Mortgage Lending and Consumer Behavior in Mexico

David Camposeco, Adrian G. de la Garza*, and Carolina Rodriguez-Zamora
Banco de Mexico

Abstract
This paper analyzes how credit policies imposed by mortgage-granting institutions in Mexico affect consumer behavior. In particular, we use a regression discontinuity framework to study how specific income thresholds that determine to which product mortgagors get access may impact default in the mortgage market. In addition, we investigate which factors influence the probability that a loan become nonperforming. To this end, we estimate a semiparametric hazard model to study how various loan characteristics and macroeconomic conditions, which vary throughout the life of the loan, may affect the risk of nonperformance. Our preliminary results suggest that, at least for certain cofinanced credit products, limiting the supply of credit based on arbitrary income thresholds has no effect on delinquency rates or default patterns, which may suggest that current institutional credit policies could be relaxed without negatively affecting the rate of nonperforming loans in bank portfolios. In agreement with the literature has researched the determinants of default in other countries and for other mortgage products, we also find that loan characteristics, such as the loan-to-value ratio at origination, may be crucial to explain delinquency hazards.

Keywords: Mortgage lending, credit default.

1. Introduction

It is difficult to overestimate the importance of mortgage lending in today’s modern economies. From the point of view of consumers, purchasing a home represents the opportunity to invest in a tangible asset, accumulate wealth and build a safety net for the homeowner and his family. In several industrialized countries, housing wealth represents over one half of total household net worth and mortgage debt constitutes the largest liability of households (Drudi et al., 2009; Iacoviello, 2011). Although data for developing countries are relatively scant, it is considered that residential property comprises an even larger percentage of total household assets or net worth, particularly since households in developing economies do not typically invest in liquid assets.
such as bonds or equity. In addition, although household debt levels are relatively low in middle-income countries such as those in Latin America, mortgage lending has recently expanded and is expected to deepen even further given the development of financial markets in the region and the increasing purchasing power of a young population (Softec, 2009; Cubeddu et al., 2012). From the perspective of lenders, mortgage credit allows them to expand their loan portfolios relatively safely given the existence of physical collateral that lending institutions can seize in case of default.

However, mortgage lending is clearly not a risk-free venture. The housing market is prone to large swings in prices, which implies that a decrease in these prices may lead to situations in which the credit balance of the mortgagor exceeds the value of the property. In this case, consumers may decide to default on their loans, thereby increasing the share of nonperforming loans in the lender’s balance sheet. As we know, the subject has received considerable attention in recent years, as the burst of an asset price bubble in the U.S. housing market is widely credited as one of the main drivers of the global financial crisis of 2008-09. Moreover, recent research by Eggertsson and Krugman (2012) and Guerrieri and Lorenzoni (2011) has uncovered important effects of large household debt accumulation on real economic activity, which further strengthens the case for understanding mortgage markets and the corresponding factors that may trigger default.

This paper has two main objectives. First, it studies how credit policies imposed by mortgage-granting institutions in Mexico may affect consumer behavior. In particular, we use a regression discontinuity (RD) framework to analyze how specific income thresholds that determine to which product mortgagors get access may impact default in the mortgage market. Second, we investigate the factors that influence the probability that a loan become nonperforming. To this end, we estimate a semiparametric hazard model to study how various loan characteristics and macroeconomic conditions, which vary throughout the life of the loan, may impact the risk of nonperformance. Although a vast literature has investigated the determinants of default in mortgage credit markets, there is evidence that the relevant factors may vary depending on the various products offered (von Furstenberg, 1969; Furstenberg and Green, 1974; Campbell and Dietrich, 1983). Moreover, we employ a proprietary data set that contains information on the universe of mortgage products offered by commercial banks in Mexico, some of which are offered jointly with Infonavit— a public fund and the main mortgage lending institution in the country. While most of the relevant literature has typically focused on mortgage lending in industrialized countries, particularly in the U.S., our study is the first to comprehensively investigate default in the Mexican mortgage market.

The article is organized as follows. First, in Section 2, we provide an overview of the literature on the determinants of default in mortgage credit markets and place our study in this context. Section 3 then discusses the empirical framework we utilize to analyze the impact of credit policies on aggregate default and introduces the data. Next, Section 4 presents preliminary results for the RD framework. Section 5 then introduces the semiparametric hazard model that we employ to shed some light on how loan-specific characteristics may impact default probabilities. Finally, Section 6 provides some concluding remarks.

---

1For example, Mian et al. (2012) find that the combination of a decline in house prices and a rise in household debt provides the main explanation for the onset, strength, and duration of the current consumption slump in the United States. Glick and Lansing (2010) show that several industrialized countries also exhibited a combination of growth in household leverage and house prices in the run-up to the crisis.
2. Literature Review

The theoretical underpinnings of the drivers of both prepayment and default risks lie in viewing a mortgage loan as an option. First introduced by Merton (1974), this approach implies that loan termination is a rational decision of the borrower. However, Merton’s original framework only allowed for default to take place at maturity. To address this issue, Black and Cox (1976) extended the model to consider the possibility that default occur before maturity—a possibility that, according to the model, takes place once the default option is in the money. More recently, the literature on mortgage default found that, when the call option is in the money, the prepayment hazard increases substantially (Yongheng, 1997; Archer et al., 2002; Ciochetti et al., 2002; Ciochetti, 2003). Analogously, the default hazard increases when the put option is in the money; that is, when negative equity arises. Yongheng (1997) argues that initial conditions, such as the loan-to-value ratio (LTV) at origination, may reveal the borrower’s risk preferences, while fluctuations of these conditions over time may give away to which extent it is profitable for the mortgagor to prepay or default. Using a competing risks framework, Deng et al. (2000) estimate both prepayment and default hazards jointly and propose a model that considers default as an irreversible decision, so that agents do not necessarily default as soon as home equity becomes negative. Under this framework, an individual may thus choose to defer his default decision in case house prices increase in the near future.

From a mortgagor’s perspective, the decision to default hinges on how onerous the monthly mortgage payment is, relative to the possibility that the house value eventually exceeds the balance on the mortgage (Foote et al., 2008). In this context, liquidity is essential in determining whether to either pay and wait for prices to recover or default. In this line of thought, Bhutta et al. (2010) show that strategic default has a threshold lower than zero for underwater default, which suggests that default costs are non-negligible and may account for the discrepancies between the option theory predictions and empirical tests of different models of default. Furthermore, they find that default is better explained by income shocks than by negative equity.

The set of explanations that depart from the original Merton-Black-Cox theory of default, such as income shocks, is often termed “trigger event explanations”. The alleged importance of such events has spawned a strand of literature that incorporates measures of loan affordability and attempts to rationalize the importance of such variables. Thus, the loan to income ratio (LTI), for instance, has been deemed a relevant measure of loan affordability, which captures the sensitivity of the probability to pay given a certain trigger event. Moreover, Campbell and Cocco (2011) argue that there are important interactions between the LTV and the LTI in determining the probability of default since the LTI determines the level of negative home equity that triggers default and thus operates through the same option theory mechanism, albeit indirectly.

Although the empirical literature seems to have reached some consensus on which factors may have a significant impact on mortgage credit default, the evidence suggests that their relevance depends on inherent characteristics of the specific mortgage market under study and of the products available to consumers. For instance, some of the earlier studies found that homeownership and default outcomes are highly sensitive to the evolution of house prices and to the LTV at origination (Furstenberg and Green, 1974; Morton, 1975). Although these results have been further confirmed in the subsequent literature, Vandell (1978) demonstrated that these results do hold for fixed nominal-interest rate constant-payment loans, but that they cannot be generalized for alternative instruments. In fact, Zorn and Lea (1989) argue that, in the case of the Canadian mortgage market, the estimated effects of some of the classical determinants of default tend to be smaller due to the adjustable-rate nature of the loans. Moreover, using mortgage loan data
in 12 E.U. countries, Diaz-Serrano (2004) shows that income volatility may be more important in explaining mortgage credit delinquency than traditional determinants of default, particularly for countries in which insurance markets are inefficient. These results emphasize the need to investigate the potential drivers of default in the Mexican context.

Another variable that has been shown to be crucial to explain mortgage loan default is the time since origination. In general, the empirical literature finds that the default hazard rate increases rapidly after the loan is granted, reaches its maximum a few months later, and then declines gradually (von Furstenberg, 1969; Gerardi et al., 2007). This is particularly evident for mortgages with high LTV ratios. As Ashenfelter and Layard (1987) and Hakim (1997) explain, an inverted U-shape in the hazard function can be generated by a mixture of hazard functions, but this imposes unobserved heterogeneity in the model. Therefore, using a semi-parametric or non-parametric approach has become standard in the literature.

A recurrent theme—implicit or explicit—in most studies in the earlier literature is whether individual characteristics of the mortgagor are important determinants of default. Although a large number of papers has addressed the issue of discrimination in credit markets (particularly based on race in the U.S.), there is substantial evidence that both lending decisions and default outcomes hinge mainly on economic indicators (Shaffer, 1996; Berkovec et al., 1998). When non-economic individual characteristics—such as the race of the mortgagor—have been found to be significant, it has been shown that this is attributable to unobserved economic factors that may be correlated with the non-economic variable, such as the credit history of the debtor. These economic variables are of particular importance when analyzing the impact on default of trigger events, such as divorce or unemployment. However, most data sets that are frequently used in papers that analyze default in mortgage markets generally exclude important economic characteristics of the borrowers and of the property. To address these concerns, Deng et al. (2005) have incorporated spatial heterogeneity in hazard models in order to capture unobserved factors such as culture, education or access to information. Their analysis finds important geographical effects, which they attribute to the geographical clustering of unobserved factors in the population, and thus they conclude that aggregation bias is responsible for some of the attenuation in the LTV coefficient. Nonetheless, Campbell et al. (2011) has argued that foreclosures sales also appear to have had negative feedback effects on the values of neighboring properties, worsening the decline in house prices. This line of reasoning would imply that geographical clustering is somewhat endogenous.

An important final consideration when studying default in mortgage markets is that the definition of default is important, and that relevant measures vary depending on the context and the question of interest of different studies. For example, some studies have defined default as non-payment or delinquency of over 60, 90, or more days (Ambrose and Capone, 2000); foreclosure proceedings initiation (Yongheng, 1997; Deng et al., 2000); and foreclosure sale of the property (Foote et al., 2008). Ciochetti et al. (2002) use different definitions on delinquency and finds that being delinquent for 60 days or more increases the likelihood of terminating the loan either by prepayment or by default. Danis and Pennington-Cross (2008) expand on this literature by recognizing that there are multiple states of delinquency and that these states are not independent of each other. Moreover, the span between delinquency and foreclosure may be driven by other considerations. As Ambrose and Yavas (2010) note, there are conflicting incentives between the master and special servicers when handling troubled loans in a CMBS deal. Having noted this, Chen and Deng (2012) focus on the determinants of the span between initial delinquency and foreclosure in commercial real estate and find that the expected cash flow is a key determinant of the strategy the special servicer will pursue when terminating the loan. In light of these re-
results, we consider using alternate definitions of default to ensure that our findings are robust to differences in measurement.

As can be inferred from the references discussed above, studies on default in mortgage markets in developing countries are scarce. By studying mortgage lending in Mexico, our paper attempts to fill this void in the literature, as well as better understand the factors that may lead consumers to delay (or ultimately default on) their mortgage payments. Given the existence of specific, well-defined credit restrictions imposed by credit-granting institutions in the country, our study also contributes to our understanding of how such restrictions may affect consumer behavior and, in particular, individual decisions to stop servicing mortgage debt. The following section explains in more detail the methodology that we use to this end.

3. Empirical Framework and Data

3.1. Regression Discontinuity

Estimating whether certain loan characteristics may lead to higher or lower default by the borrower (henceforth, he) is not necessarily straightforward, as different loan covenants, interest rates, loan amounts, and other terms of the lending agreement may depend themselves on individual borrower traits that may signal a higher or lower probability of nonpayment. For instance, whenever a bank or other credit institution (henceforth, she) issues a loan, the level of reserves that she is required to set apart to cover for potential losses is linked to her expected loss, which is calculated as the estimated probability of default for an individual borrower times the amount of funds lent by the bank. In this case, the overall amount granted by the credit institution is endogenous to the ex ante probability of default. Moreover, these calculations may have further impacts on other loan characteristics, for instance, if we believe that the size of the loan may in turn determine the value of the house an individual can afford and ends up purchasing. By observing individual characteristics of the borrower, such as his monthly income, the condition and location of his current dwelling, and his employment history and stability, a rational lender may restrict the amount that is lent out to cap her expected losses and required reserves, as well as the ex post probability of default of the mortgagor.

In the Mexican mortgage market, there exists at least one product type for which there is an exogenous (as will be shown below) monthly minimum wage threshold that changes certain loan characteristics that are available to the borrower. Using an RD design, these discontinuities allow us to estimate the causal effect that these characteristics, particularly the size of the loan and the value of the dwelling, have on the probability of default. In what follows, we explain the general modeling approach. The specific data used in the empirical estimation are introduced in Section 3.2.

In the model, a bank offers a loan to borrower $i$ based on his payment capacity, $\kappa_i$. Whether the individual borrower will be able (or willing) to actually serve his debt according to the loan agreement is unobserved by the econometrician. What is observed is a noisy signal of the borrower’s payment capacity, $I_i$. Given an unobserved random component $\nu$, this implies,

$$I_i = \kappa_i + \nu_i.$$

Let $L_i$ represent the set of covenants and loan-specific characteristics offered to borrower $i$, which is an increasing function of the individual’s payment capacity. That is,

$$L_i = \phi(\kappa_i), \quad \phi' > 0.$$
Now, assume that there exists an exogenous bank-imposed policy that leads to a discontinuous break in the characteristics of the loan offered to the borrower. In particular, if borrower $i$ exceeds a specific income threshold $c$, then the loan function is shifted by an amount $\gamma$:

$$L_i = \phi(k_i) + \gamma \cdot 1[I_i \geq c] + \eta_i,$$

where $\eta_i$ represents some unobserved heterogeneity component. Following Lee and Lemieux (2010), the discontinuity in income helps us identify (at least locally) the effects of changes in loan characteristics on the probability of default, $\delta_i$, our outcome of interest. Thus, the equation we estimate is:

$$\delta_i = f(I_i - c) + \tau \cdot 1[I_i \geq c] + X_i'\beta + \epsilon_i,$$

where $f(\cdot)$ is a polynomial in the distance between the borrower’s income and the cutoff, $X_i$ is a vector of covariates, and $\epsilon_i$ is an error term.

In principle, we would expect the probability of default $\delta_i$ to be a continuously decreasing function of an individual’s payment capacity, which is proxied by his income. In particular, would-be borrowers who earn roughly the same income, should all have roughly the same payment capacity, and so their probability of default should be about the same. However, as we argue below, this should not be the case for individuals with a monthly income of $11 \pm \epsilon$ mmw, for small $\epsilon > 0$, since those at and above the 11 mmw threshold will obtain a larger loan and purchase a more expensive property, which should in turn reduce their ability to serve their mortgage debt. If this did not occur, that is, if all borrowers within the $\epsilon$ neighborhood continued observing the same payment behavior, it would suggest that the credit restriction could then be relaxed without having a negative impact on delinquency and on creditors’ expected losses.

3.2. Data Description

The data used to estimate the model described above come from a proprietary database, the Mortgage Reports database, which gathers information that banks are required to report to the Mexican National Commission on Banking and Securities (CNBV)—the main institution in charge of regulation and supervision of financial intermediaries in Mexico. Banks are required to provide monthly information on all individual mortgage credits granted, including whether they were co-financed with other credit-granting institutions. In particular, a large share of bank mortgage loans are offered jointly with Infonavit, the main mortgage-granting institution in Mexico.\(^2\)

For each loan, banks must report several loan, property, and mortgagor characteristics, including the LTV, the debt-to-income ratio, the payment-to-income ratio, the price of the house, state and municipality in which the property is located, the frequency of payments, the issuing bank, the commercial name of the loan product, the interest rate and whether such interest rate is fixed or variable, the days of delay of payment, the date when the credit contract will expire (once credit is paid in full), the borrower’s income at the time of origination, the outstanding balance, the location of the dwelling, among many other relevant features of the credit contract and its development over time. Variables such as the commercial name of the product and the name of the co-financing institution allow us to identify which loans are operating under Infonavit’s specific rules and which ones are not.

The Mortgage Reports database began to be fully operational on July 2009 and is updated every month. These updates contain information on new credit contracts as well as any possible

\(^2\)Of the total amount of housing credit granted in Mexico, about $\frac{3}{4}$ involves Infonavit funds. These loans target insured formal employees of the private sector.
changes on existing loan characteristics—for instance, the remaining balance or whether the loan became delinquent in a given month. The database contains about 1 million records, of which approximately 400,000 originated after July 2009.

The specific mortgage product that interests us is Infonavit Total (IT), a loan that is offered jointly by Infonavit and a commercial bank. The advantage of this particular product over a traditional Infonavit loan is that borrowers get additional funds from the commercial bank, while facing the same costs (interest rate plus commissions) and term to maturity than they would otherwise. These additional funds allow individuals to obtain larger loans and potentially purchase more expensive properties. Under a traditional Infonavit loan, for example, borrowers may obtain a maximum loan amount equivalent to 180 times the current monthly minimum wage (or 180 mmws)—about 300,000 pesos, or 23,000 US dollars in 2012. In contrast, recipients of an IT loan may get up to 700 mmws. The interesting feature of this product is that both the size of the loan and the original value of the dwelling depend on the borrower’s income level. In particular, if the individual earns less than 11 mmws, there are caps of 305 mmws for the IT loan amount he may get and 350 mmws for the price of the property he may choose. Individuals whose monthly income is equal or exceeds the 11 mmw mark obtain a variant of the IT loan called Infonavit Total AG (ITAG), which offers loan amounts of up to 700 mmws and faces no cap in terms of the price paid for the dwelling.

Thus, assuming would-be borrowers in the neighborhood around the 11 mmw mark are otherwise identical, the Infonavit-imposed income threshold then determines the size of the loan (and, presumably, the price of the sought-after property) individuals end up getting. Within the RD framework, exceeding the cutoff \( c = 11 \) mmw leads to the treatment \( \tau \), that is, getting access to the “unconstrained” credit product. Under the assumption that the probability of default \( \delta_i \) is continuous in the income variable, \( I_i \), the average treatment effect \( \tau \) satisfies:

\[
\tau = \lim_{\iota \downarrow 11} \mathbb{E}[\delta_i | X_i, I_i = \iota] - \lim_{\iota \uparrow 11} \mathbb{E}[\delta_i | X_i, I_i = \iota]
\]

where \( X \) represents the set of covariates and is not affected by belonging to the treatment.

### 3.3. Summary Statistics

Our sample consists of all new mortgage credits granted by commercial banks between July 2009 and April 2012 for the purpose of acquiring a new or existing dwelling. Although mortgage loans destined to other purposes, such as land procurement or home repairs, are also contained in the original database, these are excluded from the working data set given that their characteristics differ significantly from those used for purchasing a home—and because Infonavit Total loans can only be used for home purchases. The total number of mortgage credits given out by commercial banks within this period was 293,661. Descriptive statistics for the working sample appear in Table 1.

On average, the loan-to-value ratio is 0.738. Approximately, the first 25th percentile of the sample has loan-to-value less than 0.654, the 50th percentile has LTV less than 0.8, whereas

---

3In addition to the loan amount, the borrower must use the outstanding balance in his Infonavit housing account, an individual fund to which employers must make regular deposits to their active workers, which can eventually be used by individuals to obtain a mortgage.

4In principle, applicants are accepted or rejected according to the same criteria that Infonavit uses for all loans. Potential borrowers need not take their application documents to individual banks; the whole application and enrollment process is carried out by Infonavit.
Table 1: Summary Statistics

<table>
<thead>
<tr>
<th>Variables</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Loan-to-Value (LTV)</td>
<td>0.738</td>
<td>0.206</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0&lt;LTV≤0.654</td>
<td>0.266</td>
<td>0.441</td>
<td>0.003</td>
<td>1</td>
</tr>
<tr>
<td>0.654&lt;LTV≤0.8</td>
<td>0.231</td>
<td>0.421</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0.8&lt;LTV≤0.88</td>
<td>0.258</td>
<td>0.437</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0.88&lt;LTV≤0.9499</td>
<td>0.144</td>
<td>0.351</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>LTV&gt;0.9499</td>
<td>0.099</td>
<td>0.299</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Co-financed loan==1</td>
<td>0.347</td>
<td>0.476</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Monthly income</td>
<td>31907.71</td>
<td>30052.38</td>
<td>4500</td>
<td>255794</td>
</tr>
<tr>
<td>Monthly amount due for payment</td>
<td>7012.077</td>
<td>6473.382</td>
<td>28</td>
<td>47722</td>
</tr>
<tr>
<td>Unpaid balance</td>
<td>556852.5</td>
<td>436816.1</td>
<td>38527</td>
<td>2903723</td>
</tr>
<tr>
<td>Interest rate</td>
<td>10.888</td>
<td>1.260</td>
<td>9</td>
<td>20.7</td>
</tr>
</tbody>
</table>

Total number of observations is 293,661.
Statistics correspond to information as of April 2012.

the 75th percentile has LTV less than 0.88. The percentage of the sample that has loan-to-value
between 0.88 and 0.94 corresponds to 14%. Finally, the last 10% has LTV greater than 0.94.
From all credits in our sample, 34% of them are co-financed either with Infonavit or Fovissste. As
explained earlier, the monthly income of the mortgagor is an important economic characteristic
that banks explicitly use to determine to which mortgage products consumers have access. The
average monthly income as of April 2012 was 31,000 pesos. In this dataset, banks report the
monthly amount due for payment for each credit. On average, this amount is 7,000 pesos. Banks
also report the unpaid balance left for each loan, this variable is on average 550,000 pesos as of
April 2012. The average interest rate charged in April 2012 for each credit is 10%.

4. Effect of Credit Policies on Default

Before we estimate the model, we need to show the RD design is valid. We follow Imbens
and Lemieux (2008) and present graphical evidence to this end. First, we show that there is no
cluster of borrowers with an IT/ITAG loan whose income is just above 11 mmws. In case a cluster
of mortgage loans was actually observed past the 11 mmw mark, it would suggest potential
self-selection into the ITAG product or that individuals are somehow rigging their application
documents to get access to the mortgage product that offers them significantly looser credit terms
and the possibility of purchasing a more expensive property. Figure 1 shows this is not the case.
The figure on the left shows the whole income distribution of borrowers that obtain any variant
of the IT loan. As can be observed, the density function is continuously decreasing around the
11 mmw income cutoff. This is further confirmed in the right panel, where we zoom in around
the threshold of interest. Thus, the evidence seems to suggest that the forcing variable, income,
is not being manipulated by individuals.
Next, we show that, in spite of no apparent differences in borrowers’ observable characteristics, such as income, there exist significant differences in the mean value of the dwelling purchased by individuals below and above the income threshold. This is illustrated in Figure 2. The figure shows the average value of the house at the time of loan origination for individuals within very small income bins. The difference in value for people just below and just above the 11 mmw mark is about 50,000 pesos, or about 10% of the original price of the property, and this discontinuity is statistically significant. As explained earlier, this jump is explained by Infonavit’s credit policy and the fact that borrowers with incomes lower than 11 mmw have a restriction on the value of the property they can purchase with the obtained credit, while individuals at and above this cutoff point, do not.

---

5The income bandwidth is 0.05 mmws. Results are robust to different income bin sizes.
As explained before, the fact that individuals who face looser credit conditions are able to purchase a more expensive property should reduce their payment capacity, which should translate into a higher incidence of nonpayment. The evidence in Figure 3 suggests that this is not the case. The graph is similar to 2, except here we illustrate the average non-performing loan ratio (by income bins) as a proxy for default. Notice that for individuals with incomes below the cutoff of interest, the mean non-performing loan ratio is inside the 95% confidence interval constructed for those individuals with incomes above the threshold. This implies that the alleged differences in default by income level are not significant around the 11 mmw cutoff. In fact, if anything, the average probability of default is slightly lower for individuals who earn 11 mmws relative to those who earn, say 10.9 mmws. Our interpretation of this partial evidence is that credit restrictions could then be loosened without any negative consequences on creditors’ expected losses.

To better understand what may be driving these similarities in delinquency rates in spite of the looser credit restrictions for borrowers at or above the 11 mmw cutoff, in the next section we investigate the determinants of default.
5. Determinants of the Probability of Default

5.1. Semiparametric Hazard

To study the determinants of the probability of default, a flexible hazard model is estimated for the duration of a mortgage credit before it becomes a non-performing loan, that is, before it completes 90 consecutive days without making the corresponding payments. Our interest is on the effects of different factors on the probability that a mortgage credit become non-performing between month $t$ and $t+1$, given that it has been in good standing up to $t$. The method used here is the same presented in Meyer (1990). In this case the unit of observation are the mortgages. In this paper a mortgage is censored if it is still a performing loan during the last month it is observed, which in this case corresponds to April 2012. Also, mortgage credits for which no payment has been registered, but they have not completed 90 days in this situation, are censored. Finally, mortgage credits observed for more than 31 months are censored at 31.

The main benefit of using a flexible hazard model is that no assumptions about the distribution of the duration spell are necessary to estimate the hazard, which contrasts with standard parametric hazard models like the Weibull or the logistic models. The semiparametric hazard model naturally allows for time-dependent covariates, which may not be relevant for time-invariant loan characteristics, such as the type of interest rate (fixed or variable) or the amount of the monthly payment if the loan is fixed-rate. However, this flexibility is important to assess the relevance of time-dependent variables, such as macroeconomic conditions.\footnote{For more detailed explanations of these and other advantages of the semiparametric hazard model described below, please refer to Meyer (1990).}

Let $T_i$ be the duration of credit $i$ as a performing loan, that is, the time the mortgagor keeps making the corresponding payments or if he has stop making these payments but has not yet completed 90 days without paying. Then, the hazard in this case is defined as the probability
that credit $i$ becomes a non-performing loan between month $t$ and month $t + 1$, given that mortgage $i$ has survived as a performing loan through month $t$. With this definition, the hazard is parameterized using a proportional hazard form in the following way.

Let $\lambda_i(t)$ be the baseline hazard at time $t$, $x_i(t)$ be the vector of (possibly) time varying explanatory variables for mortgage $i$ to become a non-performing loan, and $\beta$ be the vector of parameters. Then the hazard function for loan $i$ is:

$$\lambda_i(t) = \lambda_0(t) \exp(x_i(t)'\beta).$$

(1)

Using equation (1) we can write down the probability that a duration spell lasts until time $t + 1$ given that it has lasted until $t$.

Using the fact that $x_i(t)$ is constant in the interval $[t, t + 1)$ and the following definition from Meyer (1990),

$$\gamma(t) = \log \int_t^{t+1} \lambda_0(u) du,$$

(2)

the probability of a loan becoming a non-performing loan in the first $k_i - 1$ intervals can be written as:

$$\prod_{t=1}^{k_i-1} \exp[\exp[\gamma(t) + x_i(t)'\beta]].$$

(3)

Moreover, the probability that duration $T_i$ falls into interval $k_i$, is given by:

$$1 - \exp[-\exp[\gamma(k_i) + x_i(k_i)'\beta]].$$

(4)

Using the probabilities defined in equations (3) and (4), the log-likelihood function for a sample of $N$ mortgages is:

$$L(\gamma, \beta) = \sum_{i=1}^{N} \left[ \delta_i \log[1 - \exp(-\exp[\gamma(k_i) + x_i(k_i)'\beta])] - \sum_{t=1}^{k_i-1} \exp[\gamma(t) + x_i(t)'\beta] \right]$$

(5)

where $\gamma = [\gamma(1), ..., \gamma(T_i)']$, $C_i$ is the censoring time for loan $i$, $\delta_i = 1$ if $T_i \leq C_i$, i.e. the observation is censored, and 0 otherwise, and $k_i = \min(int(T_i), C_i)$.

5.2. Estimation

We now estimate the determinants of the probability of a mortgage loan to become non-performing. Before explaining our results, two clarifications must be made. First, notice that we used the information available for each loan as of April 2012. We have not yet used the time variant features of the data set. In principle, we have time variant information for each loan since its inclusion in the database, such as the interest rate applied each period the loan is observed. Second, the hazard is constructed according to the number of months the loan has been a performing loan since it first entered the database, although loans entered the sample in different calendar months. In other words, all durations are aligned to the same starting point because all that matters is the number of periods before they become non-performing loans.

Table 2 summarizes the data to construct the empirical hazard function. The risk set at the beginning of month $t$ refers to the number of loans for which the spell has not ended or has been censored at the beginning of month $t$. Total exits refers to the number of mortgages that
Table 2: Failures, Censoring, and the Kaplan-Meier Empirical Hazard

<table>
<thead>
<tr>
<th>Month / Risk Set for sale</th>
<th>Exits</th>
<th>Censoring</th>
<th>Hazard</th>
<th>Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 286,040</td>
<td>2</td>
<td>7621</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>2 277,233</td>
<td>4</td>
<td>8805</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>3 269,310</td>
<td>10</td>
<td>7919</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>4 264607</td>
<td>2248</td>
<td>4693</td>
<td>0.008</td>
<td>0.001</td>
</tr>
<tr>
<td>5 254932</td>
<td>4042</td>
<td>7427</td>
<td>0.016</td>
<td>0.001</td>
</tr>
<tr>
<td>6 241905</td>
<td>3517</td>
<td>8985</td>
<td>0.015</td>
<td>0.001</td>
</tr>
<tr>
<td>7 232358</td>
<td>3564</td>
<td>6030</td>
<td>0.015</td>
<td>0.001</td>
</tr>
<tr>
<td>8 222807</td>
<td>3222</td>
<td>5987</td>
<td>0.014</td>
<td>0.001</td>
</tr>
<tr>
<td>9 213660</td>
<td>3718</td>
<td>5925</td>
<td>0.017</td>
<td>0.002</td>
</tr>
<tr>
<td>10 204740</td>
<td>3162</td>
<td>5202</td>
<td>0.015</td>
<td>0.001</td>
</tr>
<tr>
<td>11 195476</td>
<td>3651</td>
<td>6102</td>
<td>0.019</td>
<td>0.002</td>
</tr>
<tr>
<td>12 186278</td>
<td>3389</td>
<td>5547</td>
<td>0.018</td>
<td>0.002</td>
</tr>
<tr>
<td>13 178272</td>
<td>2742</td>
<td>4617</td>
<td>0.015</td>
<td>0.002</td>
</tr>
<tr>
<td>14 168922</td>
<td>3258</td>
<td>6608</td>
<td>0.019</td>
<td>0.002</td>
</tr>
<tr>
<td>15 160948</td>
<td>2892</td>
<td>4716</td>
<td>0.018</td>
<td>0.002</td>
</tr>
<tr>
<td>16 154226</td>
<td>1582</td>
<td>3830</td>
<td>0.010</td>
<td>0.001</td>
</tr>
<tr>
<td>17 145820</td>
<td>4140</td>
<td>6824</td>
<td>0.028</td>
<td>0.002</td>
</tr>
<tr>
<td>18 135083</td>
<td>10602</td>
<td>6597</td>
<td>0.078</td>
<td>0.004</td>
</tr>
<tr>
<td>19 117802</td>
<td>351</td>
<td>6679</td>
<td>0.003</td>
<td>0.001</td>
</tr>
<tr>
<td>20 112351</td>
<td>384</td>
<td>5100</td>
<td>0.003</td>
<td>0.001</td>
</tr>
<tr>
<td>21 106493</td>
<td>1007</td>
<td>5474</td>
<td>0.009</td>
<td>0.001</td>
</tr>
<tr>
<td>22 100385</td>
<td>3696</td>
<td>5101</td>
<td>0.037</td>
<td>0.003</td>
</tr>
<tr>
<td>23 91298</td>
<td>3818</td>
<td>5391</td>
<td>0.042</td>
<td>0.004</td>
</tr>
<tr>
<td>24 82585</td>
<td>3304</td>
<td>4895</td>
<td>0.040</td>
<td>0.004</td>
</tr>
<tr>
<td>25 75080</td>
<td>3390</td>
<td>4201</td>
<td>0.045</td>
<td>0.004</td>
</tr>
<tr>
<td>26 66494</td>
<td>2644</td>
<td>5196</td>
<td>0.040</td>
<td>0.004</td>
</tr>
<tr>
<td>27 59712</td>
<td>2890</td>
<td>4138</td>
<td>0.048</td>
<td>0.005</td>
</tr>
<tr>
<td>28 53692</td>
<td>2407</td>
<td>3130</td>
<td>0.045</td>
<td>0.005</td>
</tr>
<tr>
<td>29 45041</td>
<td>10879</td>
<td>6244</td>
<td>0.242</td>
<td>0.011</td>
</tr>
<tr>
<td>30 29570</td>
<td>292</td>
<td>4592</td>
<td>0.010</td>
<td>0.003</td>
</tr>
<tr>
<td>31 1934</td>
<td>1934</td>
<td>27344</td>
<td>1.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>
became nonperforming during month $t$. For instance, according to the table, a total of two loans became nonperforming during the first month after origination. As mentioned before, censored exits refer to the number of loans that were still performing in the last period they were observed. For example, being censored in the first month implies that the credit is considered a performing loan during the first month, and that was the last period they were recorded in the sample.

The hazard shown in the table corresponds to the Kaplan-Meier empirical hazard for our sample. The empirical hazard is the fraction of spells ongoing at the start of a month which end during the month. The empirical hazard is almost zero for the first three periods. This is related to the definition of nonperforming, since it takes 90 days without making the corresponding payments to become a non-performing loan. Then the hazard increases to stay between 0.010 and 0.016 for periods 5 to 16. At month 18, the hazard increases to 0.078. Then the hazard decreases to almost zero for the next two periods. Between periods 22 to 28, the hazard increases to almost 0.050. Surprisingly, the hazard increases up to 0.242 at the 29th month. Finally, the hazard is one in the last period because no observations last more than 31 periods. Therefore, once subtracting the censored observations from the risk set, all credits left become non-performing loans in that period.

To obtain the empirical hazard, the log-likelihood function of Equation 5 is maximized through standard techniques with respect to the 31 elements of $\gamma$ and the vector $\beta$. Using a random sample of 10,000 observations, our results suggest that the hazard of a mortgage loan becoming nonperforming is, in general, an increasing function of the LTV—at least for LTVs lower than 0.95. The results in column (1) of Table 3 show the coefficients of our flexible hazard model, using all loans with LTV above 0.95 as a reference point. In particular, in our preferred specification, column (3), the hazard decreases 38% for loans with lowest LTV (between 0 and 0.65) relative to those with highest LTV (above 0.95). Loans in each of the subsequent LTV categories have a larger probability of becoming nonperforming, according to the hazard function, than loans in previous categories. This is not surprising: as individuals obtain a bigger loan to acquire a property of a given price, the value of collateral is relatively lower and individuals may have more incentives to default. It is interesting to note that mortgage loans in the middle of the LTV distribution have a higher hazard relative to credits in the top LTV category. This implicit inverted U-shape may be related to the differences in risk profiles of individuals in different LTV categories. For instance, it is possible that very large mortgage credits are only given out to high net worth individuals who have an inherently lower risk profile. Although we do not observe individual characteristics (other than income and location of the mortgaged property) that could confirm this hypothesis, it is possible that other loan-specific characteristics, such as the interest rate, may provide some relevant information.\footnote{We are currently working on a version of the model that takes into account this and other elements.}

The results in Table 3 also show that cofinanced loans—that is, those that are jointly granted between a commercial bank and a public institution such as Fovissste or Infonavit—are significantly more likely to become nonperforming. The fact that cofinanced credits are, in general, given out to lower-income individuals who may not have access to only-bank loans could potentially explain this result.\footnote{This will also be verified in a future version of this paper.}

It is important to emphasize that the results shown in Table 3 are preliminary and that a number of additional considerations must be made. For instance, none of the columns in the table include state or loan origination date indicators. The former are necessary to control for potential unobserved heterogeneity at the regional level. The latter are useful to control for
<table>
<thead>
<tr>
<th>Variables</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0&lt;LTV≤0.654</td>
<td>-0.012</td>
<td>–</td>
<td>-0.379**</td>
</tr>
<tr>
<td></td>
<td>(.079)</td>
<td>–</td>
<td>(.079)</td>
</tr>
<tr>
<td>0.654&lt;LTV≤0.8</td>
<td>0.583**</td>
<td>–</td>
<td>0.246**</td>
</tr>
<tr>
<td></td>
<td>(.071)</td>
<td>–</td>
<td>(.072)</td>
</tr>
<tr>
<td>0.8&lt;LTV≤0.88</td>
<td>0.511**</td>
<td>–</td>
<td>0.381**</td>
</tr>
<tr>
<td></td>
<td>(.070)</td>
<td>–</td>
<td>(.071)</td>
</tr>
<tr>
<td>0.88&lt;LTV≤0.9499</td>
<td>1.251**</td>
<td>–</td>
<td>0.751**</td>
</tr>
<tr>
<td></td>
<td>(.069)</td>
<td>–</td>
<td>(.071)</td>
</tr>
<tr>
<td>Cofinanced loan==1</td>
<td>–</td>
<td>1.456**</td>
<td>1.408**</td>
</tr>
<tr>
<td></td>
<td>–</td>
<td>(0.039)</td>
<td>(0.040)</td>
</tr>
</tbody>
</table>

Log-likelihood  -1.326  -1.279  -1.258
Sample Size 10000 10000 10000

Standard errors in parentheses.
Estimated γ vector not included in this table, but available upon request.
** p<0.01; * p<0.05.

differences in duration distributions that may be attributed to seasonal factors or to the phase of the business cycle in which a loan is granted. We expect to cover these relevant aspects in full detail in the near future.

6. Final Remarks

The evidence presented above suggests that, at least for a particular mortgage product in the Mexican housing loan market, credit restrictions do not seem to be limiting the extent to which low-income individuals default on their loans. On the contrary, the preliminary results from the RD design may be indicative of potential benefits reaped by both credit institutions and would-be borrowers should credit constraints be relaxed and lower-income individuals were granted larger loans, that would allow them to purchase more expensive dwellings, if so desired.

These findings point towards the need to better understand the determinants of delinquency in the Mexican mortgage market. The results from the hazard model are indicative of important effects of loan-specific characteristics on the probability that a loan becomes nonperforming. In particular, and in agreement with the literature, the LTV seems to be an important determinant of default. Although this may seem at odds with our RD results, this is not necessarily the case, as IT borrowers past the 11 mmw threshold may be offered a larger loan that could potentially be used to finance a more expensive property, hence having an ambiguous effect on LTV. The findings in the hazard estimation also show that cofinanced products are more susceptible of becoming delinquent, perhaps due to the fact that these loans are generally offered to lower-income individuals with potentially higher risk profiles.

As was acknowledged since the beginning of this article, the analyses presented thus far in this version of the paper are not conclusive. We are currently working on an estimation of the RD model that takes into account other mortgage products that also use income-based thresholds to restrict credit conditions. A more thorough description of the RD estimation that includes
all other controls should help validate our preliminary results. In addition, the hazard model should include other potentially-relevant determinants, according to the literature. In the near future, this extension of the paper will include variables such as the interest rate on the loan, the remaining balance on the mortgage, the value of the property, the original loan amount, the frequency and the amount of periodic payments, the time to maturity, as well as state and starting date indicators.
References


