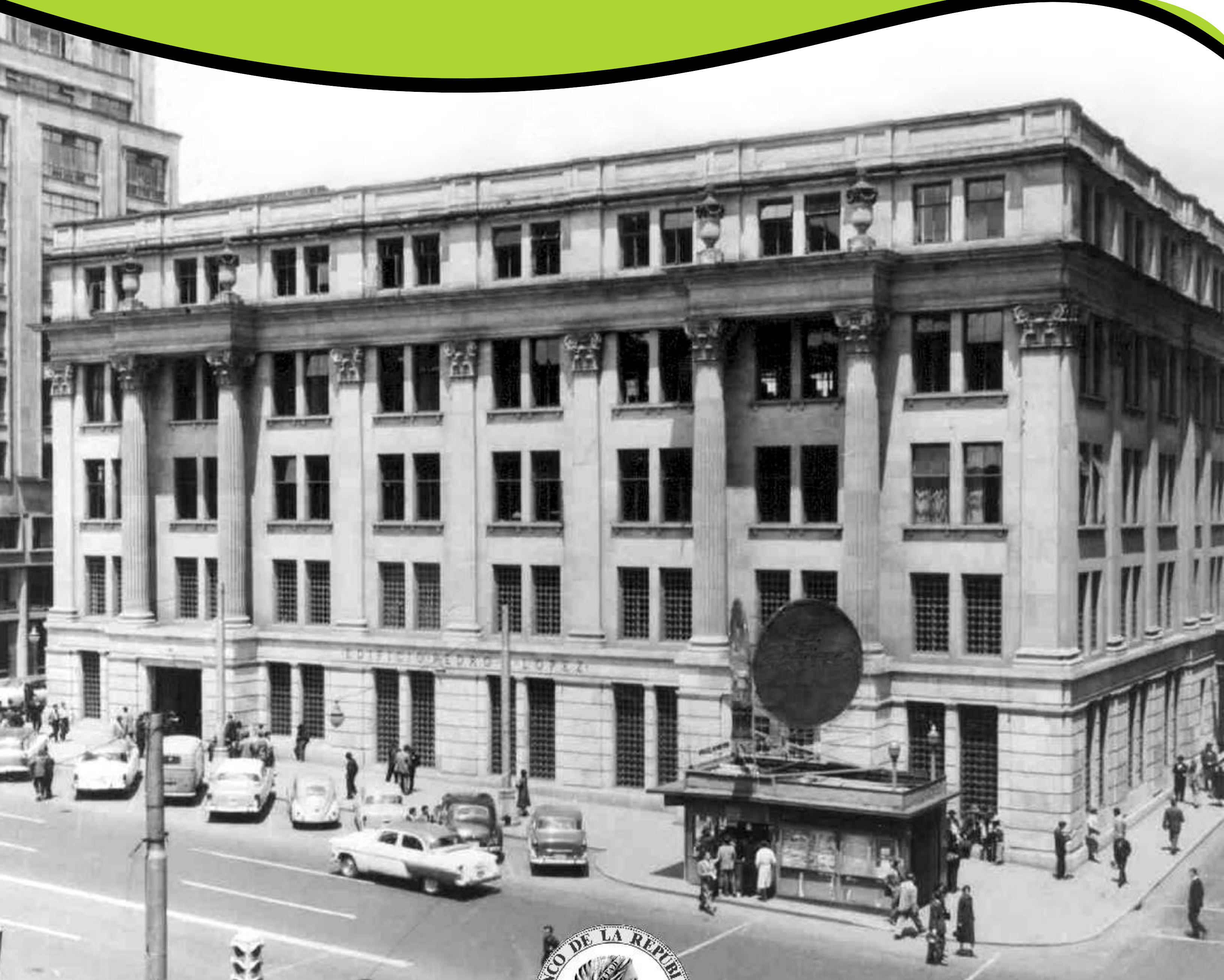
An empirical characterization of mortgage default in Colombia between 1997 and 2004

Por : Juan Esteban Carranza
Dairo Estrada

Borradores de ECONOMÍA

No. 450
2007



tá - Colombia - Bogotá - Colombia - Bogotá - Colombia - Bogotá - Colombia - Bogotá - Colombia - Bogotá - Colombia - Bogotá - Col

An empirical characterization of mortgage default in Colombia between 1997 and 2004

Juan Esteban Carranza

Dairo Estrada*

Abstract

This paper examines the relationship between mortgage default decisions and relevant observable variables under the light of a random utility model. The focus of the study is the Colombian mortgage market between 1997 and 2004 using two separate data sets that are matched using simulation techniques. The estimation allows the computation of mortgage default probabilities that are directly related to an underlying model of optimal default. Results are sharp and indicate that variation in current income has little effect on mortgage default, compared to housing prices and mortgage balances.

JEL Classification: D4; G21; L13; R12

Keywords: Banking; Location; Competition; Colombia.

---

*University of Wisconsin and Banco de la República-Cali and Banco de la República-Bogotá, respectively. We are grateful to Oscar Leiva at Titularizadora Colombiana for providing the primary data set. We also thank Julio Escobar, José Eduardo Gómez and Mario Gonzalez and staff at Banco de la República in Cali and Bogotá for data support and insights. Only the authors are responsible for remaining errors.

1 Introduction

During the late 1990's the Colombian economy, similarly to several other emerging economies, experienced a severe financial crisis and economic slowdown. The effects of such crisis were fuelled by a dramatic increase in the default rates of mortgage holders, leading to the collapse of several major financial institutions and a major crisis that persisted for years after the end of the crisis. The behavior of debtors was affected by several separate factors: on one hand, incomes fell, making it difficult for many households to fulfill their payment obligations; on the other hand, debt balances which were tied to a market interest rate increased as the monetary authority stepped in to contain the exchange rate. Simultaneously, and as the crisis ensued, the prices of real estate, which had risen to unprecedented levels since the mid 1990's, fell dramatically. The estimated cumulative cost of the crisis is above 10% of GDP.

The specific objective of this paper is to evaluate the relationship between individual default and other observable variables under the light of a simple structural model. The estimates can the be used to evaluate counterfactual equilibria under the assumptions of the model. The estimates can also be related to a more general dynamic default model.

Under given assumptions, this general model nests simpler models that can be estimated using variations of standard discrete choice models. It can be shown, for example, that the standard logit and probit models correspond to restricted versions of the model that ignore unobserved correlated shocks and cannot incorporate non-matching survey information. This paper presents estimations of default probabilities using two variations of the logit model that account for unobserved correlated state variables and for the variation of income as measured in household surveys.

Even though some of the factors mentioned above, such as the interest rate, are relatively exogenous in the sense that they were regulated by the economic authority, they are the result of a complicated macroeconomic equilibrium, whose understanding is beyond the scope of this work. This is therefore a partial equilibrium analysis, in the sense that it doesn't incorporate the higher order effects of economic policy and default behavior on the economic activity, but is nevertheless a very useful benchmark for understanding the short-run mechanics of the crisis and provides a formal framework to a discussion that is not only relevant in Colombia but also in other emerging economies.

The empirical literature on mortgage default is dominated by variations of duration models as in Deng et al (2002). The advantage of duration models over empirical models based on individual likelihood functions is that they can be estimated even if the default rates are very low. On the other hand, the estimated parameters obtained from a duration model have no clear connection to a behavioral model, much less a dynamic behavioral model. Therefore, its use as a tool for counterfactual analysis is limited.

The structural methodology proposed in this paper provides estimates of the underlying behavioral model, so that counterfactual experiments are conceptually clearer. In a manner that is consistent with the existing literature based on duration models, the model on which the estimation is based is static, in the sense that no attempt is made to solve the dynamic problem of debtors. In contrast to standard duration models, it will be clear from the model what the implications of the assumptions are.

The application is based on a data set that contains the basic characteristics and payment histories of more than 15000 mortgages that were outstanding between 1998 and 2004. The estimation of the model based on individual likelihood functions is feasible, thanks to the very high rates of default observed during the late 1990's.

An additional difficulty of the empirical exercise lies on the fact that no matching information is available on the income histories of debtors in the data set. Therefore, a secondary data set obtained from household survey that matches housing payments and the value of real estate is used to compute the default probabilities from simulated draws of income.

The estimation of the model yields default probabilities with highly significant coefficients. The results indicate that short- term income variation has significant but very small impact on default. Default is mostly affected by the variation of housing prices and of mortgage balances. The implications of this conclusion are discussed at the end of the paper.

The sharp separate identification of the structural parameters relies partly on some singular institutional features of the Colombian mortgage market during the time span of the sample. First, the terms of the individual mortgages didn't use to be negotiated between debtors and financial institutions. In general, mortgage terms were negotiated between developers or construction companies and the mortgage banks, and the terms of the mortgage were transferred to any house buyers, who qualified according to a simple income rule. Second, while mortgage payments were based on a fixed rate on the balance of the loan, this balance was indexed to the market interest rates according to a formula that was determined by the Central Bank[1]. Therefore, the data contains enough exogenous variation to identify a detailed structural model.

The next two sections describe the Colombian mortgage market and the general formal framework. The later section discusses the data, the specifics of the empirical model and the results of the empirical exercise.

2 The macroeconomic environment

As indicated above, the collapse of the mortgage market was part of a larger macroeconomic crisis that hit Colombia during the late 1990's. In this section we illustrate three macroeconomic features of this crisis that were arguably interrelated: the evolution of

[1]Specifically, payments were made according to a fixed rate over the balance of the loan. But then the balance of the loan increased month by month according to a rate fixed by the Central Bank; it used to be that this rate of increase was tied to the rate of inflation but since the early 1990's it was tied to a market interest rate. In any case, the rate was set somewhat arbitrarily by the Central Bank.

GDP, interest rates, asset prices and the exchange rate.

Figure 1 shows the drastic decrease of GDP during 1998-1999 and the low rate of growth after the crisis. Simultaneously, the interest rate increased dramatically: figure 2 displays the annualized overnight interest rate charged by banks, which reached levels above 40% in 1998. As the model will illustrate, the sudden and dramatic rise of the interest rates was associated with an equally dramatic worsening of the quality of banks' assets due to increased default. Many banks became insolvent and had to be intervened by the government

The collapse of the market and the fall of households' income was accompanied by a decrease in the value of homes. During the first years of the decade, the elimination of financial restrictions allowed a rapid growth of home prices. This caused a bubble that took prices far away from its long run values. The evolution of housing prices before, during and after the crisis is clear in figure 3[2].

Finally it is important to highlight the relationship of the conditions in the mortgage market with the evolution of foreign exchange market, even though they didn't affect directly the Colombian households. The financial crisis was in fact triggered by a massive run on the Colombian peso by international investors that forced the Central Bank to first increase interest rates to defend the exchange rate. Once the exchange rate was allowed to float, it depreciated dramatically as illustrated in figure 4.

The correlation of the difficulties faced by the mortgage market and other macroeconomic variables is evident. The model below will illustrate and quantify the causal linkages between the different factors at the household-level using an economic model.

[2]Home price indices were estimated by Escobar et al. (2006)

3 A behavioral model of mortgage default and the estimation strategy

The model below is to be implemented with data from the Colombian mortgage market between 1997 and 2004. The most salient features of the Colombian mortgage financing system in those days was that all mortgages had variable rates tied to a regulated interest rate determined by the Central Bank with a formula that was tied to the market deposit rate. Therefore, refinancing was not an easy option for mortgage holders. The (total or partial) prepayment option will be ignored, as prepayments did not seem to be empirically significant (relative to defaulting) during the time-span of the sample and have less social implications than default[3].

We will study the behavior of mortgage holders ("debtors") who live in the mortgaged piece of real estate ("home"). The utility that a debtor i gets from the home depends on a measure x_i of subjective home quality. It also depends on the difference between household income and mortgage payments $Y_{it} - R_{it}$ and an idiosyncratic preference shock ε_{it} which incorporates unobserved (to the econometrician) variables that affect default, e.g. home attributes that are only valued by its owner and other preference shocks that vary across consumers and time. The estimation of the model is ultimately based on the properties of these unobserved variables.

A debtor will choose to default on her mortgage if the utility of making the mortgage payments and staying in her home is lower than the utility generated by not making the loan payment at the time. This alternative, which we will broadly call "default", gives rise to a complex scenario. Specifically, the individual may just be waiting to see whether the following period she can pay back her dues; she may try to sell the home and cash the difference between price and loan balance; she may return it to the bank

[3]From the perspective of the lending institutions, prepayments are quite relevant. The discussed methodological framework can incorporate prepayments easily.

to cover her obligation; finally, she could also just stop making payments indefinitely and face forfeiture or a renegotiation with the bank. Given the complexity of this alternative (which may also lead to moral or reputational costs), we will use a reduced form for its associated payoff.

Let V_{it} be the value for a debtor of not defaulting on her loan at time t. For debtor i we can write the value of the decision problem at time t as follows:

$$V_{it} = max\{u(x_i, Y_{it} - R_{it}, \varepsilon_{it}) + \beta EV_{it+1}, W_{it}\} \tag{1}$$

where $u(.)$ is the instant payoff from consuming the home at period t. W is the value of default which is the weighted sum of payoffs across the complex set of random scenarios discussed above.

An important feature of (1) above is the presence of the continuation payoff βEV_{it+1}, which can be understood to be the value of the option of defaulting in the future; on the other hand, defaulting gives also rise to continuation values that depend in general on the expected evolution of the state variables. We will consider first a model in which debtors are myopic in the sense that they ignore the dynamics of the state variables. This simplification may lead to biased estimates in certain circumstances. For example, if the expected evolution of the state space is such that individuals may choose to delay default decisions that appear to be optimal in a static environment, the model will wrongly infer that the consumer has a low marginal utility of income. It is therefore going to be a good approximation of reality in circumstances where the state variables are stationary or where the correlation of the continuation payoffs and the states is stable across equilibria.

In order to fix ideas, let the static utility be additively separable on observable and unobservable states:

$$u(x_i, Y_{it} - R_{it}, \varepsilon_{it}) = \theta_0 + \gamma_i x_i + \alpha_i(Y_{it} - R_{it}) + \varepsilon_{0it} \tag{2}$$

A debtor i will choose to continue paying her dues if the utility of doing so is higher than the utility of default:

$$\theta_0 + \gamma_i x_i + \alpha_i(Y_{it} - R_{it}) + \varepsilon_{0it} + \beta EV_{it+1}(\bar{P}_{it+1}, K_{it+1}, Y_{it+1}, \varepsilon_{0it+1}, \varepsilon_{1it+1})$$

$$\geq W_{it}(\bar{P}_{it}, K_{it}, Y_{it}; \omega_i) + \varepsilon_{1it} \quad (3)$$

where the payoff of default is assumed to be a function of the expected price $\bar{P}_{it}$ of the home at time t, the balance K_{it} of the debt and the debtor's income Y_{it}. These are variables that enter directly the payoffs of the individual scenarios that may arise after a default decision, as discussed above. In consequence, the continuation payoff depends on all observed and unobserved state variables.

Let $N_{it} = 1$ be the event that debtor i does *not* default at time t. The individual probability of defaulting is the probability that (3) is true. By specifying a parametric distribution for ε we can obtain the individual choice probabilities:

$$Prob[N_{it} = 1] = Prob[\theta_0 + \gamma_i x_i + \alpha_i(Y_{it} - R_{it}) + \varepsilon_{0it} + \beta EV_{it+1}(.) \geq W_{it}(.) + \varepsilon_{1it}]$$

$$= Prob[\theta_0 + \gamma_i x_i + \alpha_i(Y_{it} - R_{it}) - W_{it}(.) + \beta EV_{it+1}(.) \geq \bar{\varepsilon}_{it}] \quad (4)$$

Potentially, the model above can be estimated using maximum likelihood after parameterizing $W(.)$ and specifying the distribution of the error term $\bar{\varepsilon}_{it} = \varepsilon_{1it} - \varepsilon_{0it}$. For example if the errors $\bar{\varepsilon}_{it}$ are assumed to be *iid* draws from an extreme value distribution and are assumed to be conditionally independent from the observable states (as in Rust, 1987), the choice probabilities (3) have an analytical solution given by the usual logit form[4]:

$$Prob[N_{it}] = \frac{e^{\theta_{i0} + \gamma_i x_i + \alpha_i(Y_{it} - R_{it}) - W_{it}(.) + \beta EV_{it+1}(.)}}{1 + e^{\theta_{i0} + \gamma_i x_i + \alpha_i(Y_{it} - R_{it}) - W_{it}(.) + \beta EV_{it+1}(.)}} \quad (5)$$

The problem of using a simple logit model as in (5) is that it assumes that all randomness is captured by the *iid* extreme value errors. It rules out the existence of

[4]Other parametric distributions can also be adopted. For example, if $\bar{\varepsilon}$ are assumed to be standard normal, a standard probit model ensues.

unobserved variables that are correlated with the preferences of the individual[5]. In general, it is desirable to allow the error term to have a richer pattern of correlation across the several dimensions of the data than the simple *iid* errors in (5). We will accomplish that by decomposing the error into time- and individual-specific components that can be treated in the estimation as fixed or random effects, or that can be conditioned on observed states.

Notice that the described empirical model allows the identification of the structure of debtors' preferences up to its difference with the outside utility W_{it}. This outside payoff cannot be normalized because it presumably depends on the same variables that affect utility.

The estimation of the general model above involves two technical difficulties. First, it requires the computation of the continuation payoffs for every debtor along the estimation algorithm which is a nontrivial computational task with a sample that contains literally thousands of observations. Second, it requires matching data of all observable states at the micro level –specially it requires matching data on individual income over time, which is something that we don't have.

To circumvent the need to compute the dynamic problem along the estimation algorithm, the estimation below is based on a set of structural restrictions that turn (1) into a static problem. Otherwise, maximization of the sample likelihood function corresponding to (5) would require the numerical computation of the continuation payoffs EV_{it} for each individual mortgage, at every point in time, throughout the estimation algorithm.

Loosely speaking, if the relationship between the states and the continuation payoffs is stable across equilibria, then the estimates of a simple logit estimation are going to be "stable". This stability condition would hold, for example, if the distribution of the state variables is stationary so that the continuation payoffs are a constant.

[5]The same limitation is true for the case of the standard probit model.

On the other hand, non-matching survey data will be incorporated into the estimation by integrating the predicted individual default probabilities over the empirical distribution of income, conditional on the observed states. Specifically, the sample likelihood is going to be simulated from draws of income taken from survey data that contains matching information on income and housing values. The details of the simulation method are discussed below.

4 Estimation of a model of default with Colombian data (1997-2004)

4.1 The data

The model above is estimated with two separate non-matching panel data sets. The first (or "main") data set contains information on 16000 random mortgages that were outstanding between 1997 and 2002. The monthly payment history of each mortgage, its original and current value and term of the mortgaged home are included. On the other hand, the expected prices of individual homes at any point in time $\bar{P}_{it}$ are computed using home price indices constructed by the Colombian Central Bank following Escobar et al (2006). All data is aggregated into quarters, so that default observations are not confounded with missed payments or coding errors.

Since this main data set contains no information on the income of debtors over the span of the sample, survey data were collected with information on the joint distribution of households income and mortgage holdings (the "secondary" data set). Specifically, annual surveys conducted by DANE contain large samples of individual household incomes and matching housing payments that can be used to simulate the joint distribution of income and the other state variables.

Table 1 contains some summary statistics of the main data set, which goes from the second quarter of 1997 to the second quarter of 2004[6]. Notice that the number of loans in the data set changes over time as loans are paid off completely or new loans start; this number fluctuates roughly between 5000 and 8000. Columns (3) and (4) of the table contain the percentage of loans in the data at each point in time with more than 3 and 6 months of past due payments, illustrating the dramatic prevalence of default during the crisis. After 2000 until the end of the data set, more than 20% of all loans in the data set had past due payments of more than 3 months reaching 23% in the second quarter of 2003. The percentage of loans with past due payments of more than 6 months reaches its peak of more than 16% in the first quarter of 2003.

In the data it is observed that sometimes debtors temporarily stop making their payments. Therefore what 'default' means has to be defined. Specifically in the estimation below, loans that accumulate past due payments of more than 3 months are assumed to be defaulted and are dropped off from the data set. Therefore, 'default' is defined as the event in which the number of past due payments in a loan history changes from 3 or less to more than 3 between two quarters. After a loan is defined to be defaulted, it is dropped off the sample[7].

The default rate based on this definition (i.e. the number of 'defaults' over the total number of outstanding loans) is displayed in column (5) of the table. This rate reached its peak of more than 6% in the midst of the financial crisis, during the earlier quarters of 2000. Notice, though, that this rate is generally decreasing over the time span of the sample, due to the fact that defaulted loans are dropped from the sample. A dramatic reflection of the depth of the crisis in these years is the fact that by the end of the sample debtors had defaulted on more than 80% of the loans included in

[6]Since default is inferred from the change in the number of past due mortgage payments, the first observation in the first quarter of 1997 had to be dropped from the data set.

[7]The default rate based on this definition is highly correlated with default rates based on longer default periods. The 3-month threshold was chosen in order to observe as much default as possible and in order to capture *all* defaulted loans, including those that are terminated soon after default.

this random according to our definition of default.

There's no direct information on the size of the required monthly payments. It is known though that they were directly tied to the rate of mortgage balance over the remaining term of the loan. The balance of the loan was tied to the market interest rate through a formula established by the Central Bank. Column (6) of Table 1 contains the average *real* value of the ratio of mortgage balance to remaining term among outstanding loans. Notice that this value increases throughout the whole span of the sample. This increase might have been partly driven by the price of new homes, specially before 2000. Nevertheless, as can be seen in column (7) the real value of the homes in the sample is decreasing throughout the whole time span of the sample, in particular after 2000.

Table 2 characterizes the observed correlations contained in the data. Specifically, a linear probability model of non-default was estimated using the definition of default described above. Dependent variables include the mortgage balance, the expected price of the collateral and the remaining term of the loan at each point in time. As expected, default is positively correlated with the balance of mortgages at any point in time and with their remaining term. It is, on the other hand, negatively correlated with the expected price of the collateral[8].

It must remembered that the whole point of this paper is to isolate the *causal* effects of each variable on default behavior. The contention is that the described data sets contain sufficient *exogenous* variation so that the underlying relationships can be uncovered. The correlations described above may be just a reflection of unobserved states that are correlated with the included explanatory variables. In particular, the variation of the unobserved income of individual debtors may be driving the results of the regressions. The last two columns of Table 2 contain results of the model which

[8]Notice that the table reports the estimates of a regression of *non-default* on covariates, which is consistent with the specification of the structural model below.

includes fixed time-effects that capture the component of the unobserved states that is common to all debtors. Notice that the magnitude and statistical significance of the correlations don't change much after the inclusion of the fixed effects which suggests that the unobserved component of the error that is common to all debtors is not correlated with the observed variables included in the regression.

The estimates of the time effects, which are measured with respect to the constant in the second quarter of 1997, are mostly significant. The coefficient of correlation of these estimates and the average income of mortgage holders in the secondary data set is 0.41, which is consistent with the presumption that the time effects are capturing a lot of the common variation in household income.

The literature on mortgage default (e.g. Deng et al, 2003) has documented the fact that the initial loan to value (LTV) ratio of loans is correlated with the risk attitude of debtors who select themselves to different mortgage contracts. As seen on the left hand side columns of Table 2, there is a significant negative correlation between default and the initial LTV ratio, controlling for current home values and mortgage balances. As seen on the two right hand side columns of the table, this correlation becomes tenuous and insignificant when using fixed time controls. This implies that the sharp correlation detected in the first set of estimates is not strong within time periods. It is suggestive, though, of the importance of heterogeneity to explain observed default behavior.

Even though the results of these regression estimates are somewhat consistent with the described behavioral model, they are only descriptive. Nevertheless, the significant correlations described above are the basis for the econometric identification of the structural model below.

4.2 The empirical model of default

As indicated above, the data sets contain no information on the characteristics of the individual homes. Therefore, it will be assumed that the unobserved "quality" of homes x_i is random:

$$x_i \equiv \kappa + \varepsilon_{it}^x \tag{6}$$

where ε_{it}^x is a random error that is potentially correlated over time and across debtors.

There is no information on the required monthly payments R_{it} of each debtor. It is known, though, that payments are linear functions of mortgage balances K_{it} and remaining term L_{it}, with some random variation across debtors:

$$R_{it} = \rho_0 + \rho_1 K_{it} + \rho_2 L_{it} + \varepsilon_{it}^r \tag{7}$$

where ε_{it}^r is an error term.

It will also be assumed that the payoff of default $W_{it}(.)$ is a linear function of relevant states:

$$W_{it} = \omega_0 + \omega_1 Y_{it} + \omega_2 \bar{P}_{it} + \omega_3 K_{it} + \varepsilon_{it}^w \tag{8}$$

where ε_{it}^w is the structural error. Recall that this payoff is a linear combination of payoffs across random outcomes payoffs; if these payoffs are linear functions of states, then the linear payoff function $W_{it}(.)$ should be stable across counterfactual equilibria, as long as states don't affect the probabilities of individual outcomes. Notice that a careful interpretation of the function W_{it} is important because the usefulness of the model for counterfactual analysis relies on the assumption that this function will not change when we change the values or the transition probabilities of the state variables.

As explained above the full estimation of the model (1) requires the computation of the continuation payoffs for every debtor at every point in time along the estimation algorithm. The estimation in this paper is going to be based on the assumption that the expected continuation payoff is a linear function of current observed states. Formally:

$$\beta E\left[V_{it+1}(\bar{P}_{it+1}, K_{it+1}, Y_{it+1}, \varepsilon_{0it+1}, \varepsilon_{1it+1}) | \bar{P}_{it}, K_{it}, Y_{it}, \varepsilon_{0it}, \varepsilon_{1it}\right]$$

$$= \upsilon_0 + \upsilon_1 \bar{P}_{it} + \upsilon_2 Y_{it} + \upsilon_3 K_{it} + \varepsilon^{v}_{it} \tag{9}$$

which implies that the continuation payoff can be written in terms of the observed states is a linear function of the states, and that this linear relationship is stable. Admittedly, this assumption is very strong. It will nevertheless enable the estimation of the model using relatively standard techniques. Obtained results, in turn, can be easily related to the underlying general model.

Substituting (6), (7), (8) and (9) in condition (3), the non-default condition for debtor i at time t can be obtained:

$$\theta_0 + \gamma(\kappa + \varepsilon^{x}_{it}) + \alpha_i(Y_{it} - (\rho_0 + \rho_1 K_{it} + \rho_2 L_{it} + \varepsilon^{r}_{it})) + \varepsilon^{u}_{it}$$

$$+\upsilon_0 + \upsilon_1 \bar{P}_{it} + \upsilon_2 Y_{it} + \upsilon_3 K_{it} + \varepsilon^{v}_{it} \geq \omega_0 + \omega_1 Y_{it} + \omega_2 \bar{P}_{it} + \omega_3 K_{it} + \varepsilon^{w}_{it} \tag{10}$$

Grouping terms the condition above can be rewritten as:

$$\zeta_0 + \zeta_1 \bar{P}_{it} + \zeta_2 Y_{it} + \zeta_3 K_{it} + \zeta_4 L_{it} + \bar{\varepsilon}_{it} \geq 0 \tag{11}$$

Therefore the non-default probability will depend on the distribution of the error term $\bar{\varepsilon}_{it} \equiv \gamma\varepsilon^{x}_{i} - \alpha\varepsilon^{r}_{it} + \varepsilon^{u}_{it} + \varepsilon^{v}_{it} - \varepsilon^{w}_{it}$. In order to allow a rich correlation across choices we will consider models in which the error is decomposed as follows:

$$\bar{\varepsilon}_{it} = \xi_t + \mu_i + \epsilon_{it} \tag{12}$$

where the term μ_i is an individual-specific unobservable state and ϵ_{it} is an *iid* idiosyncratic disturbance. This specification allows individual choices to be correlated over time and across debtors; in addition, this unobserved heterogeneity can be allowed to depend on other observed states such as income which would be equivalent to a model with heterogenous ζ coefficients.

Assume that ϵ is distributed according to an extreme value distribution. Then, the individual non-default probability (5) is given by:

$$Prob(N_{it} = 1) = \frac{e^{\zeta_0 + \zeta_1 \bar{P}_{it} + \zeta_2 Y_{it} + \zeta_3 K_{it} + \zeta_4 L_{it} + \xi_t + \mu_i}}{1 + e^{\zeta_0 + \zeta_1 \bar{P}_{it} + \zeta_2 Y_{it} + \zeta_3 K_{it} + \zeta_4 L_{it} + \xi_t + \mu_i}} \tag{13}$$

Suppose first that $\mu_i = \xi_t = 0$. Again, $N_{it} = 1$ stands for the event of debtor i not defaulting on her mortgage at time t. Estimating the parameters ζ of the model above requires the maximization of the sample non-default likelihood predicted by the model. This likelihood is computed by multiplying the likelihood of observed choices across debtors and over time as follows:

$$L(\zeta) = \prod_{i \in S_t} \prod_{t \in T} (Prob(N_{it} = 1))^{N_{it}} (Prob(N_{it} = 1) - 1)^{1 - N_{it}} \tag{14}$$

where S_t is the random set of loans that is outstanding at time t.

The likelihood (14) cannot be computed directly due to the unavailability of matching income data. The non-matching income data from household surveys can be incorporated into the estimation above by integrating the likelihood over the empirical joint distribution of income and mortgage payments. Notice that the individual unobserved effects can also be incorporated into the estimation by assuming they come from a pre-specified parametric distribution and integrating them out throughout the estimation.

Specifically, if we assume that the individual effects μ_i are distributed according to some known parametric distribution $\Phi(\sigma_\mu)$, the "expected" non default probability is:

$$\hat{Prob}[N_{it} = 1] = \int \frac{e^{\zeta_0 + \zeta_1 \bar{P}_{it} + \zeta_2 Y + \zeta_3 K_{it} + \zeta_4 L_{it} + \xi_t + \mu}}{1 + e^{\zeta_0 + \zeta_1 \bar{P}_{it} + \zeta_2 Y + \zeta_3 K_{it} + \zeta_4 L_{it} + \xi_t + \mu}} dG_t(Y \mid K) d\Phi(\sigma_\mu) \tag{15}$$

where $\xi = \{\xi_{t=1 \ldots T}\}$ is treated as a vector of fixed time-effects that can be estimated for each t. $G(. \mid K)$ is the empirical distribution of household income at time t, conditional on mortgage balances, which can be inferred from the survey data.

Given any set of parameters $\{\zeta, \xi, \sigma_\mu\}$ the probabilities above can be obtained via simulation and the simulated sample likelihood can be computed just like in (14) above:

$$\hat{L}(\zeta, \xi, \sigma_\mu) = \prod_{i \in S_t} \prod_{t \in T} (\hat{Prob}(N_{it} = 1))^{N_{it}} (\hat{Prob}(N_{it} = 1) - 1)^{1 - N_{it}} \tag{16}$$

4.3 Estimation and results

To estimate the model, the simulated likelihood (16) was maximized by computing the predicted probabilities (15) using simulation. The first issue to be addressed is the specification of the unobserved debtor heterogeneity. Following the previous empirical literature on mortgage default (e.g. Deng et al, 2003), debtor heterogeneity will be tied to the initial "loan to value" $LTV_i = K_{i0}/P_{i0}$ of the mortgage, where $t = 0$ stands for the moment at which the loan was first started. This specification presumes that debtors select themselves to mortgages with different LTV according to their attitude towards risk.

Accordingly, it is assumed that $\bar{\varepsilon}_{it} = \xi_t + \sigma_{mu} LTV_i \bar{\mu}_i + \epsilon_{it}$, so that the unobserved component of utility has a common element ξ_t that varies over time, a consumer-specific component $\sigma_{mu} LTV_i \mu_i$ and an extreme value consumer- and time-specific shock ϵ_{it}.

The consumer-specific shock $\sigma_{mu} LTV_i \bar{\mu}_i$ is assumed to be correlated with the initial leverage of the mortgage, so that its distribution can be separated from the distribution of the idiosyncratic shock. Specifically, $\bar{\mu}_i$ is assumed to be a standard normal error, so that the consumer-specific error is normal with zero mean and variance $\sigma_{mu}^2 LTV_i^2$. Higher absolute realizations of this unobserved error are associated with a higher initial LTV and are a consumer specific constant that shifts the individual utility function.

On the other hand, the common component of the error $\xi_{t=1,\ldots,T}$ was estimated as a fixed time effect. Therefore, for any set of parameters $\{\zeta, \xi, \sigma_\mu\}$, a consistent estimator of such integral is given by:

$$P\hat{r}ob[N_{it} = 1] = \frac{1}{J} \sum_{j=1}^{J} \frac{e^{\zeta_0 + \zeta_K K_{it} + \zeta_P \bar{P}_{it} + \zeta_L L_{it} + \zeta_Y Y_j + \xi_t + \sigma_\mu \bar{P}_{it} \bar{\mu}_i}}{1 + e^{\zeta_0 + \zeta_K K_{it} + \zeta_P \bar{P}_{it} + \zeta_L L_{it} + \zeta_Y Y_j + \xi_t + \sigma_\mu LTV_i \bar{\mu}_i}} \tag{17}$$

where μ_j are independent standard normal draws and Y_j are income draws taken from the empirical distribution of income, conditioned on housing payments, contained in yearly surveys. The average is taken over J simulations.

Specifically, the survey data contains random observations of households' income and mortgage payments of homeowners[9], while the main data set contains information on the balances and maturities of outstanding mortgages. It is assumed that the distribution of monthly mortgage payments is the same as the distribution of balances over the remaining maturity of mortgages K_{it}/L_{it}. Therefore, the quantiles of the distribution of K_{it}/L_{it} in the main data set correspond to the quantiles of the distribution of income for the households that are making mortgage payments. In computing (11), draws of $\{\bar{P}_{it}, K_{it}, L_{it}\}$ are therefore matched with random draws of Y_{it} from the same conditional distribution quantile. Given that surveys with mortgage payment data are only available at the yearly level, the distribution of income conditional on mortgage payments is interpolated to remaining quarters, by assuming that the income distribution was constant within years.

Four versions of the model were estimated with results reported in Table 3. In model 1 it is assumed that $\mu = \xi = 0$; in model 2 $\mu = 0$ and in model 3 $\xi = 0$. Model 4 is the full model as in (17). The displayed results of models 1 and 3 were obtained simulating 20 income draws for each observation from the corresponding income quantile in the secondary data set. Due to the size of the involved matrices, models 2 and 4 were estimated from 10 income draws per observation; in addition, a random subsample of 1/4 of the simulated sample was taken to alleviate computer memory restrictions. To give an idea of the computational magnitude of the estimation, the size of the matrix of regressors after subsampling in model 4 was (572600x34). The reported standard errors were obtained using the standard formula and are very robust to alternative specifications.

Given the rich variation of the data documented above, it is not surprising that the estimates are highly significant. An exception is the income coefficient in models 2 and 4, whose high standard errors are presumably due to the lower number of in-

[9]More precisely, the surveys ask whether people are financing the home they live in and how much do they pay.

come draws used in its estimation. Consistently with the correlations reported above, higher balances and longer remaining terms induce lower non-default probabilities. On the other hand, higher expected home prices induce higher non-default probabilities. Similarly, higher income is associated with higher non-default, which is not surprising.

Table 3 also reports the computed average marginal effects of each variable on the non-default probabilities. As can be seen, these average effects are all comparable to the reported results from the linear probability model and mostly similar. One likely exception is the marginal effect of the expected home price which is consistently higher in the structural estimation than in the linear model, the magnitude of the difference being both statistically and economically significant.

Given the environment faced by debtors during the time span of the sample, the salient feature of the results is the estimate of the income coefficient. The statistical significance of the coefficient is not surprising given the sharp correlation of income and default reported above. What is somewhat surprising is its very low economic significance. As indicated in the table, in average a marginal increase of COL$10 million a month in 1998 which is well above the mean of COL$0.6 million, induces an increase in the non-default probability of 0.2%. On the other hand, an decrease in the price of home of COL$10 million in 1998, which is less than the average loss of housing values in the sample between 1998 and 1999, induces an increase in the default probability of around 0.2%[10]. This magnitude is not insignificant, given the magnitude of the default rates, which is between 1% and 6%.

Notice that this results are robust to specifications that control for the random heterogeneity that might be correlated with the observable covariates. Specifically, it might be argued that income, balances and housing prices might be correlated with an unobservable variable that hit the default probabilities of *all* debtors. It is difficult ascertain what such a variable can be, but it might be the case that once default rates

[10]Remember that the table displays the results of the model of non-default.

increased, debtors anticipated that renegotiating the terms of the mortgage was easier. On the other hand, it might also be argued that the observable variables might be correlated with unobserved individual-level heterogeneity due to endogenous selection of debtors into specific mortgage terms.

Nevertheless, the results of the estimation are robust to the inclusion of fixed time-effects in models 2 and 4 and to the inclusion in models 3 and 4 of a normal unobserved heterogeneity that is correlated with the initial LTV and, at least indirectly, with current realizations of K_i and P_i . As indicated above, the addition in models 3 and 4 of the normal error that is correlated with LTV_i aims to capture the fact that even conditional on the unobservables, default rates might vary across consumer types who select themselves into homes with different prices.

Notice that the coefficient of the individual-specific error is significant and negative, which means that default rates of debtors are negatively correlated across initial LTV values due to underlying heterogeneity. The estimates of the fixed time-effects, which account for unobserved aggregate shocks that are not correlated with observed variables are large. As seen in Figure 1, this time varying constant which is measure with respect to the constant at the initial period exhibits a large correlation with the aggregate default rates.

It has to be said that this time effects are quite large. In other words, the observed variables in the model cannot account for a big portion of the time variation of default behavior. This is not surprising given that most of the variation on which the identification of the model is based is cross sectional and dynamic effects have been ignored. The results are nevertheless reassuring of the econometric validity of the obtained estimates, in the sense that the the observed states variables don't appear to be correlated with the unobservables.

The results above strongly suggest that, conditional on the specified behavioral model, household income variation was not the driving force behind the dramatic increases in mortgage defaults during the late 1990's. The variation in the price of the collateral and the increases in the size of the mortgage balances seemed to have played a more important role. In the following section a more precise evaluation of these impacts is performed. The role of other unobserved factors is also discussed.

4.4 Fit of the model and additional results

As seen in figure 1, the model can trace satisfactorily the aggregate default rates. Aggregated over time, the default rate reached 86%, which is a dramatic figure; it means that 86% of the household in this random sample accumulated past due mortgage payments for more than 3 months[11]. The difference between this observed overall default rate and the rate predicted by the model is of less than 0.5%. As indicated above, though, much of the time variation of default behavior is driven by an unobserved aggregate shock, equivalent to a time-changing model constant (more on this below). In this sense, the model is better suited to understand the variation in default across debtors at any point in time.

In order to isolate more clearly the impact of individual factors on default probabilities, the model can be used to compute them directly. Table 4 contains default probabilities for an "average" debtor with different values of the observed states, keeping the other values at the average value they had at the beginning of the sample[12]. The probabilities are computed for values of L, P and K that lie at the center of the quintile of their distribution.

Notice that predicted default probabilities for this "average" debtor are very sen-

[11]If instead of defining default as the accumulation of 3 or more months of past due payments we had used a threshold of 6 months, the accumulated default rate over the time span of the sample would have been 49%, which is still a staggering figure.

[12]Specifically, default probabilities were computed for $\bar{K} = 1.614$, $\bar{P} = 4.493$, $\bar{L} = 42$ and $\bar{Y} = 0.0865$ which were the average values of these states in the 3rd quarter of 1997; recall that K, P and Y are measured in tens of millions of constant 1997 Colombian pesos.

sitive to changes in the observed states. For example, at any point in time increasing the number of remaining periods of mortgage maturity, L, from the center of the lower quintile of the distribution of L ($L = 27$) to the center of the upper quintile of the distribution of L ($L = 57$) increases the predicted default probability by around 50%. Keeping everything else constant, increasing the balance of the mortgage K from the center of the lower quintile ($K = 0.36$)to the center of the upper quintile of its distribution ($K = 3.8$) increases the predicted default probabilities at any point in time by more than 60%. The effect of price is just as significant: increasing the price from the lower quintile ($P = 1.3$) to the upper quintile ($P = 10.8$) of its distribution increases the predicted default probability by more than 50%.

The model also allows the characterization of default probabilities across individuals according to their unobserved income. This is important, because income is the most important individual random state that affects default behavior and that is not directly observed by banks or policy makers. Figure 6 illustrates the percent difference in the default rates between debtors in the upper 20% and lower 20% of the income distribution at each point in time. For comparison purposes, figure 7 also displays the percent difference in the default rates of households with home prices in the upper and lower 20% distribution of home prices at each point in time. Figure 7 also also display the percent difference in default rates between loans with remaining maturities in the upper and lower tails of the distribution of mortgage age.

In figure 6 it can be seen that, perhaps surprisingly, the predicted default rates of wealthier households are consistently higher than the predicted default rates of poorer households, despite the fact that income has a negative effect on the probability of default. This difference is almost 15% at the beginning of the sample period and tends to disappear over time as the pool of debtors shrinks. In fact, the predicted aggregate rate of default is 90% for debtors in the upper tail of the income distribution and 84% for debtors in the lower tail. For debtors located around the median of the distribution this rate was 86%.

This effect is not an artifact of the wealthier households having more expensive homes, which itself induces higher default rates; as seen in figure 7, the difference between the average predicted default rates of households in the upper and lower tails of the distribution of housing prices is around zero at the beginning of the sample period and negative in the latter periods, which means that, if anything, default rates were higher for debtors with relatively low-priced collateral. In fact, the accumulated default rate is 81% for homes in the lower tail of the price distribution, whereas it is 86% in the upper tail. For debtors located around the median this aggregate default rate was 88%.

In figure 7 it can be seen that default rates are consistently higher for loans with longer remaining maturities. This is of course a direct implication of the fact that it is easier to default on young mortgages that have small accumulated equity. But then the positive correlation of income and default has to be a result of an underlying concentration of mortgages with long remaining maturities in the hands of relatively wealthier households. This, in turn, is a reflection of the credit boom that preceded the time span of the sample.

This section finishes with a discussion about the variation over time of the predicted default probabilities. As indicated above, the model is not very well suited for predicting the variation of default probabilities over time, as much of it is explained by an unobserved aggregate shock which was estimated taking advantage of the panel structure of the data. The estimation results implied that this "error" is not correlated with the observed states. It is difficult to argue that these shocks are random, but it is also difficult to infer from the data what drives their evolution.

The model doesn't explicitly include aggregate variables, because it would be difficult to establish any meaningful causal connection to the observed default behavior. For example, figure 8 displays the fixed time-effects estimated in Models 2 and 4 and the average 90-days deposit interest rate, which is regarded as a good measure of the opportunity cost of liquid assets in Colombia. As can be seen, the time-effects and the interest rate seem to be somewhat negatively correlated: their correlation coefficient is around -0.1. Inferring a causal relationship from this correlation is not possible, due to the fact that the interest rate is presumably correlated with other unobserved aggregate variables that drive the variation of default over time. Notice, though, that the estimated time-effects isolate the effects of aggregate variables on default. Therefore, they can be used to construct a statistical model relating aggregate shocks to default, which is beyond the scope of this study.

4.5 Summary of the estimation algorithm

The estimation of the model above is based on the computation of the simulated likelihood of the sample across observations:

- Organize the income data from highest to lowest housing payments using the surveys that contain both. Separate observations into quantiles; this joint distribution of income and housing payments is assumed to be equivalent to the joint distribution of income and K_{it}/L_{it}. Organize the observations of $\{\bar{P}_{it}, K_{it}, L_{it}, LTV_i\}$ from highest to lowest K_{it}/L_{it} and separate it into quantiles.

- For each loan in the sample generate a number J of standard normal draws ε_i that are constant over time. Match these draws with J random draws with replacement from the corresponding quantile of the distribution of Y_{jt}, conditional on the mortgage payments. Keep these draws constant throughout the estimation.

- Set the vector of parameters $\{\zeta_0, \sigma_{\mu 0}\}$. Compute $\hat{L}_0$ using (16) and (17) using the "simulated" sample described above.

- Look numerically for the set of parameters $\{\zeta^*, \sigma_\mu^*\}$ that maximize the likelihood of the sample. Compute the standard errors of the estimates using the usual methods.

5 Concluding remarks and further research

This paper has developed an empirical model of mortgage default, whose estimates can be related to an underlying behavioral model. Methodologically, the paper illustrated the kind of assumptions needed to obtain a model that can be estimated using variations of a simple logit model. Still, the model improved upon the standard logit/probit framework in two ways: First, the model is estimated using non-matching panels of income and mortgage payment data. Second, the model allows for the presence of persistent heterogeneity across consumers.

The resulting technique is similar to a standard discrete choice model, except that the non-matching income data and the unobserved heterogeneity are incorporated into the model using simulation techniques. The main result of the paper is that income variation had a very small effect on the default probabilities, compared with the effect of housing prices and mortgage balances (which were tied to a market interest rate), which is consistent with a model of rational default behavior.

It is also found persistent debtor heterogeneity is found to be statistically significant. Moreover, it is found that at any point in time and given the joint distribution of states, default rates were higher for higher income households. The estimation also allows the estimation of an aggregate time-varying common shock that seems to be the driving force of the variation of default over time.

This paper leaves two open avenues for continuing research. First, the estimated aggregate shock can be used to construct a model that ties it to the evolution of the macroeconomic environment. Second, the estimation of the model relied on a simplified treatment of dynamics. Due to the size of the sample, estimating a fully dynamic model is complicated. It would require the computation of consumer-specific continuation payoffs along the estimation algorithm, which is computationally difficult.

References

Cuellar, M. M. (2006): “A la vivienda quién la ronda?. Situación y Perspectivas de la política de vivienda en Colombia”, *ICAV y Universidad Externado de Colombia*.

Deng, Y.,J. Quigley and R. Van Order (2000). “Mortgage Terminations, Heterogeneity and the Exercise of Mortgage Options”, *Econometrica*, 68(2), 275-307

Escobar J., C.A. Huertas, D.A. Mora and J.V. Romero (2006): “Indice de precios de la vivienda usada en Colombia - IPVU - Metodo de ventas repetidas” *Borradores Semanales de Economia 368* Banco de la Republica, Bogota.

Rust, J. (1987): “Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher”, *Econometrica*, 55(5), 999-1033.

Table 1: Summary statistics (main data set)

(1) Quarter	(2) Number of loans	(3) Past due 3 months	(4) Past due 6 months	(5) Default rate	(6) Balance/ term	(7) Average price
1997 : 2	4965	6.4 %	1.6 %	4.0 %	370770	72482000
1997 : 3	4958	7.1 %	1.8 %	3.2 %	387160	73308000
1997 : 4	5101	8.1 %	2.4 %	3.0 %	399720	75518000
1998 : 1	7197	8.2 %	2.8 %	3.4 %	435920	64634000
1998 : 2	7365	7.9 %	3.1 %	2.3 %	447770	61548000
1998 : 3	7502	8.8 %	3.8 %	1.9 %	469250	58331000
1998 : 4	7569	10.6 %	4.6 %	2.8 %	493660	56400000
1999 : 1	7482	14.1 %	6.3 %	4.1 %	523170	57867000
1999 : 2	7809	16.3 %	7.8 %	4.7 %	504910	53608000
1999 : 3	8060	11.8 %	6.3 %	3.7 %	483630	47878000
1999 : 4	7827	19.0 %	8.3 %	6.3 %	495370	50559000
2000 : 1	8594	18.1 %	10.6 %	6.4 %	503100	47984000
2000 : 2	8020	16.1 %	9.3 %	5.3 %	478060	49014000
2000 : 3	7505	19.0 %	9.0 %	5.5 %	479880	48540000
2000 : 4	7053	19.5 %	10.6 %	3.5 %	481980	46981000

Continues in next page

Prices and balances are in 1997 COL$

Table 1, continued

(1)	(2)	(3)	(4)	(5)	(6)	(7)
2001 : 1	6786	20.4 %	12.1 %	2.6 %	488410	46750000
2001 : 2	6601	22.1 %	13.8 %	2.7 %	512340	38483000
2001 : 3	6416	22.9 %	14.7 %	2.6 %	520730	40771000
2001 : 4	6253	22.8 %	15.1 %	2.2 %	525090	36298000
2002 : 1	6140	22.2 %	15.3 %	1.8 %	528920	32062000
2002 : 2	6060	22.0 %	15.6 %	1.6 %	541360	34959000
2002 : 3	6028	23.4 %	16.5 %	2.5 %	553380	33041000
2002 : 4	5891	22.6 %	15.9 %	1.7 %	554660	36579000
2003 : 1	5862	23.0 %	16.6 %	1.6 %	563210	32067000
2003 : 2	5816	23.2 %	16.4 %	1.8 %	581450	32043000
2003 : 3	5580	22.6 %	16.4 %	1.3 %	584720	31138000
2003 : 4	5666	23.3 %	16.9 %	1.8 %	576500	31256000
2004 : 1	5553	22.7 %	16.8 %	1.2 %	571580	29534000
2004 : 2	5450	22.0 %	16.4 %	1.0 %	582490	31386000

Prices and balances are in 1997 COL$

Table 2: Linear Probability Regressions

Variable	Est.	t-stat	Est	t-stat
Constant	0.0239	14.8909	0.0271	7.8372
Balance	0.0059	12.1878	0.0053	10.7226
Price	-0.0016	-10.0016	-0.0013	-8.3808
Term	0.0004	10.8136	0.0003	5.6982
LTV	-0.0075	-3.2854	-0.0029	-1.2497
Time-effects				
1997 : 3			-0.0089	-2.1328
1997 : 4			-0.0037	-2.5864
1998 : 1			-0.0142	-1.1562
1998 : 2			-0.0182	-4.4177
1998 : 3			-0.0097	-5.6867
1998 : 4			0.0044	-3.0346
1999 : 1			0.0109	1.3566
1999 : 2			0.0011	3.3789
1999 : 3			0.0271	0.3345
1999 : 4			0.0293	8.3613

Continues in next page

Time-effects, continued		
2000 : 1	0.0197	9.1465
2000 : 2	0.0216	6.0505
2000 : 3	0.0028	6.5592
2000 : 4	-0.0061	0.8397
2001 : 1	-0.0051	-1.8137
2001 : 2	-0.0058	-1.4952
2001 : 3	-0.0091	-1.6932
2001 : 4	-0.013	-2.6326
2002 : 1	-0.0151	-3.7193
2002 : 2	-0.0057	-4.3098
2002 : 3	-0.0134	-1.6115
2002 : 4	-0.0151	-3.8027
2003 : 1	-0.0125	-4.254
2003 : 2	-0.018	-3.5268
2003 : 3	-0.0123	-5.0115
2003 : 4	-0.0185	-3.4212
2004 : 1	-0.0196	-5.1382
2004 : 2	0.0194	-5.4116

Table 3: Structural Estimation Results

Coefficient	Model 1	Marginal effects	Model 2	Marginal effects
Constant	3.7933 (0.0081)	0.1144	3.6990 (0.0630)	0.1079
Balance	-0.1776 (0.0031)	-0.0054	-0.1693 (0.0089)	-0.0049
Price	0.0589 (0.0013)	0.0018	0.0492 (0.0037)	0.0014
Term	-0.0128 (0.0003)	-0.0004	-0.0127 (0.0010)	-0.0004
Income	0.0847 (0.0252)	0.0026	0.0859 (0.0723)	0.0025
σ_μ	0		0	
Coefficient	**Model 3**	**Marginal effects**	**Model 4**	**Marginal effects**
Constant	3.7911 (0.0081)	0.1144	3.0973 (0.0569)	0.1011
Balance	-0.1738 (0.0031)	-0.0052	-0.1038 (0.0091)	-0.0034
Price	0.0579 (0.0013)	0.0017	0.0645 (0.0034)	0.0021
Term	-0.0129 (0.0003)	-0.0004	-0.0007 (0.0009)	-0.00002
Income	0.0825 (0.0251)	0.0025	0.0453 (0.0631)	0.0015
σ_μ	-0.0169 (0.0051)		-0.264 (0.0128)	

Models 2 and 4 contain time-effects (not shown).

Table 4: Predicted Default Probabilities (selected quarters; evaluated at mean values as of 1997:3)

Variable	1997:3	1998:3	1999:3	2000:3	2001:3	2002:3	2003:3
L							
27	0.0435	0.0331	0.0665	0.0586	0.0106	0.0023	0.0121
40	0.0504	0.0384	0.0769	0.0678	0.0123	0.0026	0.0141
45	0.0533	0.0407	0.0812	0.0717	0.0131	0.0028	0.0149
51	0.0571	0.0436	0.0868	0.0766	0.0140	0.0030	0.0160
57	0.0611	0.0467	0.0926	0.0818	0.0151	0.0032	0.0172
K							
0.3644	0.0439	0.0305	0.0588	0.0509	0.0089	0.0019	0.0099
0.7956	0.0468	0.0325	0.0626	0.0542	0.0095	0.0020	0.0106
1.1973	0.0496	0.0344	0.0663	0.0574	0.0101	0.0021	0.0112
1.7831	0.0539	0.0375	0.0720	0.0624	0.0110	0.0023	0.0122
3.8742	0.0727	0.0509	0.0964	0.0838	0.0151	0.0032	0.0168
P							
1.3044	0.0639	0.0433	0.0778	0.0652	0.0111	0.0023	0.0121
2.1439	0.0607	0.0411	0.0740	0.0620	0.0105	0.0022	0.0114
3.1892	0.0570	0.0385	0.0694	0.0581	0.0098	0.0021	0.0107
4.9138	0.0512	0.0346	0.0625	0.0523	0.0088	0.0019	0.0096
10.8746	0.0353	0.0237	0.0432	0.0360	0.0060	0.0013	0.0065

The probabilities were evaluated at $\bar{K} = 1.614$, $\bar{P} = 4.493$, $\bar{L} = 42$ and $\bar{Y} = 0.0865$; prices, balances and income are measured in tens of millions of Colombian pesos.

Figure 1: Yearly rate of growth of GDP (1996−2003)

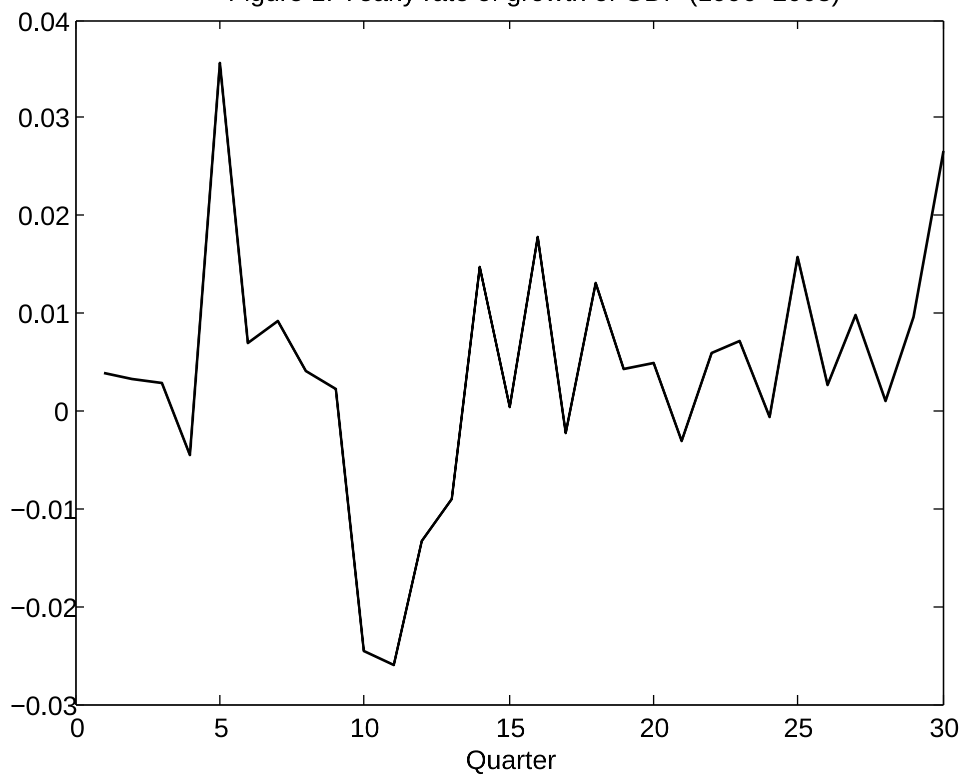


Figure 2: Annualized overnight interest rate (1996−2003)

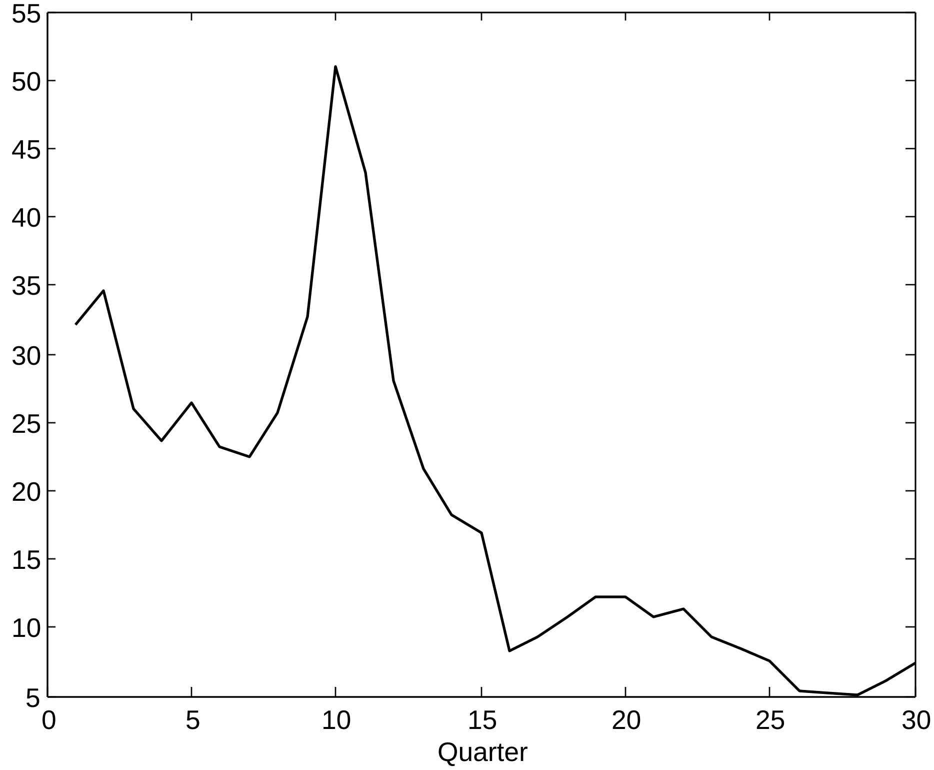


Figure 3: Home price index (1996−2003)

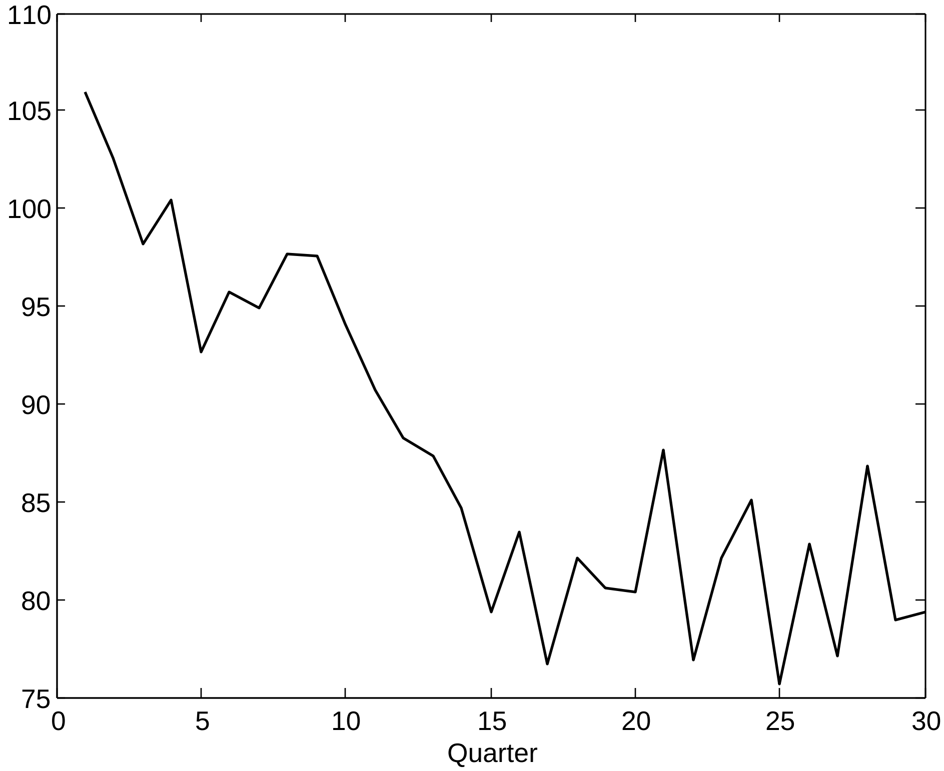


Figure 4: Exchange rate COL$/US$ (1996−2003)

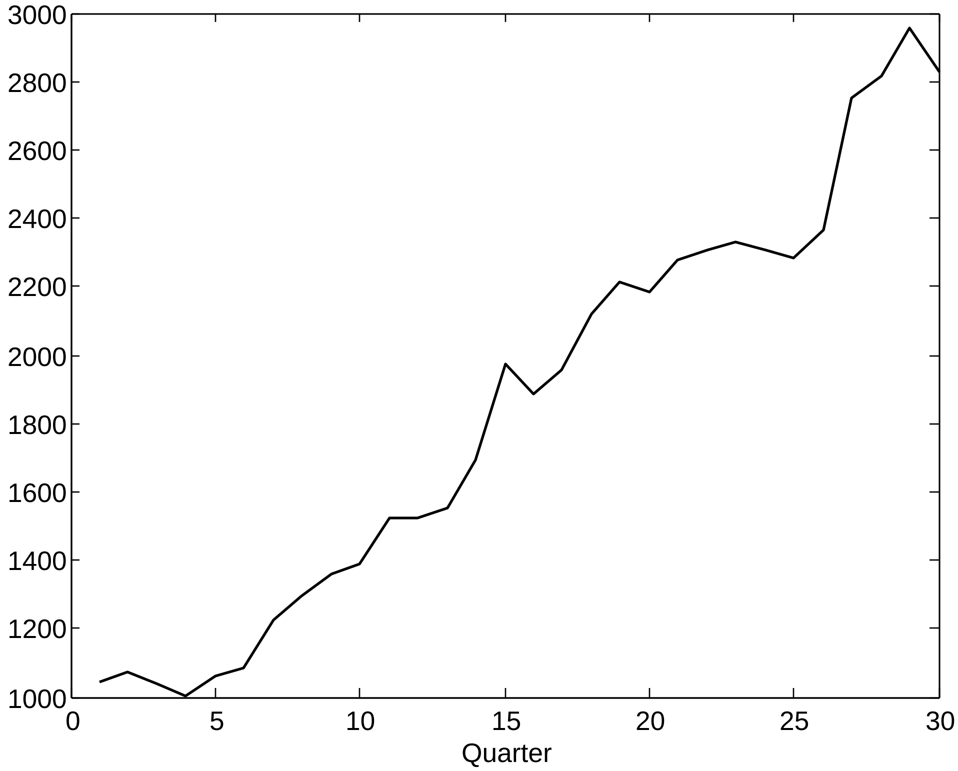


Figure 5: Observed and predicted default (Model 4)

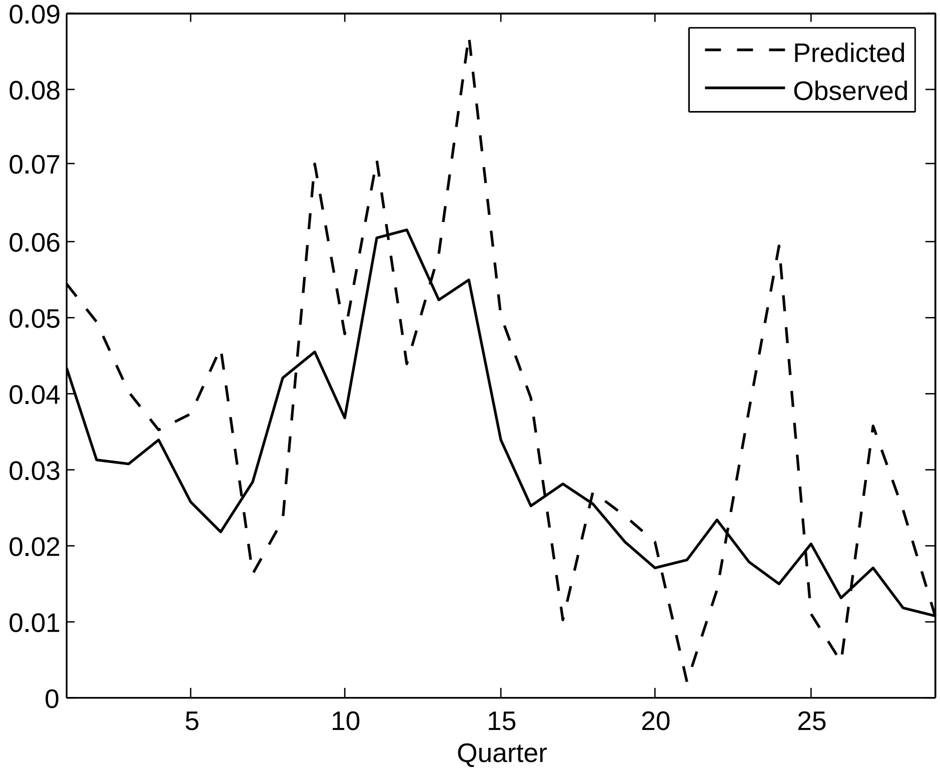


Figure 6: Percent differences in default rates across income levels
(top 20% minus bottom 20%)

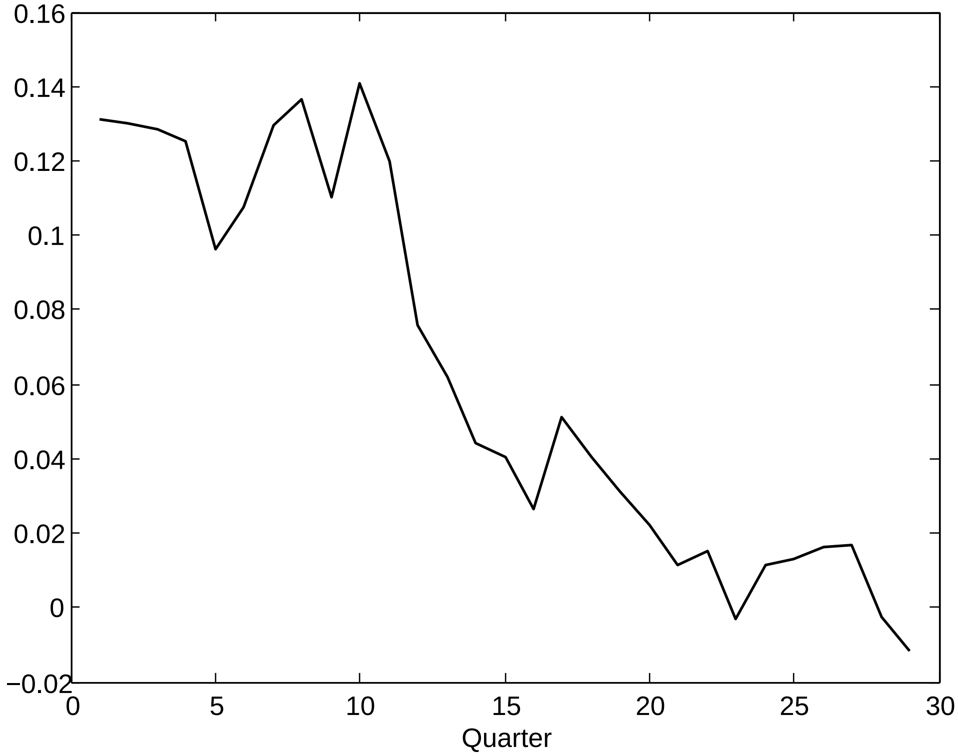


Figure 7: Differences in default rates
across house prices and mortgage maturities
(top 20% minus bottom 20%)

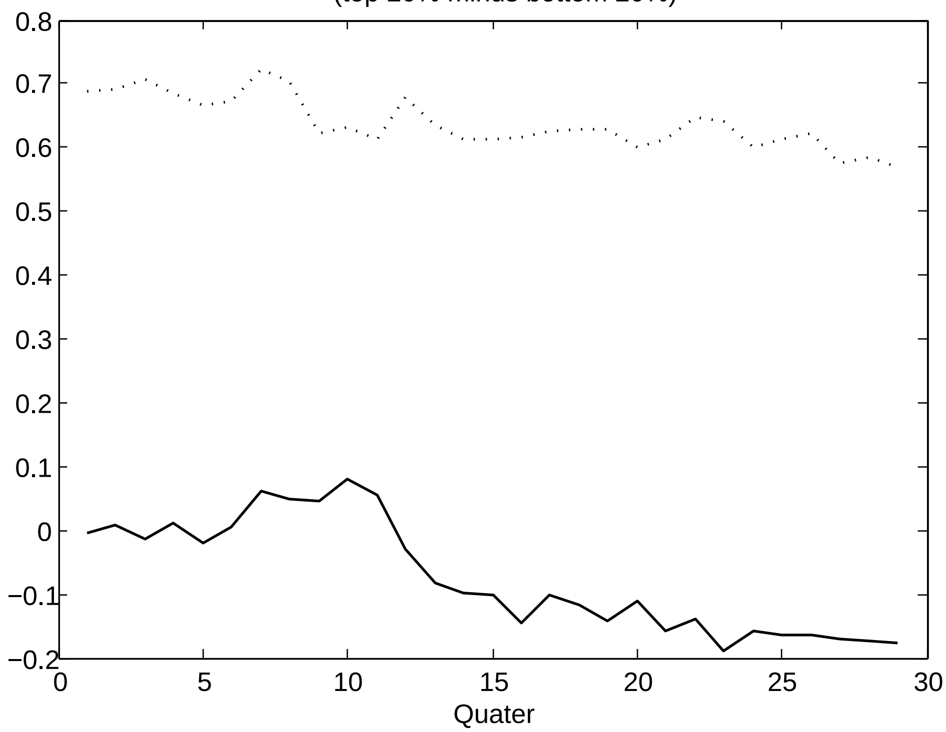


Figure 8: Real interest rate and fixed time effects (from model 4)