Press, forthcoming.

Basle Committee on Banking Supervision, 1999a, Credit Risk Modeling: Current Practices and Applications, BIS.

Basle Committee on Banking Supervision, 2001a, The Internal Rating Approach, BIS.

Basle Committee on Banking Supervision, 2001b, The New Basle Capital Accord, BIS.

Chandler, Gary G. and Lee E. Parker, 1989, “Predictive Value of Credit Bureau Reports”, Journal of Retail Banking, Vol. XI, No 4, pp 47-54.

Galindo, Arturo and Margaret Miller, 2002, “Can Credit registries Reduce Credit Constraints? Empirical Evidence on the Role of Credit Registries in Firm Investment decisions”, presented in the Seminar ‘Towards Competitiveness: the Institutional Path’, Annual meetings of the Board of Governors, Inter-American Development Bank and Inter-American Investment Corporation, Santiago de Chile, March.

Green, William H., 1990, Econometric Analysis. Macmillan Publishing Co.

Gruner, Jens, Lars Norden and Martin Weber, 2002, The role of non-financial factors on internal credit rating. Center for Economic Policy Research, Discussion Paper No. 3415

Hagle, Timothy M. and Glen E. Mitchell, II. 1992. “Goodness-of–Fit Measures for Probit and Logit” American Journal of Political Science, 36(3), 762-84.

Hayden, Evelyn (2003), “Are Credit Scoring Models Sensitive With Respect to Default definitions? Evidence from the Austrian Market, University of Vienna, Department of Business Administration.

Jappelli, Tulio and Marco Pagano, 1999, “Information Sharing, Lending and Defaults: Cross-Country Evidence”, mimeo, Centre for Studies in Economics and Finance, Working Paper No. 22.

Kallberg, Jarl G. and Gregory F. Udell, 2003, The value of private sector business credit information sharing: The US case. Journal of Banking and Finance 27, 449-469.

Klein, Daniel B., 1992, “Promise Keeping in the Great Society: A Model of Credit Information Sharing”, Economics and Politics, vol. 4, no. 2, 117-136.

Long, Scott J. and Jeremy Freese, 2001, Regression Models for Categorical Dependent Variables Using Stata. Stata Press, USA.

Madala, G.S., 1983, Limited Dependent and Qualitative Variables in Econometrics, CUP.

Márquez, Javier, José L. Negrin, Pascual R. O’Dogherty and Alejandro M. Werner Credit (2003), Information, Credit Risk Measurement and the Regulation of Bank Capital and Provision Requirements in Mexico. Banco de México. Unpublished document.

Miller, Margaret, 2000, “Credit Reporting Systems Around the Globe: The State of the Art in Public and Private Credit Registries”, World Bank. Presented at the Second Consumer Credit Reporting World Conference, held in San Francisco, California, October.

Negrin José L., 2001, “Mecanismos para Compartir Información Crediticia. Evidencia Internacional y la Experiencia Mexicana”, El Trimestre Económico, V68(3), 405-466.

Padilla, A. Jorge and Marco Pagano, 1999, “Sharing Default Information as a Borrower Discipline Device”, Centre for Studies in Economics and Finance Working Paper no. 21.

Padilla, A. Jorge and Marco Pagano, 1997, “Endogenous Communication Among Lenders and Entrepreneurial Incentives”, Review of Financial Studies, V10, no.1, 205-236.

Pagano, Marco and Tullio Japelli, 1993, “Information Sharing in Credit Markets”, Journal of Finance, vol. 48, no. 5, 1693-1718.

Raftery, A.E. 1996, Bayesian model selection in social research. In Sociological Methodology, ed. P.V. Marsden, 111-163.Oxford:Basil Blackwell.

Rothschild M. and Stiglitz S. 1976, “Equilibrium in Competitive Insurance Markets: an Essay on the Economics of Imperfect Information”. Quarterly J. of Economics 80, 629-49.

Stiglitz, Joseph E. and Weiss, Andrew 1981, “Credit Rationing in Markets with Imperfect Information”, American Economic Review, vol. 71, no. 3, 393-410.

Appendix 1

Descriptive Statistics

Table A.1

	(1) Loans (#)	(2) % of Total loans	(3) Non-performing Loans (#)	(4)= (3)/(1) Non-Perf/Total (%)
Total Number of Loans	90,805	100.0	3,588	4.0
Collateralized loans	54,892	60.5	1,877	3.4
Borrowers with:				
Previous Non-Performing Loans	5,901	6.5	3,013	51.1
Previous Missing Payments	3,309	3.6	1,198	36.2
Previous Restructures	980	1.1	156	15.9
Previous Renewals	1,418	1.6	61	4.3

Table A.2

	Total Number of Loans		Defaulted Loans	
Market Participation	(1) Number of loans	(2) (%) of Total loans	(3) Non-Performing	(4)=(3)/(1) (%) of Loans
Total Number of Loans	90,805	100.0	3,588	100.0
First market participation	69,863	76.9	2,882	80.3
Previous market participation	20,942	23.1	706	19.7

Table A.3

	Total loans		Non-performing loans		Non-per./ total loans (%)
Loan Denomination	#	Structure (%)	#	Structure (%)	
Pesos	60,610	66.7	3,289	91.7	5.4
Dollars	7,857	8.7	292	8.1	3.7
UDIS	22,338	24.6	7	0.2	0.03
Total	90,805	100.0	3,558	100.0	4.0

Table A.4

	Loans		Non-performing Loans	
Bank	**#**	**(%) of Total Loans**	**#**	**(%) of Bank Loans**
Bank 1	29,124	32.1	1,025	3.5
Bank 2	22,022	24.3	806	3.7
Bank 3	9,857	10.9	660	6.7
Bank 4	9,117	10.0	119	1.3
Bank 5	4,363	4.8	196	4.5
Bank 6	4,057	4.5	190	4.7
Other 30 banks	12,265	14	592	4.8
Total	90,805	100.0	3,588	4.0

Table A.5

	Loans		Non-performing Loans	
Economic Activity	**#**	**(%) of Total Loans**	**#**	**(%) of Activity Loans**
Housing (21)	47,053	51.8	1,447	3.1
Professional Services (18)	7,028	7.7	321	4.6
Finished products trade (3)	5,804	6.4	285	4.9
Raw materials and machinery trade (4)	4,852	5.3	209	4.3
Agriculture (1)	3,665	4.0	305	8.3
Other 17 activities	22,403	25	1,021	4.6
Total	90,805	100.0	3,588	4.0

Table A.6

	Loans		Non-performing Loans	
State	**#**	**(%) of Total Loans**	**#**	**(%) of State Loans**
Federal District (9)	32,351	36.0	606	1.9
Nuevo Leon (19)	10,497	12	536	5.1
Estado de Mexico (15)	5,235	6	129	2.5
Jalisco (14)	6,254	7	369	5.9
Guanajuato (11)	3,370	4	116	3.4
Other 27 states	33,098	36	1,832	5.5
Total	90,805	100	3,588	4.0

Appendix 2

Economic Activities

1	Agriculture
2	Livestock, fishing and hunting
3	Extraction of coal, graphite, oil, natural gas, minerals and salt.
4	Food, beverages and tobacco
5	Textiles and apparel includes leather industries
6	Cork, wood and paper products
7	Chemical industry
8	Fabrication of rubber, plastic and non-metallic minerals
9	Basic metallic industries
10	Manufacturing, assembling and repairs of machinery and equipment
11	Other manufacturing industries
12	Construction
13	Clothing and home products retail trade
14	Gas, fuels, lubricants, raw materials, machinery and transportation equipment trade
15	Real estate retail trade
16	Communications and transport
17	Financial services, insurance and leasing
18	Professional and technical services
19	Commerce, restaurants and hotels
20	Educational, health and public services, social assistance and others.
21	Housing

Appendix 3

In Probit models, the presence of heteroskedasticity, represents a more serious problem than in OLS since the estimated parameters become inconsistent. We use Harvey's general formulation to correct this problem(Greene, 2000). The estimated model is the following

$$\hat{P}_i = \int_0^{\frac{x_i'\hat{\beta}}{\exp(\hat{\gamma}'z)}} \frac{1}{\sqrt{2\Pi}} e^{-\frac{1}{2}(\frac{x_i'\hat{\beta}}{\exp(\hat{\gamma}'z)})} dx$$

where z_i is the vector of explanatory variables that belong to the conditional variance specification and $\hat{\gamma}$ is the vector of estimated coefficients; the general formulation for the conditional variance is the following:

$$V[\boldsymbol{\varepsilon}] = \left[\exp(\hat{\boldsymbol{\gamma}}'\mathbf{z}_i)\right]^2$$

$$\hat{\gamma}'z_i = \hat{\gamma}_1 D_i + \hat{\gamma}_1 X_i + \hat{\gamma}_2 (D_i X_i)$$

Notice that $\mathbf{z}_i$ includes the variables that may be causing the heteroskedasticity. In our case, we considered that both, initial loan amount (X_i) and the type of loan could be generating the problem. To capture for the latter, we included the dummy variable that separates commercial loans from mortgages (D_i). We also introduced in the conditional variance estimation the interaction between these variables $(D_i X_i)$.

Given that our model corrects for heteroskedasticity, we use the following set of expressions to compute the marginal effects (2000, p.830), assuming that we are dealing with a continuous independent variable w_k, which could appear in $\mathbf{x}_i$, $\mathbf{z}_i$ or both:

$$\frac{\partial \text{Prob}(Y=1)}{\partial w_k} = f\left(\frac{\boldsymbol{\beta}'\mathbf{x}}{\exp(\boldsymbol{\gamma}'\mathbf{z})}\right)\frac{\beta_k - (\boldsymbol{\beta}'\mathbf{x})\gamma_k}{\exp(\boldsymbol{\gamma}'\mathbf{z})} \qquad \text{where } f(\bullet) \text{ is the normal pdf.}$$

The appropriate marginal effect for a binary independent variable, *d*, would be computed as follows:

$$\frac{\Delta \text{Prob}(Y=1)}{\Delta d} = \text{Prob}\left[Y=1 \middle| \bar{\mathbf{x}}^*, \bar{\mathbf{z}}^*, d=1\right] - \text{Prob}\left[Y=1 \middle| \bar{\mathbf{x}}^*, \bar{\mathbf{z}}^*, d=0\right]$$

Appendix 4

Goodness of Fit Measures (taken from Long and Freese, 2001)

For binary dependent variable models the pseudo-R2 goodness of fit measures are usually reported. However, these measures are not relevant in the presence of heteroskedasticity. The appropriated goodness of fit measures for this setting are constructed by testing the model predicted probabilities against the actual dichotomous dependent variable. Such measures are count and adjusted count R^2. We also report the Akaike Information Criterion (AIC) used in model selection for non-nested alternatives; this criterion states that the smaller AIC values the better the model. Finally, we report the Bayesian information criterion (BIC) which has been proposed as a measure of overall fit; the more negative the BIC, the better the fit. In the following paragraphs we describe these measures.

Count and adjusted count R^2 Observed and predicted values can be used to compute the count R^2. Consider the binary case where the observed y is 0 or 1 and $\pi_i = \hat{\Pr}(y = 1/x_i)$. Define the expected outcome as

$$\hat{y}_i = \begin{cases} 0 \; if \pi_i \leq 0.5 \\ 1 \; if \pi_i > 0.5 \end{cases}$$

Another measure is the proportion of correct predictions, referred to as the *count* R^2:

$$R^2_{Count} = \frac{1}{N} \sum_j nji$$

where the n_{jj}'s are the number of correct predictions for outcome j. The count R^2 can give the faulty impression that the model is predicting very well. In a binary model *without* knowledge about the independent variables, it is possible to correctly predict at least 50 percent of the cases by choosing the outcome category with the largest percentage of observed cases. To adjust for the largest row marginal,

$$R^2_{AdjCount} = \frac{\sum_j njj - \max_r(nr+)}{N - \max_r(nr+)}$$

where n_{r+} is the marginal for row r. The *adjusted count* R^2 is the proportion of correct guesses beyond the number that would be correctly guessed by choosing the largest marginal.

Information measures This class of measures can be used to compare models across different samples or to compare non-nested models. The Akaike information criterion is defined as

$$\text{AIC} = \frac{-2\ln \hat{L}(M_k) + 2P}{N}$$

where $\hat{L}(M_k)$ is the likelihood of the model and P is the number of parameters in the model (e.g., $K + 1$ in the binary regression model where K is the number of regressors). All else being equal, the model with the smaller AIC is considered the better fitting model.

The BIC was proposed by Raftery (1996) as a measure of overall fit and a means to compare nested and non-nested models. Consider the model M_k with deviance $D(M_k)$. BIC is defined as

$$BIC_k = D(M_k) - df_k \ln N$$

where df_k is the degrees of freedom associated with the deviance. The difference in the BIC from the two models indicates which model is more likely to have generated the observed data. If $BIC_1 - BIC_2 < 0$, then the first model is preferred. If $BIC_1 - BIC_2 > 0$, then the second model is preferred.