NBER WORKING PAPER SERIES

AN EMPIRICAL MODEL OF SUBPRIME MORTGAGE DEFAULT FROM 2000 TO 2007

Patrick Bajari
Chenghuan Sean Chu
Minjung Park

Working Paper 14625
http://www.nber.org/papers/w14625

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2008

We thank Narayana Kocherlakota, Andreas Lehnert, Monika Piazzesi, Tom Sargent, and Dick Todd for helpful conversations. Bajari would like to thank the National Science Foundation for generous research support. Thanks also go to Sean Flynn for helpful research assistance. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

An Empirical Model of Subprime Mortgage Default From 2000 to 2007
Patrick Bajari, Chenghuan Sean Chu, and Minjung Park
NBER Working Paper No. 14625
December 2008
JEL No. G01,G18,G2,G33,R51

ABSTRACT

The turmoil that started with increased defaults in the subprime mortgage market has generated instability in the financial system around the world. To better understand the root causes of this financial instability, we quantify the relative importance of various drivers behind subprime borrowers' decision to default. In our econometric model, we allow borrowers to default either because doing so increases their lifetime wealth or because of short-term budget constraints, treating the decision as the outcome of a bivariate probit model with partial observability. We estimate our model using detailed loan-level data from LoanPerformance and the Case-Shiller home price index. According to our results, one main driver of default is the nationwide decrease in home prices. The decline in home prices caused many borrowers' outstanding mortgage liability to exceed their home value, and for these borrowers default can increase their wealth. Another important driver is deteriorating loan quality: The increase of borrowers with poor credit and high payment to income ratios elevates default rates in the subprime market. We discuss policy implications of our results. Our findings point to flaws in the securitization process that led to the current wave of defaults. Also, we use our model to evaluate alternative policies aimed at reducing the rate of default.

Patrick Bajari
Professor of Economics
University of Minnesota
4-101 Hanson Hall
1925 4th Street South
Minneapolis, MN 55455
and NBER
bajari@econ.umn.edu

Chenghuan Sean Chu
Federal Reserve Board of Governors
20th Street and Constitution Avenue NW
Washington, DC 20551
sean.chu@frb.gov

Minjung Park
Department of Economics
University of Minnesota
4-101 Hanson Hall
1925 Fourth Street South
Minneapolis, MN 55455
mpark@umn.edu

1 Introduction

Subprime mortgages are made to borrowers who have a higher probability of default due to low credit quality or risk factors associated with the loan, such as a small downpayment. The subprime market experienced substantial growth starting in the mid- to late 1990s. The percentage of all mortgages that were subprime grew from less than 5% in 1994 to 20% in 2005.[1] Much of this growth was made possible by an expansion in the market for private-issue mortgage-backed securities (MBS). Securitization through MBS and related credit derivatives made it less costly to originate and fund loans that did not conform to the underwriting standards of the government-sponsored enterprises (GSEs), Fannie Mae and Freddie Mac, which are the chief securitizers of prime mortgages.

Beginning in late 2006, the United States subprime mortgage market experienced a sharp increase in delinquencies and foreclosures. In the third quarter of 2005, 10.76% of all subprime mortgages were delinquent and 3.31% were in the process of foreclosure. By comparison, the corresponding figures were 18.67% and 11.81% in the second quarter of 2008. The turmoil in the housing market has also generated broader instability in financial markets. Because securitization transfers ownership of the stream of mortgage payments from the originator to noteholders—chiefly other financial institutions—the capital structures of these other institutions became seriously impaired when the unexpected spike in default rates caused the value of MBS to plunge. Thus, not only have subprime lenders such as New Century Financial Corporation been forced to declare bankruptcy, but also commercial- and investment banks have experienced substantial losses from write-downs on the value of MBS and collateralized debt obligations. A further consequence has been the collapse of major institutions including Bear Stearns and Lehman Brothers. The resulting reduction in economywide lending is linked to what many forecast could be the worst recession since the Great Depression.

Policymakers have initiated a number of responses to the rise in defaults and worsening conditions in credit markets. The United States government has earmarked $700 billion to fund capital injections into financial institutions, instituted a credit facility to swap MBS for treasury securities, and placed the previously independently operating Fannie Mae and Freddie Mac under conservatorship. The Federal Reserve Board has announced a $600 billion program to purchase the direct debt of Fannie Mae and Freddie Mac as well as MBS issued by the two corporations, with the goal of lowering mortgage rates and increasing the availability of credit for housing purchases. The Federal Deposit Insurance Corporation

[1] Source: Moody's Economy.com.

(FDIC) has also advocated modifying mortgages to reduce monthly payments to no more than 31 percent of borrowers' monthly pretax income as a way to mitigate foreclosures. In addition, the banking industry itself has led efforts to stem foreclosures by modifying loan terms to make payments more affordable. Because problems in the housing market were at the origin of this cascade of events, identifying the underlying causes behind the recent increase in mortgage defaults is key to formulating appropriate policy.

Financial innovations leading to the development of the subprime MBS market have been subject to two chief criticisms. The first objection is that existing models used by the financial industry to price subprime MBS have been too optimistic and have placed insufficient weight on sources of systematic (nondiversifiable) risk. Although bundling individual mortgages certainly reduces idiosyncratic risk, the pools are not immune to aggregate shocks such as nationwide declines in home prices. Through our unified econometric framework, we analyze how subprime borrowers' default decisions respond to home price declines, thus providing a key input into a more accurate pricing model for securitized debt, which in turn is necessary for capital markets to function properly.

A second concern is that the MBS market is plagued by adverse selection and agency problems. Originators are unaffected by the ex post outcomes of bad mortgages that they have sold off but generate income by offloading them. As a result, securitization gives lenders a stronger incentive to issue risky loans, to the extent that certain markers of risk are unobserved to other market participants. Thus, financial innovation may in equilibrium lead to lower lending standards, causing the composition of borrowers receiving loans to shift over time toward riskier types. Understanding the drivers of default based on commonly observed characteristics is an initial step toward determining the magnitude of such agency problems, and our analysis allows us to quantify the impact of changes in borrower composition on default rates.

In this paper, we explore four potential explanations for the increase in mortgage defaults. Our analysis uses a unique data set from LoanPerformance that tracks the universe of subprime and Alt-A mortgages that were securitized between 2000 and 2007. The unit of observation is an individual mortgage observed at a point in time. At the loan level, we observe information from the borrower's loan application, including the term of the loan, the initial interest rate, interest rate adjustments, the level of documentation, the appraised value of the property, the loan-to-value ratio, and the borrower's FICO score at the time of origination. We also observe the month-by-month stream of payments made by the borrower as well as whether the mortgage goes into default. We merge the LoanPerformance data with the Case-Shiller home price indices in 20 major U.S. cities. The merge allows us to track the current value of a home, by inflating the original appraisal value by the applicable disaggregated price index.

One potential explanation for the rise in defaults is falling home prices. Consider a frictionless world in which there are no transaction costs from selling a home and no penalties for defaulting on a mortgage (including any limits on the household's ability to immediately buy back the same house or a similar one). If the current market value of the home is less than the outstanding mortgage balance, it is optimal for the borrower to default. In the literature, the option to default is referred to as the put-option component of the mortgage (see Crawford and Rosenblatt, 1995; Deng, Quigley, and van Order, 2000; Foster and van Order, 1985; Quigley and van Order, 1995; Vandell, 1993).

A second explanation is changes in expectations about home prices. In a world in which agents have dynamic incentives, expectations about home price appreciation affect the value of keeping a mortgage alive, and therefore influence the default decision. When home prices are expected to appreciate rapidly, borrowers have a reduced incentive to default, because default would entail forgoing the capital gains from the increased value of the home.

A third potential explanation attributes the observed rise in defaults to increases in contract interest rates relative to market rates, particularly for adjustable-rate mortgages (ARMs). When the contract interest rate is less than the current market rate, the incentive to default is lower because a borrower who defaults would lose access to the discounted interest rate. Conversely, when the contract interest rate rises relative to the market rate, the incentive to default increases.

In addition to these financial incentives for default, increased defaults may also be due to short-term liquidity constraints on households. Borrowers who select into subprime loans are presumably more likely than other types of borrowers to be unable to make their monthly payments, due to insufficient income and lack of access to other forms of credit. Moreover, when interest rates reset for adjustable-rate mortgages, monthly mortgage payments can rise by large amounts and make it difficult for borrowers to meet their monthly debt obligations.

We build an econometric model that nests these four possibilities and thereby permits us to quantify the relative importance of each factor. The dependent variable in the model is the decision to default. Households act as utility maximizers and default if either the expected utility from continuing to make mortgage payments falls below the utility from defaulting or if the household becomes liquidity-constrained. The former comparison is based on an equation that depends on home prices, expectations about future home prices, and the interest rate environment. We also include a second equation capturing the borrower's ability to continue making payments on the mortgage, in order to allow for the possibility of default due to liquidity constraints. We show that our structural equations can be specified as a bivariate probit model with partial observability, whose general features were first studied by Poirier (1980). As robustness checks, we also estimate two alternative specifications: a competing hazards model with unobserved borrower heterogeneity—similar to the approach in Deng, Quigley, and van Order (2000)—as well as a univariate probit model.

We find evidence for each of the hypothesized factors in explaining default by subprime mortgage borrowers. In particular, our results suggest that declining house prices and borrower and loan characteristics affecting borrowers' ability to pay are the two most important factors in predicting default. The effect of declining home prices on default is substantial. For a hypothetical borrower who purchased a home one year earlier with a 30-year fixed-rate mortgage and no downpayment, a 20% decline in home price makes the borrower 15.38% more likely to default than an otherwise identical borrower whose home price remained stable. However, liquidity constraints are as empirically important a driver as declining house prices, and the recent increase in subprime defaults is closely linked to changes over time in the composition of mortgage recipients. In particular, elevated default rates in the subprime market are to a large extent driven by the worsening credit quality of subprime borrowers, as indicated by higher numbers of borrowers who provide little or low documentation in their loan applications, have low FICO scores, make only small downpayments, or have multiple liens on their properties. Although less important, the increasing prevalence of adjustable-rate mortgages also contributed somewhat to rising foreclosures. Periodic resets for ARMs sometimes resulted in large increases in required monthly payments, forcing liquidity-constrained borrowers to default.

There is a wealth of literature examining various aspects of mortgage borrowers' decision to default. One strand of research has focused on the put-option nature of default by studying how net equity or home prices affect default rates (Deng, Quigley, and van Order, 2000; Foote, Gerardi, and Willen, 2008a; Gerardi, Shapiro, and Willen, 2008). Other studies have examined the importance of liquidity constraints and the ability of borrowers to pay, as measured by their credit quality (Archer, Ling, and McGill, 1996; Carranza and Estrada, 2007; Demyanyk and van Hemert, 2008), as well as the role of rate resets for adjustable-rate mortgages (Pennington-Cross and Ho, 2006).

We build on the literature by considering each of the factors proposed by the above researchers. However, our analysis differs from the previous literature in at least four respects. First, our econometric model nests the various potential incentives for default inside a unified framework. In particular, we depart from the previous literature by allowing for default to result from either of two latent causes: financial incentives making default the action that maximizes lifetime utility and binding household liquidity constraints. The likelihood function of our model takes into account the fact that we do not observe which of the two underlying causes actually triggers default in each particular case. Carefully distinguishing between these two causes is important, because unlike prime borrowers, subprime mortgage borrowers tend to have poor credit quality and thus are likely to face liquidity constraints in making monthly payments. Second, our data set includes recent observations from a nationally representative sample of subprime mortgages, allowing us to focus on the drivers behind the recent wave of mortgage defaults. In contrast, a closely related paper by Deng, Quigley, and van Order (2000) examines prime mortgage borrowers, for whom default is much less common. Third, the level of detail in our data allows us to control for loan terms and borrower risk factors that some previous work could not adequately take into account. Moreover, our paper systematically examines the effects of several variables that economic theory suggests ought to affect the decision to default, including expectations about home prices, the volatility of home prices, the amount of time remaining until the next rate reset for ARMs, and the ratio of monthly mortgage payments to monthly income. By controlling for a more comprehensive list of potential drivers of default, we are better able to assess the relative importance of various factors, as compared to the existing literature. Finally, in contrast to more descriptive work such as Demyanyk and van Hemert (2008), we estimate structural equations derived from a model of default in which borrowers maximize their utility and face liquidity constraints.

The rest of this paper proceeds as follows. In Section 2, we present a model of borrower default on mortgage loans. In Section 3, we describe the data. Section 4 presents model estimates and other empirical findings. In Section 5, we discuss policy implications of our results. Section 6 concludes.

2 Model

Our model of housing default builds on the empirical frameworks proposed by Deng, Quigley, and van Order (2000), Crawford and Rosenblatt (1995), and Archer, Ling, and McGill (1996). The empirical literature has traditionally modeled mortgage default using option pricing theory, where the decision to default is treated as a put option. In this framework, it is optimal for a homeowner to default if and only if doing so increases her wealth. Following this earlier literature, we begin by considering the case of a frictionless environment without transaction costs or credit constraints. We next incorporate expectations about home prices, interest rates, and credit constraints into our model. We demonstrate that a household's optimal decision rule takes the form of a system of two inequalities and can be represented as a bivariate probit with partial observability, a type of model first studied by Poirier (1980).

A natural alternative to our framework would be to estimate a completely specified, structural dynamic model of the decision to default in the spirit of Rust (1987). We do not follow this approach in our paper for three reasons. First, our data set contains 2.6 million observations of default decisions made by 135,000 borrowers over multiple months. The approach of Rust (1987) is computationally intensive and would require computing the optimal default decision for each of these borrowers. This is not computationally feasible without the use of multiple processors and supercomputing. Second, this approach requires us to fully specify the model. In particular, we would need to estimate an auxiliary time series model of home price dynamics in order to specify an agent's beliefs about the future evolution of home prices. This is difficult in our application because home price dynamics in the last decade were atypical. Such large, nationwide increases and then decreases in home prices have not been observed in the post-war period. Misspecifying beliefs about home prices could lead to large biases in our parameter estimates and potentially lead us to misinterpret the causes of the current default wave. Finally, our data set has a large number of covariates, which capture heterogeneity in borrowers' ability and willingness to continue paying their mortgages. Because of its computational complexity, the approach of Rust (1987) typically requires the researcher to limit attention to just a few state variables. We believe that this is not appropriate for a first set of estimates, since it would limit our ability to learn about the influence of this rich set of covariates on default decisions.

Our approach instead is to build on the papers listed above. Our model of optimal default decisions is rigorously derived from economic theory. However, we rely on more parsimoniously specified models which have been widely used in the empirical literature. As a result, we can estimate our model using standard techniques from the discrete choice literature. This more parsimonious specification has two advantages. First, we can consider multiple causes for the current default wave. Second, we can include a large number of variables in our model to control for borrowers' ability and willingness to pay. A fully specified structural model would not have this flexibility because of computational costs.

In future work, we plan on extending our results by estimating a fully specified structural model as in Rust (1987). This will allow us to study counterfactuals, such as how borrowers would alter default decisions in response to different mortgage contracts. We believe that the framework in this paper will help us to justify the modeling restrictions that are required to estimate a more complicated, fully structural model.

2.1 Optimal Default without Liquidity Constraints

Let i index borrowers and t index time periods. Let V_{it} and L_{it} refer to the value of borrower i's home and the outstanding principal on i's mortgage at time t, respectively. We can normalize the time period in which i purchases her home to $t = 0$. Let g_{it} denote the nominal rate of increase in home prices between time periods $t-1$ and t. Thus,

$$V_{it} = V_{i0} \prod_{t'=1}^{t} (1 + g_{it'}) \tag{1}$$

That is, the current home value is the initial home value times the rate of increase in home prices between time periods 0 and t. In our empirical analysis, we define g_{it} using the Case-Shiller price index corresponding to the location (MSA) and tercile of the appraised value of i's house at the date of origination. Empirically, there has been considerable time-series variation in V_{it}. In general, $V_{it} > V_{i0}$ for buyers who have held their homes for many years. However, for more recent buyers it may be the case that $V_{it} < V_{i0}$ because of a nationwide decline in home prices starting in mid-2006. There has been considerable cross-sectional variation in the magnitude of home price declines as well. San Francisco and Las Vegas have experienced 33% and 37% declines from their peaks respectively, while Dallas and Charlotte have witnessed flat home prices. Moreover, the magnitude of price declines varied substantially across houses in different price tiers. From April 2006 to April 2008, the Case-Shiller index averaged across cities declined by 21.3% at the bottom tercile, by 18.6% at the middle tercile, and by 14.3% at the top tercile.

The evolution of the outstanding principal, L_{it}, is more complicated. L_{it} depends on the original loan amount, loan term, contract interest rate, rules for interest rate resets, and the history of mortgage payments. In order to economize on notation, we shall not write down an explicit formula for L_{it}. However, the empirical analysis makes use of the fact that we observe in the data the outstanding principal as well as a complete specification of the contract terms that determine how L_{it} evolves.

2.1.1 Frictionless Environment

First consider optimal default in a "frictionless" environment in which there are no penalties from default (either explicit or in terms of damaged credit), no transaction costs (including search costs of finding a new house), and no credit constraints. By assumption, a borrower who defaults is able to immediately repurchase another house. In this extremely stylized model, i will choose to default if and only if

$$V_{it} - L_{it} < 0 \tag{2}$$

If $V_{it} - L_{it} > 0$, then default would be suboptimal, because the borrower's overall wealth would decline by the amount of her net equity, $V_{it} - L_{it}$. If $V_{it} - L_{it} < 0$, the borrower could default, thereby increasing her wealth by $L_{it} - V_{it}$, and then repurchase an identical house.[2]

Even this highly stylized model has testable predictions for cross-sectional and time-series variation in default behavior. First, the model predicts that ceteris paribus, default is more likely for homeowners in markets such as San Francisco and Las Vegas than in markets like Dallas and Charlotte, because homeowners in the former cities have experienced larger drops in V_{it}. If the decline is sufficiently large, inequality (2) will hold, triggering default. Similarly, due to recent price drops, default should be more likely for recent home buyers, whose homes are more likely to be worth less currently than at the time of purchase. Third, our model predicts that borrowers who have made only small downpayments (and therefore have higher L_{it}) are more likely to default.

2.1.2 Expectations about Home Prices

Next, we generalize our model to include expectations about future home prices. The relevant home price as far as optimal default is concerned is the market value *at the time of sale*. In typical housing markets, there is at least a three-to-six month lag between when a home is listed and when the home is sold. As a result, current default will depend on the household's expectations. Let Eg_{it} represent borrower i's expectation in period t, given her current information, about the future growth rate in home prices. The

[2] This argument holds even for exotic loans such as interest-only loans. One might think that since borrowers do not make any principal payments for some months under interest-only loans, the borrowers would not have an incentive to default on their mortgages even if $V_{it} - L_{it} < 0$. However, if the borrowers default on their current mortgage and obtain a new, identical interest-only mortgage for a home worth L_{it}, they can enjoy a greater flow of housing services from a more valuable asset while still not making any principal payments.

borrower will be able to sell the home in the next period for an expected value of $V_{it}(1 + Eg_{it})$. If she is risk-neutral and there is no discounting, it is optimal for her to default if and only if

$$V_{it}(1 + Eg_{it}) - L_{it} < 0 \tag{3}$$

Our model predicts that default will be more likely in cities where homeowners forecast steep price declines than in cities where borrowers expect prices to remain flat. We shall describe our approach to measuring Eg_{it} in the next subsection.

A richer model might also allow the default decision to depend on higher-order moments of future home prices. For example, the higher the variance in home prices over time, the greater the potential gains in housing wealth if prices go up, while the potential downside is limited by the option to default. The added option value generated by higher price volatility decreases the incentive to default even if inequality (3) holds, to a degree depending on the borrower's level of risk aversion. Modeling the impact of variance on consumer utility in a completely structural manner is beyond the scope of this paper. As a compromise, we modify (3) to also control for the reduced-form effect of the variance of g_{it}:

$$V_{it}(\alpha_1 + \alpha_2 Eg_{it} + \alpha_3 Vg_{it}) - L_{it} < 0 \tag{4}$$

The terms α_1, α_2, and α_3—to be estimated in our empirical application—are free parameters that allow the default decision to depend flexibly on V_{it}, Eg_{it}, and Vg_{it}. We include the parameter α_1 because the presence of transaction costs—for instance, the typical 6% commissions paid to real estate agents—causes the actual value of the home to the borrower to potentially deviate from V_{it}.

Economic theory suggests additional reasons why expectations should enter into the default decision. First, there are costs to defaulting, including the transaction costs associated with finding a new house to rent or buy and the cost of having a damaged credit history. The addition of these costs makes the default decision a dynamic optimization problem whose solution depends on expectations about future states of the world, including the evolution of housing prices. Second, option pricing theory suggests that if agents are not risk-neutral, the appropriate pricing kernel depends on higher moments of the process by which home prices evolve over time. Fully modeling these complications is beyond the scope of this paper. Our approach instead is to capture the first order effects of expectations by including the first two moments as in (4).

An important empirical problem is that it is not clear how to derive expectations of home prices from the data. The home prices in our sample were atypical, with large nationwide increases and then decreases that have not been observed before in the post-war period. Our parsimonious approach allows us to focus on the problem of alternative strategies for recovering expectations from the data, which we describe below. In future work, having determined the correct empirical model for beliefs about home price dynamics, we will estimate a fully specified structural model in the spirit of Rust (1987). This will allow us to endogenize the impact of price expectations on default more completely.

Measuring Expectations about Home Prices We construct three measures of Eg_{it}: a measure based on the user cost of housing, another that uses recent price trends (constructed from realizations of housing prices prior to t) as a proxy for expectations, and a third measure based on future price trends (constructed from ex post realizations of housing prices, in periods after t). We derive the first measure using the standard formula for the user cost of housing, which is based on the observation that in a housing market in which people can either rent or buy their homes, the marginal buyer must be indifferent between buying and renting. This implies that the user cost of homeownership must equal the annual rent:

$$\begin{aligned} \text{Cost of ownership at time } t \quad &= \\ V_{it}r_t^{rf} + V_{it}\omega_{it} - V_{it}\tau_{it}(r_{it}^c + \omega_{it}) + V_{it}\delta_{it} - V_{it}Eg_{it} + V_{it}\gamma_{it} \quad &= \quad R_{it} \end{aligned} \tag{5}$$

In this equation, V_{it} is the house price and R_{it} the annual rent. The term r_t^{rf} is the risk-free rate of return at time t, and therefore $V_{it}r_t^{rf}$ is the forgone interest from owning a home. ω_{it} is the property tax rate, τ_{it} is the effective tax rate on income, and r_{it}^c is the contract interest rate. The term $V_{it}\tau_{it}(r_{it}^c+\omega_{it})$ represents savings to the homeowner due to the tax-deductibility of mortgage payments and property taxes. The term δ_{it} represents the depreciation rate of the house, Eg_{it} the expected capital gain, and γ_{it} the risk premium. Using observed values of V_{it}, R_{it}, r_t^{rf}, ω_{it}, τ_{it}, r_{it}^c, δ_{it}, and γ_{it}, we can impute expected housing price appreciation by solving for Eg_{it}:

$$Eg_{it} = r_t^{rf} + \omega_{it} - \tau_{it}(r_{it}^c + \omega_{it}) + \delta_{it} + \gamma_{it} - \frac{R_{it}}{V_{it}} \tag{6}$$

Himmelberg, Mayer, and Sinai (2005) impute Eg_{it} using equation (6) and decompose it into two components. The first component, Eg_i^f, is the expected growth due to "fundamentals," which they proxy using the average annual home price growth rate between 1950 and 2000. We can think of this as the long-run price trend in each market. Note that this term is fixed within an MSA and therefore is captured by the MSA fixed effects in our empirical specification. The remainder, Eg_{it}^b, captures expected growth that is unexplained by fundamental factors, and reflects short-run deviations due to speculative bubbles. Himmelberg, Mayer, and Sinai (2005) report a variant of Eg_{it}^{b} in their paper, which we use in our empirical model. See Himmelberg, Mayer, and Sinai (2005) for more details.

For our second measure of expected house price appreciation, we assume that $Eg_{it} = g_{i,t-1}$. That is, expectations about home prices are adaptive and equal to the previous period's home price appreciation. In principle, we could use a more elaborate time series model to construct a backward-looking measure of Eg_{it}. However, the price trend during the last part of our sample was atypical because of the nationwide home price declines. Therefore, we prefer a simpler specification that does not place weight on observations from the distant past in forming price forecasts.

For the third measure, we assume that $Eg_{it} = g_{i,t+1}$. That is, households have perfect foresight about home price movements—an extreme form of rational expectations—and have expectations equal to the ex post realized home price appreciation in the next period.

2.1.3 Interest Rates

Finally, we allow the optimal default decision to depend on interest rates. Theory predicts that when market interest rates are high relative to the contract rate, the incentive to default is lower. If the contract rate is less than the current market rate available to the borrower, default implies losing the future value of the discount. Following Deng, Quigley, and van Order (2000), we compute the normalized difference between the present value of the payment stream discounted at the contract rate and the present value discounted at the current market interest rate. For borrower i in period t, define

$$IR_{it} = \frac{\sum_{t'=1}^{TM_{it}} \frac{P_{it}}{(1+r_{it}^{m}/1200)^{t'}} - \sum_{t'=1}^{TM_{it}} \frac{P_{it}}{(1+r_{it}^{c}/1200)^{t'}}}{\sum_{t'=1}^{TM_{it}} \frac{P_{it}}{(1+r_{it}^{m}/1200)^{t'}}} = \frac{\sum_{t'=1}^{TM_{it}} \frac{1}{(1+r_{it}^{m}/1200)^{t'}} - \sum_{t'=1}^{TM_{it}} \frac{1}{(1+r_{it}^{c}/1200)^{t'}}}{\sum_{t'=1}^{TM_{it}} \frac{1}{(1+r_{it}^{m}/1200)^{t'}}} \tag{7}$$

In (7), the term P_{it} is the monthly payment for the mortgage, TM_{it} is the number of remaining months until maturity, r_{it}^{m} is the market rate for i at time t, and r_{it}^{c} is the contract rate.[3] Note that IR_{it} is an increasing function of r_{it}^{c}, a decreasing function of r_{it}^{m}, and an increasing function of TM_{it} if $r_{it}^{c} > r_{it}^{m}$. A higher value of IR_{it} implies a stronger incentive to default. For example, households locked into a lower

[3] For adjustable-rate mortgages, P_{it} and r_{it}^{c} may vary over the course of the loan, but for simplicity, we assume that P_{it} and r_{it}^{c} remain constant at the levels of the current month t. We also assume that r_{it}^{m} remains constant at the level of the current month t.

rate are less likely to default. Accounting for IR_{it} yields the optimal default rule

$$V_{it}(\alpha_1 + \alpha_2 Eg_{it} + \alpha_3 Vg_{it}) - L_{it}(1 + \alpha_4 IR_{it}) < 0 \tag{8}$$

The market interest rate r_{it}^m should vary across households because of differences in credit histories and other risk factors, but is not directly observed in our data. Fortunately, the LoanPerformance data cover a large majority of all subprime mortgage originations and detailed borrower characteristics. We can therefore form a very precise estimate of r_{it}^m that controls for both observed and unobserved household-level heterogeneity. Details behind our procedure for estimating r_{it}^m are described in Appendix A.

We would ideally also like to control for expectations about interest rates, just as we control for house price expectations. Doing so in a fully structural way is difficult, particularly because mortgage rates changed in an atypical manner during our sample period, and is beyond the scope of our research. However, we do incorporate one prominent source of interest rate changes: rate resets for adjustable-rate mortgages. If a borrower expects that her contract interest rate will reset to a higher level in the near future, ceteris paribus, she will have a stronger incentive to default. Let MR_{it} denote the number of months before the next rate reset for borrower i in period t.[4] We can then write our default decision as

$$V_{it}(\alpha_1 + \alpha_2 Eg_{it} + \alpha_3 Vg_{it}) - L_{it}(1 + \alpha_4 IR_{it} + \alpha_5 MR_{it}) < 0 \tag{9}$$

Dividing both sides of the equation by L_{it} yields

$$\frac{V_{it}}{L_{it}}(\alpha_1 + \alpha_2 Eg_{it} + \alpha_3 Vg_{it}) - (1 + \alpha_4 IR_{it} + \alpha_5 MR_{it}) < 0 \tag{10}$$

2.2 Liquidity Constraints

In the previous section, we considered a model in which borrowers default whenever doing so increases wealth. This type of model is referred to as "ruthless" default in the mortgage literature (Vandell, 1995). The earlier literature has found that the ruthless default model provides an incomplete explanation of borrower behavior. Researchers have argued that liquidity constraints and access to credit are important explanatory variables in modeling default behavior in mortgage and credit markets more generally (Adams, Einav, and Levin, 2008; Deng, Quigley, and van Order, 2000; Kau, Keenan, and Kim, 1993). In this section, we consider two explanations of how credit constraints may trigger default. The first is that default is triggered by interest rate resets. The second explanation comes from theoretical models of credit constraints.

[4] For fixed-rate mortgages, we set MR_{it}=0 and then include a separate dummy for fixed-rate mortgages.

With regard to the former, an often-cited reason in the popular media for the increase in borrower default rates is that homeowners lack adequate income to make mortgage payments after interest rate resets. For example, on a $300,000 ARM with a 30-year term, an increase in the interest rate from 9% to 11% generates an increase of over $400, or a 15% increase, in monthly payments. This sharp increase in mortgage payments makes it difficult for the household to service its debt in addition to paying for other expenses such as food, gasoline, and clothing.

Let P_{it} denote i's mortgage payment, C_{it} the consumption of a composite commodity, and Y_{it} income. Household i's budget constraint at time t can then be expressed as

$$P_{it} + C_{it} \leq Y_{it} \tag{11}$$

Suppose that consumption C_{it} is fixed in each time period and does not adjust in response to changes in scheduled mortgage payments P_{it}. A household would then be forced to default if it is unable to simultaneously make its mortgage payment P_{it} and purchase C_{it}, i.e., if

$$1 - \frac{P_{it}}{Y_{it}} - \frac{C_{it}}{Y_{it}} < 0 \tag{12}$$

This motive for default is not rigorously grounded in economic theory since it assumes that consumption of the composite commodity is fixed. However, if we are willing to abstract from substitution between consumption of housing and other goods, (12) captures the popular explanation described above.

The budget constraint (12) assumes that household i is in a state of autarky, without any savings or access to credit outside of the mortgage market. We can relax this assumption by allowing households to have access to other forms of credit as well as to tap into savings. Theoretical models of credit constraints suggest that creditworthiness and future income determine the amount i can borrow (see Aiyagari, 1994; Deaton, 1991; Chatterjee, Corbae, Nakajima, and Ríos-Rull, 2007; Chatterjee, Corbae, and Ríos-Rull, 2008). Thus, let Z_{it} be a vector of covariates that serve as predictors of creditworthiness and future income, including credit score or employment status. Z_{it} also proxies for household i's savings in assets other than its house. Incorporating Z_{it} into the budget constraint (12), we assume i defaults if

$$\beta_{0i} + \beta_1 Z_{it} + \beta_2 \frac{P_{it}}{Y_{it}} + \beta_3 Z_{it}(\frac{P_{it}}{Y_{it}}) < 0 \tag{13}$$

The budget constraint (13) nests various liquidity-related triggers of default. The parameters β_1, β_2, and β_3 allow us to flexibly model i's budget constraint as a function of the payment-to-income ratio $\frac{P_{it}}{Y_{it}}$ and covariates Z_{it}. The effect of interest rate resets enters through $\frac{P_{it}}{Y_{it}}$,and the interaction term $Z_{it}(\frac{P_{it}}{Y_{it}})$ allows for the possibility that an increase in the payment-to-income ratio has a bigger impact on borrowers with low credit quality. Note that we do not explicitly include $\frac{C_{it}}{Y_{it}}$ in equation (13). In our empirical application, we assume that the ratio of consumption to income is constant over time and can therefore be captured by allowing for household-level heterogeneity, reflected in the random coefficient β_{0i}.

2.3 Empirical Framework

Equations (10) and (13) represent two drivers of default: borrowers default either because doing so increases their wealth or because credit constraints bind. As econometricians, we only observe whether a given household defaults in each period t. When we observe default, we do not know whether it is due to (10), (13), or both (10) and (13).

We formulate our econometric model by defining two latent utilities, $U_{1,it}$ and $U_{2,it}$, constructed from the left-hand sides of expressions (10) and (13) with the addition of stochastic errors $\varepsilon_{1,it}$ and $\varepsilon_{2,it}$. For household i at time t:

$$\begin{aligned} U_{1,it} &= \alpha_{0i} + \tfrac{V_{it}}{L_{it}}(\alpha_1 + \alpha_2 Eg_{it} + \alpha_3 Vg_{it}) - (\alpha_4 IR_{it} + \alpha_5 MR_{it}) + \varepsilon_{1,it} \\ U_{2,it} &= \beta_{0i} + \beta_1 Z_{it} + \beta_2 \tfrac{P_{it}}{Y_{it}} + \beta_3 Z_{it}(\tfrac{P_{it}}{Y_{it}}) + \varepsilon_{2,it} \end{aligned} \tag{14}$$

$U_{1,it}$ represents the latent utility associated with not defaulting. The term $\varepsilon_{1,it}$ is an iid shock and represents idiosyncratic differences across borrowers and time in their utility from not defaulting. $U_{2,it}$ represents the budget constraint of household i at time t, and $\varepsilon_{2,it}$ is an idiosyncratic shock to the tightness of the household budget constraint. The terms $U_{1,it}$ and $U_{2,it}$ are correlated with each other through the observable covariates V_{it}, L_{it}, Eg_{it}, Vg_{it}, IR_{it}, MR_{it}, $\frac{P_{it}}{Y_{it}}$, and Z_{it}, as well as through the distribution of the unobservables $\varepsilon_{1,it}$ and $\varepsilon_{2,it}$. We assume $\varepsilon_{1,it}$ and $\varepsilon_{2,it}$ are jointly normal with a variance of 1 and a covariance of σ. The terms α_{0i} and β_{0i} capture time-invariant, unobserved borrower heterogeneity in $U_{1,it}$ and $U_{2,it}$. For instance, if borrowers differ in their degree of emotional attachment to their homes, they will exhibit different default behavior even if they face the same financial incentive to default, and the difference is captured by α_{0i}. On the other hand, if borrowers differ in their access to informal sources of credit (such as other family members), such a difference is captured by β_{0i}. α_{0i} and β_{0i} are identified by within-borrower variation in the observable predictors of default, and accounting for the unobserved heterogeneity is important for robustness. In principle, we could estimate α_{0i} and β_{0i} as fixed effects. However, we treat them as random effects, because the large size of our sample makes it computationally costly to estimate fixed effects.

We define the outcome as the random variable ND_{it}, which equals 1 if household i does NOT default in period t and as 0 otherwise. The condition for default is as follows:

$$ND_{it} = I(U_{1,it} \geq 0) \times I(U_{2,it} \geq 0) = 0, \quad \therefore \text{Default} \Leftrightarrow \{U_{1,it} < 0 \text{ or } U_{2,it} < 0\} \tag{15}$$

where $I(\cdot)$ is an indicator function and the outside options for both $U_{1,it}$ and $U_{2,it}$ are normalized to zero. From the data, we observe the value of ND_{it}. However, when default occurs ($ND_{it} = 0$) we do not observe whether it is because $U_{1,it} < 0$, because $U_{2,it} < 0$, or both.

The data-generating process for the observed outcome corresponds to a bivariate probit model with partial observability. By modeling default as the outcome of a two-equation model, our approach contrasts with the existing literature, in which researchers have typically included in a single equation both the determinants of financial incentives and measures of liquidity (Archer, Ling, and McGill, 1996; Demyanyk and van Hemert, 2008). A single-equation model is misspecified because it fails to account for the fact that the financial incentives are relevant for default decisions only if the liquidity constraint does not bind, and vice versa.

Among the covariates Z_{it} entering the liquidity equation, we include measures of creditworthiness, such as the FICO score,[5] whether the borrower has other mortgage loans on the property, and the monthly unemployment rate at the county level. We also include observable loan characteristics that proxy for credit quality, such as the level of documentation for the loan application and the loan-to-value ratio at origination. Borrowers with low documentation on income or wealth are more likely to have low credit and have liquidity problems. Loans with higher loan-to-value ratios at origination are more likely to attract illiquid borrowers, many of whom would have been unable to obtain mortgages under tighter terms.

In principle we could specify the borrower's decision as a choice among *three* options by distinguishing between prepayment and regular continuation of scheduled payments. In the above baseline specification, the choice of no default includes both prepayment as well as the decision to make only scheduled payments. Therefore, if certain factors influence both default probability and prepayment probability, they are reflected in our coefficient estimates, which essentially capture only the "net" effect. To check whether our key findings are sensitive to this modeling choice, we estimate an alternative model in which prepayment

[5] An important feature of the data is that the FICO score is the score in the household's loan application, and does not reflect any credit risk generated by the loan itself. We can think of the FICO score as the household's creditworthiness just before it took out the mortgage.

and default are dependent competing hazards. As a separate exercise, we also estimate the baseline model after dropping all loans ending in prepayment.

3 Data

Our estimation uses data from LoanPerformance on subprime and Alt-A mortgages that were originated between 2000 and 2007 and securitized in the private-label market. The LoanPerformance data set covers more than 85% of all securitized subprime and Alt-A mortgages. According to the Mortgage Market Statistical Annual, 55%-75% of all subprime mortgages were securitized in the early- to mid-2000s. Because sample selection is based on securitization, our sample may differ from the subprime mortgage market as a whole.

For each loan, we observe the loan terms and borrower characteristics reported at the time of origination, including the identity of the originator and servicer, the type of mortgage (fixed rate, adjustable rate, etc.), the frequency of rate resets (in the case of ARMs), the initial contract interest rate, the level of documentation (full, low, or nonexistent[6]), the appraisal value of the property, the loan-to-value ratio, whether the loan is a first-lien loan, the existence of prepayment penalties, the location of the property (by zip code), the borrower's FICO score,[7] and the borrower's debt-to-income ratio. We exclude from our sample exotic mortgage types such as interest-only or balloon loans, and focus on standard fixed-rate and adjustable-rate mortgages. We further restrict our sample to first-lien mortgages. See Table 1 for variable definitions.

The data also track each loan over the course of its life, reporting the outstanding balance, delinquency status, current interest rate, and scheduled payment in each month. We define default as occurring if either the property forecloses or becomes real-estate owned (REO). Default is a terminal event, so if a loan defaults in month t, the loan is no longer in the sample starting from month $t+1$. One important time-varying variable that enters into the liquidity equation (13) is the payment-to-income ratio, $\frac{P_{it}}{Y_{it}}$.

[6] Full documentation indicates that the borrower's income and assets have been verified. Low documentation refers to loans for which some information about only assets has been verified. No documentation indicates there has been no verification of information about either income or assets.

[7] According to Keys, Mukherjee, Seru, and Vig (2007), FICO scores represent the credit quality of a potential borrower based on the probability that the borrower will experience a negative credit event (default, delinquency, etc.) in the next two years. FICO scores fall between 300 and 850, with higher scores indicating a lower probability of a negative event.

While we do not observe income at the household level in each month, we can impute household income at the time of origination based on the reported front-end debt-to-income ratio.[8] The front-end debt-to-income ratio is available only for a very small fraction (3.5%) of all loans, significantly reducing our sample. To see if our results are sensitive to this sample restriction, we also construct an alternative imputation of household income based on the back-end debt-to-income ratio,[9] which is available for 63% of all loans and therefore permits a much more representative estimation sample to be used. However, our estimation results are similar across the two specifications, so throughout this paper we focus on results based on the subsample for which income can be constructed from the front-end ratio. For more detailed discussions of the LoanPerformance data, see Chomsisengphet and Pennington-Cross (2006), Demyanyk and van Hemert (2007), and Keys, Mukherjee, Seru, and Vig (2007).

Because the LoanPerformance data do not report borrowers' demographic characteristics, we match the loan-level data to 2000-Census data on demographic characteristics at the zip-code level (per-capita income, average household size and education, median age of householder, racial composition, etc.). In addition, we utilize monthly unemployment rates reported at the county level by the Bureau of Labor and Statistics (BLS). These variables proxy for individual-level demographics and employment status. Because our proxies are not measured at the level of households, the resulting measurement error implies that we will not be able to consistently estimate the effect of individual-level demographics and employment status on mortgage default. However, since we expect the proxies to be correlated—and in many cases strongly correlated—with the correct measures, including these variables will still provide evidence about the impact of demographics and employment status on mortgage default.

To track movements in home prices, we use housing price indices at the MSA level from Case-Shiller, which covers 20 major MSAs.[10] The HPI for each MSA is normalized to 100 for January 2000. The home price indices are reported at a monthly frequency, and are determined using the transaction prices

[8] Specifically, we assume that household income stays constant over time, and approximate it by the scheduled monthly payment divided by the front-end debt-to-income ratio, both reported as of the time of origination. The front-end ratio measures housing-related principal and interest payments, taxes, and insurance as a percentage of monthly income.

[9] The back-end debt-to-income ratio measures all monthly debt obligations, including mortgage payments, car loans, student loans, and minimum monthly payments on any credit card debt, as a percentage of monthly income. Because we do not have any information on the amount of other loans that each borrower has, the income imputation based on the back-end debt-to-income ratio is noisier than the imputation based on the front-end debt-to-income ratio.

[10] Cities covered by Case-Shiller are Atlanta, Boston, Charlotte, Chicago, Cleveland, Dallas, Denver, Detroit, Las Vegas, Los Angeles, Miami, Minneapolis, New York, Phoenix, Portland, San Diego, San Francisco, Seattle, Tampa, and Washington D.C.

of the properties that undergo repeat sales at different points in time in a given geographic area. Since the index is designed to measure price changes for homes whose quality remains unchanged over time, homes are assigned different weights depending on the length of time between the two transactions, along with other rules of thumb indicating the likelihood that the home has undergone major renovations.[11]

In addition, for 17 out of the 20 MSAs covered by Case-Shiller, the HPI is also broken into three price tiers—low, medium, and high—depending on the quantile of the first transaction price of the property within the distribution of all observed transaction prices occurring during the period of the first sale.[12] We use the tier-specific HPI in constructing V_{it}, Eg_{it}, and Vg_{it} for the 17 MSAs for which it is available because doing so helps capture some of the unobserved within-MSA variation in housing price movements. For the remaining three MSAs, we use the MSA-level HPI.

Table 2 reports summary statistics for key variables. We report separate summary statistics according to the termination mode—prepaid loans, defaulted loans, and loans that are either paid to maturity or censored by the end of the sample. In the last category, virtually all of the loans are censored, so in the following discussion, we shall simply refer to the third category as the "censored" observations.

The raw relationships between the termination mode and measures of borrowers' ability to pay are generally consistent with our hypotheses. Loans that default tend to be adjustable-rate mortgages, are associated with higher initial loan-to-value ratios, and tend to be issued to borrowers with lower credit scores. For instance, fixed-rate mortgages comprise 31.5% of all loans, 26.8% among loans that prepay, and 46.3% among the censored loans, while comprising only 18.4% of loans that default. The average FICO score in the sample is 623 but the average is much lower conditional on default (584).

Table 2 also summarizes time-varying variables, both as an average over the course of each loan (the second panel) as well as for the last period in which we observe each loan (the third panel). Relative to the overall average, borrowers that default tend to have less equity as well as higher payment-to-income ratios. To be more precise about the magnitudes of these effects, $log(V/L)$ is on average 0.481

[11] The index assigns zero weight to houses that have undergone repeat transactions within a span of six months. Lower weights are also assigned to houses for which the change in transaction price is an outlier within a geographic area. Finally, houses with a higher initial sales price are assigned a higher weight.

[12] The three MSAs lacking tier-specific price indices are Charlotte, Dallas, and Detroit. Case-Shiller reports the tier cutoffs (in terms of dollar values) in each MSA for the last month in the series. To back out cutoff values in previous periods, we assume that at each point in time, the low/medium (medium/high) cutoff grows at a rate equal to the average rate of growth for the low- and medium-tier (medium- and high-tier) price indices. (CC)

over the course of each loan and 0.487 in the last observed period. The average is higher conditional on prepayment (0.524 in the last period) and much lower for loans that default (0.361). The average monthly payment-to-income ratio is 0.301 over the course of the loan and 0.312 in the final period. This ratio tends to be highest among loans that default (0.345 in the final period), somewhat lower among loans that prepay (0.320), and lowest among the censored loans (0.281).

Consistent with theory, default tends to occur at points in time when the trend in housing prices is low, as measured by the change over the previous twelve months. The measure based on user costs tells largely the same story, although the forward-looking measure does not show the same pattern. Default is also associated with lower volatility in housing prices, though of course, our measure of volatility (i.e., the normalized standard deviation of housing prices over the previous twelve months) is highly correlated with the trend. Furthermore, as we would expect, the data indicate that conditional on default, borrowers tend to have higher interest rates than the market rate. For loans that end in default, IR has an average value of 0.048 at the point of default, compared to an overall average of 0.037 for the final observation across all loans. Finally, default is more prevalent in areas with lower income and less education.

4 Results

4.1 Bivariate Probit with Partial Observability

We begin by discussing estimates from our baseline model (14), i.e., the bivariate probit with partial observability. The dependent variable in the baseline model is default or no default in a given month. We consider a wide range of alternative specifications in order to assess the robustness of the results to our modeling assumptions. The coefficient estimates and marginal effects are reported in Tables 3-4.[13] Table 3a uses the "backward-looking" measure of home price expectations, Exp_Bwd, based on price trends in the past. In Table 3b, we instead use either Himmelberg, Mayer, and Sinai's measure based on user-costs, Exp_HMS, or the "forward-looking" measure, Exp_Fwd, based on future price trends.

[13] In Tables 3-4, we express all marginal effects in terms of the effect on the probability of "no default," $P(U_1 \geq 0,\ U_2 \geq 0)$, with all independent variables set at their mean values conditional on eventual default. The reason why we evaluate the marginal effects at the mean values conditional on eventual default, instead of the more conventional unconditional sample mean, is because the probability of default in a given period is very low, meaning that we encounter numerical problems when trying to compute the marginal effects at the unconditional sample mean.

In Table 4, we add household-level random effects to the model. For each particular specification in the tables, the column labeled "Eqn 1" includes the covariates and parameter estimates that determine $U_{1,it}$ in equation (14). The column labeled "Eqn 2" includes the covariates and parameter estimates that determine $U_{2,it}$. In Table 5, we display estimates of the change in the probability of default due to an increase in each independent variable by one standard deviation, divided by the baseline default probability. The baseline default probability is defined by setting all explanatory variables equal to their mean values conditional on eventual default.

Table 3a reports the first set of results, beginning with Specification 1, a parsimonious model in which $U_{1,it}$ is only determined by the ratio of the home value to the outstanding loan balance. The discussion in Section 2.1.1 suggests that the incentives to default decrease as this ratio increases. In our empirical analysis, we choose to use the natural logarithm of the ratio of the home value to the loan balance instead of untransformed ratio, because the denominator can be very close to zero for loans that are nearing maturity. Taking the natural log prevents these observations from having unduly large influence on our estimates.

The estimates of Specification 1 in Table 3a are consistent with the predictions of Section 2.1.1: borrowers with lower value-to-loan ratio are more likely to default. Our estimates of the marginal effects in Specification 1 imply that a one-standard-deviation increase in $log(\frac{V}{L})$ is associated with a 24.22% reduction in the hazard of default in a given month. This suggests that the sharp decline in home prices played an important role in the recent increase in foreclosures. Consider a hypothetical household in Phoenix that purchases a home in February 2007 with a 30-year fixed-rate mortgage and no downpayment. The household's $log(\frac{V}{L})$ is then 0 at the time of purchase. Further assume that the household makes monthly payments such that the outstanding balance on the mortgage in February 2008 is $\frac{29}{30}$ of the original loan amount. If there is no change in home price between February 2007 and February 2008, the household's $log(\frac{V}{L})$ in February 2008 would be 0.034. During this time period, however, home prices in Phoenix fell by 21.7%. If this household's property value experienced the average home price change in Phoenix, its $log(\frac{V}{L})$ at the end of this time period would be -0.211. Thus, the decline in home price makes the household 16.9% more likely to default in February 2008 compared to the hypothetical case of no change in home price. The estimated impact of $log(\frac{V}{L})$ decreases when we add MSA- and year fixed effects (Specification 4), since these fixed effects soak up some of the variation in home price changes. However, we still find that net equity in the property plays an important role in default decisions. According to Specification 4, a one-standard-deviation increase in $log(\frac{V}{L})$ is associated with a 7.55% lower hazard of default even after we control for expectations about home price appreciation and volatility (Exp_Bwd, Vol) and MSA- and year fixed effects.

As we would expect from our discussion in Section 2.2, an increase in the ratio of monthly mortgage payments to monthly income predicts an increase in the probability of default. According to Specification 1, a one-standard-deviation increase in this ratio (0.12) is associated with a 17.15% greater hazard of default. This suggests that interest rate resets for ARMs contributed to the recent increase in foreclosures by making it difficult for ARM-holders to meet their increased monthly payments. For Specifications 2–4 of Table 3a, we interact the ratio with the borrower's credit score, and find that the effect is stronger for borrowers with low or medium credit than for those with high credit. This is consistent with the idea that liquidity constraints are less severe for high-credit households because they have greater access to the capital market.

In addition, the estimates in Table 3a indicate that the measures of borrower creditworthiness that affect households' access to the capital market are important drivers of default as well. According to Specification 1, a low-documentation loan has a 0.3 percentage point higher chance of default in a given month, or equivalently, a 39.81% increase in the hazard of default. A borrower who has more than one mortgage on the property is 125.37% more likely to default on her first-lien loan than an otherwise identical borrower with only one mortgage. The marginal effect of a one-standard-deviation increase in the FICO score—about 74 points—corresponds to a decrease in default probability of 0.518 percentage points, or 77.09% of the hazard. Similarly, a one-standard-deviation increase in the original loan-to-value (0.14) is associated with a 21.52% greater hazard, and a one-standard-deviation increase in the local unemployment rate (1.42%) is associated with a 10.09% greater hazard. The magnitudes of these estimates do not vary much across Specifications 1-4.

Following the discussions in Sections 2.1.2 and 2.1.3, we also include additional determinants of the financial incentive to default in the equation for $U_{1,it}$—namely, Eg, Vg, IR, and MR—and report the estimates in Specifications 2 and 3 of Table 3a. Specification 4 is the most comprehensive specification, with the loan age, local demographics, MSA dummies, servicer dummies, and year fixed effects all included as regressors. The estimates from Specifications 2–4 indicate that higher house price growth over the previous twelve months reduces the financial incentive to default. The estimate from Specification 4, for instance, implies that in markets where housing prices have been appreciating at an annual rate 10% above the sample average, the hazard of default is 4.22% lower than for an otherwise identical borrower in an average housing market.

Besides the expected trend, expectations about price volatility also affect default behavior. When we include the volatility of housing prices over the previous twelve months among the independent variables, along with its interaction with $log(\frac{V}{L})$, the uninteracted term has almost no effect, while the interaction decreases the propensity to default. Specifically, at the average level of $log(\frac{V}{L})$, an increase of 1.46 (one standard deviation) in the volatility measure is associated with a 2.77% lower hazard of default, according to our results in Specification 4. Therefore, our findings suggest that volatile home price movements increase the option value of holding onto the mortgage, and that this effect is larger for those borrowers with higher net equity in the property. One conjecture that would imply such a differential effect is that risk aversion declines with wealth. Assuming that households with greater net equity also tend to be wealthier overall, the option value generated by volatility would thus be greater for households with more net equity.

The estimates of the effect of interest rates are weak but consistent with model predictions. We find that IR (a measure of how "overpriced" contract interest rates are, relative to the market rate) has almost no effect on the probability of default. This is most likely because high contract interest rates increase both the incentive to prepay and to default, combined with the fact that prepayment is classified under the category of no default in our baseline specifications. As expected, borrowers with ARMs are riskier. The effect may be due in part to selection (riskier borrowers choosing ARMs). To the extent that MR, IR, and $\frac{P}{Y}$ do not fully capture causal effects of rate resets on default, the coefficient on ARM may also capture some residual causal effects of rate resets. Conditional on everything else being equal, ARM borrowers have a 12% higher hazard of default, according to Specification 4. Among ARM-holders, default is also more likely when rate resets are imminent: adding an extra eleven months between the present period and the next reset results in a lowering of the hazard of default by about 1.34% (Specification 4).

Finally, the parameter estimates for $Loan\ Age$ and $(Loan\ Age)^2$ indicate that over the life of a loan, there is an initial increase in the probability of default, but that after the first three years, older loans are much less likely to default conditional on survival. This "hump-shaped" hazard profile is consistent with the findings of other researchers (Gerardi, Shapiro, and Willen, 2008; von Furstenberg, 1969). Part of this effect could be due to unobserved heterogeneity in households' propensity to default: loans that survive are, by definition, more likely to be held by borrowers with a lower unobserved propensity to default. However, the estimated effect of loan age does not become weaker after controlling for random effects (Table 4), suggesting that the hazard of default for a given individual indeed varies over the life of the loan even after controlling for equity and interest rates.

One might expect that subprime borrowers who live in the collateral property ("owner-occupiers") exhibit different default behavior from those who purchase the property solely for investment purposes ("investors"). Owner-occupiers may have a stronger sentimental attachment to the property, making them reluctant to default even when their financial incentives dictate doing so. In other words, their valuation of the house could diverge from the market value (V) in systematic ways. On the other hand, investors are plausibly less likely to be bound by liquidity constraints because they have better access to the credit market. Obviously, the distinction between "owner-occupiers" and "investors" is not always unambiguous, because most buyers consider possible appreciation when they buy a home and also because they may not truthfully report their expected occupancy status. Despite this measurement issue, we use information in the LoanPerformance data on the occupancy status of each borrower to see whether owner-occupiers and investors behave differently and, if so, how their default decisions are differentially affected by financial incentives and liquidity constraints. Cowan and Cowan (2004) have previously found that default is more likely for properties for which the mortgage borrower is not the occupant. Interestingly, our data indicate the opposite: 8.28% of all loans held by investors end in default while the corresponding figure is 11.02% for owner-occupiers. This difference is driven by the fact that in our sample owner-occupiers have lower net equity (with $log(\frac{V}{L})$ on average at 0.45, versus 0.68 for investors), higher payment-to-income ratios (with $\frac{P_{it}}{Y_{it}}$ on average at 0.31 versus 0.23), lower probability of having a fixed-rate mortgage (28.69% versus 54.31%), higher probability of having multiple liens on the property (13.08% versus 5.68%), and lower FICO scores (615 versus 684).

However, when we re-run Specification 4 separately for owner-occupiers and investors (reported under Specifications 5 and 6 in Table 3a), the results suggest that the financial incentives tend to have stronger *marginal* effects on the probability of default for investors than for owner-occupiers. For instance, the marginal effect of a one-standard-deviation increase in $log(\frac{V}{L})$ is 2.66 times larger for investors than for owner-occupiers. On the other hand, there do not seem to be any systematic differences between owner-occupiers and investors regarding how liquidity constraints affect their default decisions. For example, the original loan-to-value ratio has larger effects on investors' default decisions, while the payment-to-income ratio is more important for the default decision of owner-occupiers.

As another robustness check, we re-run Specification 4 for subprime loans only (that is, excluding Alt-A loans). The results are reported under Specification 7 in Table 3a. Comparing the results from Specification 4 and Specification 7 reveals that the smaller sample of subprime loans does not exhibit systematically different default behavior from the overall sample, indicating that the decision to pool subprime and Alt-A loans in our sample does not drive any of the results.

We also perform some additional robustness checks: (a) adding the initial interest rate as a covariate for the liquidity equation; (b) using "time to first reset" instead of "time to next reset" for ARMs, since the first reset accounts for much of the increase in interest rates due to resets; and (c) adding originator fixed effects.[14] Results from all of these alternative specifications (not reported but available upon request) are very similar to the baseline results.

In Table 3b, we report results from re-running Specifications 3-4 using Exp_HMS and Exp_Fwd instead of Exp_Bwd. The results suggest that the relationship between default and home price expectations depends on how we measure expectations. In contrast to our earlier finding that higher price growth in the previous twelve months (Exp_Bwd) reduces the financial incentive to default, the propensity to default is actually *higher* in markets where the user-cost approach (Exp_HMS) implies stronger house price appreciation, unless we control for demographics as well as MSA, year, and originator fixed effects, in which case the estimated effect lowers the propensity but is insignificantly different from zero. Similarly, the propensity to default is higher in markets that would experience higher price growth in the next twelve months (Exp_Fwd).[15] It thus appears that borrower behavior is more consistent with beliefs that are based on extrapolation, but not with beliefs imputed from the price-to-rent ratio or perfect foresight. We do not have a clear explanation as to why this is the case, and in our upcoming work we plan to investigate the degree to which borrowers are forward-looking and how sophisticated or naive they are about future housing prices. Here we simply note that the price-to-rent ratio may not be a particularly good measure of buyers' expectations if purchase and rental prices are related by a more complicated mechanism than the one proposed by the standard user cost theory (see equation (5)).

Table 4 adds random effects to the model—α_{0i} and β_{0i} in (14)—in order to control for unobserved borrower heterogeneity. In principle, we can include fixed effects in our model. However, given the number of observations in our sample, this is computationally too burdensome. The results for the

[14] We control for the ten largest originators as measured by the number of loans. Collectively, these ten originators account for 70% of the sample.

[15] Estimated coefficients for other financial and credit quality variables are very similar in Tables 3a and 3b.

random effects specification are similar to those reported in Table 3, both qualitatively and quantitatively. As before, higher net equity, higher expectations about home prices (measured using Exp_Bwd), higher volatility in home prices, and lower contract rates all lead to a smaller hazard of default. Similarly, we still find that variables representing higher credit quality and less severe liquidity constraints predict a lower probability of default. The scale parameters for the random effects in $U_{1,it}$ and $U_{2,it}$ are significantly different from zero, suggesting that there is a substantial degree of unobserved borrower heterogeneity influencing the financial incentives to default and the tightness of budget constraint. We find it reassuring that all results carry over to the random effects specifications despite the large degree of unobserved borrower heterogeneity.

Table 5, which reports estimates of the impact of a one-standard-deviation increase in the independent variables on the probability of default, conveys the relative significance of each regressor in default decisions. Alternatively, we could pose a slightly different question. We know that 2006 vintage loans have had much worse performance than 2004 vintage loans. The cumulative empirical probability of default within the sample period is 6.48% and 10.81% for mortgages originated in 2004 and 2006, respectively, despite the fact that older loans have had more time over which to default. How much of this difference in performance between older and newer loans can be explained by observed changes across vintages in each regressor? Such a decomposition would provide an additional way to quantify the relative importance of various factors behind the recent increase in subprime defaults. Table 6 reports the mean value of each regressor conditional on vintage for 2004 and 2006 (columns 1 and 2) and the difference in means (column 3). The product of this difference with the marginal effect of each regressor, as a percentage of the empirical probability of default for 2004 vintage loans (column 4), indicates the contribution of each regressor to the higher default probability of 2006 vintage loans compared to 2004 vintage loans.[16]

Table 6 shows that the biggest contributors to the high probability of default among 2006 vintage loans are declining home prices and deterioration in the credit quality and liquidity conditions of mortgage borrowers. Declining home equity contributed to a 55.84% higher hazard of default for 2006 vintage loans compared to the 2004 vintage. A decrease in house price expectation, measured by recent price trends, made the holders of 2006 vintage loans 39.92% more likely to default than otherwise identical holders of 2004 vintage loans. A decrease in house price volatility, which could largely reflect the slowdown in home

[16] We use the marginal effects from Specification 3 in Table 3a, not Specification 4. Since our objective is to compare loans of different vintages, it makes more sense to use a specification that does not include year fixed effects.

price appreciation, resulted in a 30.26% higher hazard of default for 2006 vintage loans. Lower credit quality, as measured by FICO scores, is responsible for a 69.57% larger hazard of default among 2006 vintage mortgage holders. Moreover, having multiple liens on the property and high payment-to-income ratios among low-credit borrowers are also significant contributors to the high incidence of default among 2006 vintage loans.

4.2 Univariate Probit

In this section, we estimate a univariate probit model as a check for robustness. Similar to the baseline model, the outcome is default or no default in a given month. Here, however, we assume that both the financial incentive to default and borrowers' liquidity constraints enter into a single equation. Table 7 reports the estimates.

Specification 1 uses only the covariates included in the financial incentives equation (Eqn. 1) from Specification 3 of Table 3a. Similarly, Specification 2 includes only the covariates related to the liquidity equation (Eqn. 2) from Specification 3 of Table 3a, except for loan age. Each of these specifications is equivalent to the bivariate probit model with the constant term for one of the two equations constrained to equal infinity and the covariance of the errors constrained to equal zero. The model fit, as measured by the loglikelihood or pseudo-R^2, is higher for Specification 2 than for Specification 1 of Table 7, providing additional support for the notion that illiquidity is as empirically important a driver behind default as financial incentives.

From Specifications 3-5 of Table 7, we see that the qualitative results from the univariate probit model are generally similar to those from the bivariate probit model, with a few notable exceptions. Similar to before, default is more likely if the borrower has low net equity in the house. However, the effects of expectations about home prices are very different, with all expectation measures taking on different signs from the bivariate case. Another notable difference is the impact of home price volatility. The interaction between volatility and net equity in the univariate probit regression suggests that volatility decreases the probability of default more strongly for low-net-equity homeowners than for high-net-equity homeowners. The implications are the same as before for the measures related to interest rates. Default is less likely when the market interest rate is higher than the contract rate, because default entails losing access to the discounted rate. Likewise, for ARMs, default is more likely as the next rate reset gets closer in time. Parameter estimates for the measures that represent the liquidity constraint and the overall credit quality of the borrowers confirm prior results: a high payment-to-income ratio, low FICO score, low documentation level, having multiple liens, and high loan-to-value ratio at origination all lead to an increased probability of default.

To partially address the concern that continued payments and prepayments are rather distinct events, we re-run Specification 3 after excluding prepaid cases from the category of no default. The results are reported under Specification 6 of Table 7, which shows that most of the coefficients are qualitatively stable. Finally, we address the potential concern that the outstanding loan amount, L, is endogenous because of delinquent payments. Loans typically experience several months of delinquency prior to foreclosure (which is our definition of default). Thus, riskier loans are more likely to be delinquent as well as to default, thereby generating positive correlation between L and unobserved determinants of the propensity to default. To address this endogeneity issue, we construct for each loan in each month the hypothetical balance that would have been realized if the household had made payments according to schedule and use the hypothetical $log(\frac{V}{L})$ as an instrument for $log(\frac{V}{L})$. The results (not reported) are essentially unchanged because of near-perfect correlation between $log(\frac{V}{L})$ and the instrument.

Although the overall results are qualitatively similar between the bivariate and univariate probit models, the magnitudes of the coefficients for most of the variables tend to be larger in the bivariate probit results than in the univariate probit, and sometimes much larger. The financial incentives become directly relevant for default only when the liquidity constraint does not bind, and vice versa. Because the univariate probit model does not take this dependency into account, the univariate probit model underestimates the effect on default of certain key factors. Our bivariate probit model does not suffer from this misspecification, giving us better parameter estimates.

4.3 Competing Hazards Model

As an alternative to the probit approach, we also estimate a competing hazards model in which a mortgage can be terminated by either default or prepayment. Compared to the probit models, the hazards approach has the advantage of treating prepayment and regularly scheduled payment as separate outcomes and allowing default and prepayment to be correlated due to unobservables. Moreover, the hazards model correctly accounts for the fact that we essentially observe only one outcome for each loan—the point in time when the loan defaults (if ever)—by treating the *time* to default (or prepayment) as the dependent variable. By contrast, the period-by-period probit model treats the status of the loan in each month as a separate observation, artificially deflating the standard errors.[17] On the other hand, relative to the bivariate probit model, the hazards model has the disadvantage of being potentially misspecified because it treats default as being determined by only a single equation that includes both the covariates for the financial incentive to default and as well as the covariates for household budget constraints.

For household i, denote the time of default as T_{di} and time of prepayment as T_{pi}, where T_{di} and T_{pi} are discrete random variables (Obviously, at least one of these stopping times must be censored). The probabilities of survival past some time t in the future are

$$\begin{aligned} P[T_{di} > t] &= exp\left[-\sum_{k=1}^{t} h_{di}(k)\right] \\ P[T_{pi} > t] &= exp\left[-\sum_{k=1}^{t} h_{pi}(k)\right] \end{aligned} \tag{16}$$

Suppose the instantaneous hazards of default and prepayment of household i, $h_{di}(t)$ and $h_{pi}(t)$, follow a proportional hazards model:

$$\begin{aligned} h_{di}(t) &= exp(\lambda_d(t) + \gamma_d' X_{it} + \eta_{di}) \\ h_{pi}(t) &= exp(\lambda_p(t) + \gamma_p' X_{it} + \eta_{pi}) \end{aligned} \tag{17}$$

In other words, the hazards depend on time-dependent "baselines" $\lambda_d(t)$ and $\lambda_p(t)$, common across all borrowers; on (potentially) time-varying covariates, X_{it}; and on unobserved, borrower-specific random effects η_{di} and η_{pi}. Changing the observed covariates X_{it} results in proportional shifts of the hazard function relative to the baseline hazards, hence the name "proportional hazards."

As is well known (Lunn and McNeil, 1995), if the unobserved heterogeneity terms η_{di} and η_{pi} are independent, the two risks are independent conditional on observables, so separate estimation of the two hazard functions yields consistent estimates. When estimating the hazard of default, we would simply treat loans that end in prepayment as censored observations, similar to loans that are censored by the end of the sample period. Similarly, when estimating the hazard of prepayment, either a default or the end of the sample period would result in censoring. When η_{di} and η_{pi} are not independent, estimation becomes more involved, but we can still estimate the parameters using maximum likelihood, as in Deng,

[17] The clustered standard errors that we report partially address this problem, but not entirely. To further investigate how treating each period as a separate observation might affect our standard errors, we re-run various specifications of the univariate probit model using data aggregated by quarter, instead of using monthly observations (and still clustering our standard errors). The results are reported in Table A1. Using quarterly observations increases our standard errors only very slightly, indicating that clustering largely mitigates the problem of understated standard errors.

Quigley, and van Order (2000) and McCall (1996). The likelihood function and estimation details for the dependent competing hazards model are provided in Appendix B.

In Specification 1 of Table 8, we report the estimation results for a particular case of independent hazards. Specifically, we assume that the hazards only depend on observables (i.e., $\eta_{di} = \eta_{pi} = 0$), while making no parametric assumptions about the underlying baseline hazards, $\lambda_d(t)$ and $\lambda_p(t)$. This specification is simply the standard Cox proportional hazards model (Cox, 1972; Cox and Oakes, 1984), and we can estimate the coefficients γ_d and γ_p by minimizing the "partial loglikelihood," while essentially netting out the baseline hazards.

We also estimate a specification that allows for unobserved correlation in the hazards of default and prepayment, following Deng, Quigley, and van Order (2000). Specifically, we assume that there are two types of borrowers, where

$$(\eta_{di}, \eta_{pi}) = \begin{cases} (\eta_{d1}, \eta_{p1}) \text{ with probability } \rho \\ (\eta_{d2}, \eta_{p2}) \text{ with probability } 1-\rho \end{cases} \tag{18}$$

Results for the model with correlated unobserved hazards are reported as Specifications 2-4 in Table 8.

The hazards model estimates generate implications for default behavior that are similar to what we see in the bivariate and univariate probit models. We find that higher net equity decreases the probability of default while increasing the probability of prepayment. A higher home price growth rate in the previous twelve months decreases the hazard of default while increasing the hazard of prepayment. Specification 2 indicates that the greater the contract rate is relative to the market interest rate, the more likely the borrower is to default and prepay. However, as we add more regressors (Specifications 3–4), the impact of the interest rate option IR on default changes sign. Borrowers are less likely to default or prepay if they are farther away from the next rate reset for adjustable-rate mortgages. Again not surprisingly, fixed-rate mortgages are less likely to default or prepay than adjustable-rate mortgages. At a given loan age, both hazards are approximately two and a half times higher for ARMs than for fixed-rate mortgages. The measures of liquidity constraints display similar patterns as before: borrowers with low documentation, low FICO scores, multiple liens, and high payment-to-income ratios are more likely to default.

Finally, we find that there is a high degree of unobserved heterogeneity in both default and prepayment risk, with the unobserved heterogeneity being somewhat greater for default. This result contrasts with the findings of Deng, Quigley, and van Order (2000) who find substantial and statistically significant unobserved heterogeneity in the exercise of the prepayment option but not in the exercise of the default option. The difference between their finding and ours may be due to the differences in the composition of the borrower pools in our respective data sets: while their sample is confined to prime mortgage borrowers, who have a low probability of default in any case, we study subprime mortgage borrowers, for whom the default risk is much higher. Thus, it makes intuitive sense that our sample exhibits a much greater degree of unobserved heterogeneity in default behavior.

5 Discussion

Here, we summarize the key results and place them in a broader context. To recapitulate our main findings: First, declining home prices are an important driver of subprime mortgage default. For a borrower who purchased a home one year earlier with a 30-year fixed-rate mortgage and no downpayment, a 20% decline in home price makes the borrower 15.38% more likely to default than an otherwise identical borrower whose home price remained stable.[18] Second, borrower and loan characteristics affecting borrowers' ability to pay are as empirically important in predicting default as declining house prices, as evidenced by the magnitudes of the marginal effects in Tables 5 and 6.[19] Our results suggest that the increase in defaults in recent years is partly linked to changes over time in the composition of mortgage recipients. Higher numbers of borrowers who have little documentation, low FICO scores, or multiple liens on the same property contributed to the increase in foreclosures in the subprime mortgage market. The increasing prevalence of adjustable-rate mortgages also contributed to rising foreclosures, because the monthly payments for adjustable-rate mortgages come with periodic—and sometimes very large—adjustments, forcing liquidity-constrained borrowers to default.[20]

Viewed in the broader economic landscape, our findings provide the means to quantify the significance of housing prices in generating correlated default. The simultaneous decline in home prices nearly everywhere in the United States caused the degree of correlation across regions to greatly exceed the

[18] Based on Specification 1 in Table 3a. If we use the marginal effects from Specification 4, a 20% decline in home price would make the borrower 4.8% more likely to default, which is much smaller than 15.38%, but still economically significant.

[19] We will also compare loglikelihoods from two specifications: a univariate probit model with financial covariates only and a univariate probit model with measures of credit quality and liquidity constraints only (corresponding to Specifications 1 and 2 of Table 7). The comparison, which we will discuss in the next section, provides additional support for this claim that liquidity constraints are as important as declining home prices in explaining default.

[20] This result is not inconsistent with the claim made by Foote et al. (2008b) that rate resets for ARMs are not the main problem in the subprime market. Our finding suggests that rate resets play a role in default decisions, but that their role is not as important as falling home prices or low credit quality of subprime borrowers.

expectations of credit rating agencies and investors. Indeed, there is much evidence that markets overrewarded geographic diversity in the structuring and pricing of MBS (Nadauld and Sherlund, 2008), suggesting that rating agencies and investors either underestimated the degree of crossregional correlation or underestimated the significance of the housing price channel in driving default. While acknowledging that housing prices are an equilibrium phenomenon, our analysis points to a key way in which existing models may have led to mispricing.

As well, worsening ex post loan outcomes in combination with observed deterioration in the composition of the borrower pool—as measured by FICO scores, loan documentation status, LTV at origination, and other contract terms—informally support the notion that lenders loosened their underwriting standards and lent to a less and less creditworthy pool of borrowers over time. The importance of compositional changes in explaining the observed rise in default suggests the presence of flaws in the securitization process. One consequence of securitization is that loan originators do not typically hold the securities backed by the mortgages that they originate. As a result, they do not bear the ex post consequences of bad mortgages even as they continue to generate income by originating such loans. This agency problem, coupled with underestimation of default risks by the financial market, gave primary lenders an incentive to lower their lending standards in response to the "technological shock" of securitization and other related financial innovations. Thus, a key policy implication is that future waves of default can be averted by measures that reduce originator moral hazard, such as stronger enforcement of lending standards.

Our results also help us to evaluate the potential efficacy of various remedies under consideration by policymakers. We find empirical relevance for both of the key alternative factors that can theoretically lead to default: in our model, borrowers default if *either* their net equity falls below a certain threshold or if they cannot make their monthly payments due to credit constraints. Notably, most of the policy alternatives that have actually been proposed have been directed primarily toward payment affordability. For example, the FDIC proposal aims to reduce mortgage payments to no more than 31 percent of borrowers' pretax income, by means of loan modifications that reduce interest rates or extend the lengths of the loans. Measures taken by the banking industry similarly focus on modifying loan terms to make payments more affordable. While these loan modifications would be helpful in stanching default due to liquidity constraints, they would have little effect on borrowers' equity positions. The empirical importance of both illiquidity and net equity as drivers of default suggests that effectively mitigating foreclosures would require some combination of policies targeting each cause or a single instrument that targets both. Write-downs on loan principal amounts are an example of a measure that would address both causes simultaneously, because a reduction in loan size would increase a borrower's net equity as well as reduce the monthly payments.[21]

6 Conclusion

We estimate a model of optimal default by subprime mortgage borrowers. Our model nests four possible explanations for the recent increase in mortgage defaults: falling home prices, lower expectations about home prices, increases in borrowers' contract interest rates relative to market rates, and inability to pay due to a lack of income or access to credit. The first three factors affect borrowers' financial gains from default, while the last factor represents the possibility that liquidity constraints may force borrowers to default even when defaulting is against their long-term financial interests. We account for the fact that the financial incentives are relevant to default decisions only if the liquidity constraint does not bind, and vice versa. The structural equations of this model can be represented as a bivariate probit with partial observability.

We estimate our model using unique data from LoanPerformance that track each loan over the course of its life, and find evidence for each of the hypothesized factors in explaining default by subprime mortgage borrowers. In particular, our results suggest that borrower and loan characteristics that affect borrowers' ability to pay are as important in predicting default as the fundamental determinants of whether it makes financial sense to default. Declining home prices are indeed an important driver behind the recent surge in defaults, but for the particular segment of homeowners represented in our data, deterioration in the credit quality of the pool of borrowers is an equally important factor.

Our findings have broad macroeconomic implications. In particular, the evidence points to flaws in the securitization process that led to the current economic downturn. The estimated effect of housing prices on default behavior implies that default will be highly geographically correlated when home prices decline nationwide. Failing to take into account the correct magnitude of this source of correlation would cause existing pricing models to underestimate the degree of nondiversifiable risk, impairing the proper functioning of capital markets. Moreover, deterioration in the observed characteristics of the

[21] In a recent speech the Chairman of the Federal Reserve, Ben Bernanke, also seemed to advocate simultaneously targeting both drivers of default.

borrower pool suggests that underwriting standards became looser, perhaps an equilibrium response to worsening asymmetric information between lenders and securitizers. Finally, our findings suggest that for a foreclosure mitigation policy to produce the desired result, it must address both declining home equity as well as borrowers' ability to pay in the short run. In this regard, write-downs on loan principal amounts may be an effective measure.

The framework of this paper is essentially static: to capture the dynamic nature of borrowers' default decisions, we simply account for the reduced-form effects of various option values associated with holding a mortgage. In future work, we shall examine how our results change when we explicitly model borrowers' default decisions as an optimal stopping problem. The findings in this paper suggest the importance of including credit constraints and inform us on how best to model price expectations in the fully dynamic model.

References

[1] **Aguirregabiria, Victor, and Pedro Mira.** 2007. "Swapping the Nested Fixed-Point Algorithm: a Class of Estimators for Discrete Markov Decision Models." *Econometrica*, 70(4): 1519-1543.

[2] **Adams, William, Liran Einav, and Jonathan Levin**. 2008. "Liquidity Constraints and Imperfect Information in Subprime Lending." *American Economic Review*, forthcoming.

[3] **Archer, Wayne, David Ling, and Gary McGill.** 1996. "The Effect of Income and Collateral Constraints on Residential Mortgage Terminations." *Regional Science and Urban Economics*, 26: 235-261.

[4] **Bajari, Patrick, Lanier Benkard, and Jonathan Levin.** 2007. "Estimating Dynamic Models of Imperfect Competition." *Econometrica*, 75(5): 1331-1370.

[5] **Case, Karl, and Robert Shiller.** 2003. "Is There a Bubble in the Housing Market?" *Brookings Papers on Economic Activity*, 34(2): 299-362.

[6] **Carranza, Juan Esteban, and Dairo Estrada.** 2008. "An empirical characterization of mortgage default in Colombia between 1997 and 2004." Working Paper.

[7] **Chatterjee, Satyajit, Dean Corbae, Makoto Nakajima, and Jose-Victor Ríos-Rull**. 2007. "A Quantitative Theory of Unsecured Consumer Credit with Risk of Default." *Econometrica*, 75(6): 1525-1589.

[8] **Chatterjee, Satyajit, Dean Corbae, and Jose-Victor Ríos-Rull**. 2008. "A Theory of Credit Scoring and Competitive Pricing of Default Risk." Working Paper.

[9] **Chomsisengphet, Souphala, and Anthony Pennington-Cross.** 2006. "The Evolution of the Subprime Mortgage Market." Federal Reserve Bank of St. Louis Review.

[10] **Coppejans, Mark, Donna Gilleskie, Holger Sieg, and Koleman Strumpf.** 2007. "Consumer Demand under Price Uncertainty: Empirical Evidence from the Market for Cigarettes." *Review of Economics and Statistics*, 89(3): 510-521.

[11] **Cowan, Adrian, and Charles Cowan**. 2004. "Default Correlation: An Empirical Investigation of a Subprime Lender." *Journal of Banking & Finance*, 28(4): 753-771.

[12] **Cox, D. R..** 1972. "Regression Models and Life-Tables." *Journal of the Royal Statistical Society*. Series B, 34(2): 187-220.

[13] **Cox, D. R., and David Oakes.** 1984. *Analysis of Survival Data*, Chapman and Hall: London.

[14] **Crawford, Gordon W., and Eric Rosenblatt.** 1995. "Efficient Mortgage Default Option Exercise: Evidence from Loss Severity." *Journal of Real Estate Research*, 10(5): 543-556.

[15] **Demyanyk, Yuliya, and Otto van Hemert.** 2007. "Understanding the Subprime Mortgage Crisis." Working Paper.

[16] **Deng, Yongheng, John M. Quigley, and Robert van Order.** 2000. "Mortgage Terminations, Heterogeneity, and the Exercise of Mortgage Options." *Econometrica*, 68(2): 275-307.

[17] **Dougherty, Ann, and Robert van Order.** 1982. "Inflation, Housing Costs, and the Consumer Price Index." *American Economic Review*, 72(1): 154-164.

[18] **Foote, Christopher, Kristopher Gerardi, and Paul Willen**. 2008a. "Negative Equity and Foreclosure: Theory and Evidence." *Journal of Urban Economics*, forthcoming

[19] **Foote, Christopher, Kristopher Gerardi, Lorenz Goette, and Paul Willen**. 2008b. "Subprime Facts: What (We Think) We Know about the Subprime Crisis and What We Don't." Federal Reserve Bank of Boston Working Paper.

[20] **Foster, Chester, and Robert van Order.** 1985. "FHA Terminations: A Prelude to Rational Mortgage Pricing." *Real Estate Economics*, 13(3): 273-291.

[21] **Friedman, Milton.** 1957. *A Theory of the Consumption Function*, Princeton, NJ: Princeton University Press.

[22] **Gerardi, Kristopher, Adam Hale Shapiro, and Paul S. Willen.** 2007. "Subprime Outcomes: Risky Mortgages, Homeownership Experiences, and Foreclosures." Federal Reserve Bank of Boston Working Paper.

[23] **Gerardi, Kristopher, Andreas Lehnert, Shane Sherlund, and Paul Willen.** 2008. "Making Sense of the Subprime Crisis." Brookings Papers on Economic Activity.

[24] **Han, Aaron, and Jerry Hausman.** 1990. "Flexible Parametric Estimation of Duration and Competing Risk Models." *Journal of Applied Econometrics*, 5(1): 1-28.

[25] **Hendershott, Patric, and Joel Slemrod.** 1983. "Taxes and the User Cost of Capital for Owner-Occupied Housing." *Real Estate Economics*, 10(4): 375-393

[26] **Himmelberg, Charles, Chris Mayer, and Todd Sinai.** 2005. "Assessing High House Prices: Bubbles, Fundamentals, and Misperceptions." *Journal of Economic Perspectives*, 19(4): 67-92.

[27] **Hotz, V. Joseph, and Robert A. Miller.** 1993. "Conditional Choice Probabilities and the Estimation of Dynamic Models." *The Review of Economic Studies*, 60(3): 497-529.

[28] **Keys, Benjamin, Tanmoy Mukherjee, Amit Seru, and Vikrant Vig.** 2008. "Did Securitization Lead to Lax Screening? Evidence from Subprime Loans." Working Paper.

[29] **Krainer, John, and Chishen Wei.** 2004. "House Prices and Fundamental Value." Federal Reserve Bank of San Francisco Economic Letter.

[30] **Lunn, Mary, and Don McNeil.** 1995. "Applying Cox Regression to Competing Risks." *Biometrics*, 51(2): 524-532.

[31] **McCall, Brian.** 1996. "Unemployment Insurance Rules, Joblessness, and Part-Time Work." *Econometrica*, 64(3): 647-682.

[32] **Nadauld, Taylor and Shane Sherlund.** 2008. "The Role of the Securitization Process in the Expansion of Subprime Credit." Working Paper.

[33] **Pennington-Cross, Anthony, and Giang Ho.** 2006. "The Termination of Subprime Hybrid and Fixed Rate Mortgages." Federal Reserve Bank of St. Louis Working Paper.

[34] **Poterba, James.** 1984. "Tax Subsidies to Owner-occupied Housing: An Asset Market Approach." *Quarterly Journal of Economics,* 99(4): 729-752.

[35] **Poirier, Dale.** 1980. "Partial Observability in Bivariate Probit Models." *Journal of Econometrics*, 12(2): 209-217.

[36] **Quigley, John M., and Robert van Order.** 1995. "Explicit Tests of Contingent Claims Models of Mortgage Default." *The Journal of Real Estate Finance and Economics*, 11(2): 99-117.

[37] **Rust, John.** 1987. "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher." *Econometrica*, 55(5): 999-1033.

[38] **Vandell, Kerry.** 1993. "Handing Over the Keys: A Perspective on Mortgage Default Research." *Real Estate Economics*, 21(3): 211-246.

[39] **Vandell, Kerry.** 1995. "How Ruthless is Mortgage Default? A Review and Synthesis of the Evidence." *Journal of Housing Research*, 6(2): 245-264.

[40] **Von Furstenberg, George.** 1969. "Default Risk on FHA-Insured Home Mortgages as a Function of the Terms of Financing: A Quantitative Analysis." *Journal of Finance*, 24(3): 459-477.

Appendix A

Imputation of counterfactual refinancing interest rates for individual households

We assume that in each time period t, household i is able to refinance its mortgage at rate r^m_{it}, the market rate adjusted by a household-specific risk premium. To impute this hypothetical rate from the data, we make the following assumptions about the relationship between r^m_{it} and r^c_{i0}, the initial contract rate owed during the first month of the household's actual loan. Let t_{i0} denote the time period corresponding to the initial month of the actual loan. $j = 1, ..., J$ index covariates, with z_{ij} denoting the observable household and loan characteristics upon which the actual interest rate is determined, and z^m_{ij} denoting the covariates that determine the refinancing rate. Then,

$$\begin{aligned} r^c_{i0} &= f(t_{i0}) + \sum_{j=1}^{J} g_j(t_{i0}) \cdot z_{ij} + \zeta_i \\ r^m_{it} &= f(t) + \sum_{j=1}^{J} g_j(t) \cdot z^m_{ij} + \zeta_i \end{aligned} \tag{19}$$

The function $f(t)$ captures the time-varying "baseline" market interest rate, and the functions $g_j(t)$ capture the time-varying risk premia on characteristics $j = 1, \ldots, J$. By restricting the error, ζ_i, to be equal across equations, we are assuming that the household's risk premium is constant over time. Crucially, among the covariates z_{ij} we include controls for the type of mortgage (adjustable-rate or fixed-rate) held by household i. We assume that $z^m_{ij} = z_{ij}$ for all characteristics j *except* for dummies related to the mortgage type: to preserve comparability across households, we assume that all consumers refinance into fixed-rate mortgages.

We estimate the first equation in (19) using actual originations of *all* subprime mortgages observed in the data. Since the LoanPerformance data cover a large majority of subprime mortgage originations, we can get a very precise estimate of $f(t)$ and $\{g_j(t)\}^J_{j=1}$. The residuals from this estimation are interpreted as household-specific risk premium, ζ_i, unexplained by observable characteristics. We then predict the refinancing rate r^m_{it} using the estimated $f(t)$ and $g(t)$ functions, the observable household characteristics, and the risk premium ζ_i recovered from the first estimation.

We choose to approximate $f(t)$ and $g(t)$ by natural cubic spline functions. A natural cubic spline function $f(t)$ consists of piecewise cubic polynomials $f_n(t), n = 0, \ldots, N-1$ passing through nodes at $t_0, t_1, \ldots, t_N$, with the restriction that $f(t)$ be twice-continuously differentiable at each node and with the boundary conditions $f''(t_0) = f''(t_N) = 0$. The boundary conditions, which impose local linearity at the furthest endpoints, mitigate the tendency for cubic polynomials to take on extreme values near the endpoints.

We include the following variables among the covariates z_{ij}:

- FICO score
- "Low documentation" and "No documentation" dummies
- Dummies for mortgage type. We categorize mortgages as fixed-rate mortgages, ARMs that have a first reset less than a year after origination (which tend to have much lower initial contract rates), and other types of ARMs.
- The loan-to-value ratio at origination
- The front-end debt-to-income ratio: the ratio of monthly housing-related principal and interest payments, taxes, and insurance to monthly income
- The back-end debt-to-income ratio: similar to the front-end ratio, but including in the numerator all payments for non–housing-related debts (e.g., car loans, credit card debt) in addition to mortgage payments

Note that by setting $z^m_{ij} = z_{ij}$ for all characteristics j other than the mortgage type, we abstract from the fact that refinancing generally alters the debt-to-income ratio. Moreover, the debt-to-income and loan-to-value ratios are endogenous, because the amount of debt borrowers are willing to take on is presumably correlated with the interest rates they are able to obtain. We ignore these issues since adequately addressing them is beyond the scope of this paper.

Appendix B

Estimation details for dependent competing hazards model

As a robustness check, we estimate a model of dependent competing hazards. We assume that at time t, borrower i is described by (potentially) time-dependent observable characteristics X_{it} as well as a pair of unobservable characteristics (η_{di}, η_{pi}), which shift the hazards of default and prepayment. We follow Han and Hausman (1990), Deng, Quigley, and van Order (2000), and McCall (1996) in writing the likelihood function of this model.

Denote the time to default as T_d and time to prepayment as T_p, both being discrete random variables. For economy of notation, we omit the subscript for individual i. The joint survival function, conditional on observable characteristics X and unobservable type, is then as follows:

$$\begin{aligned} S(t_d, t_p \mid X, \eta_d, \eta_p) = \exp\big(- &\sum_{k=1}^{t_d} exp(\lambda_d(k) + \gamma_d' X_k + \eta_d) \\ &- \sum_{k=1}^{t_p} exp(\lambda_p(k) + \gamma_p' X_k + \eta_p)\big) \end{aligned} \tag{20}$$

We approximate the baseline hazards $\lambda_d(t)$ and $\lambda_p(t)$ using a third-order polynomial function of time t.

$$\begin{aligned} \lambda_d(t) &= \theta_{0d} + \theta_{1d} t + \theta_{2d} t^2 + \theta_{3d} t^3 \\ \lambda_p(t) &= \theta_{0p} + \theta_{1p} t + \theta_{2p} t^2 + \theta_{3p} t^3 \end{aligned} \tag{21}$$

For the system to be identified, we normalize θ_{0d} and θ_{0p} to zero because these parameters are not separately identified from the population means of η_d and η_p. Re-estimating Specification 1 of Table 8 using nonparametric baseline hazard functions instead of the third-order polynomials yields essentially the same estimates, suggesting that the polynomials give us a good approximation.

Default and prepayment are competing risks, so we only observe the duration associated with the first terminating event. Define $F_d(k \mid X, \eta_d, \eta_p)$ as the probability that the mortgage is terminated by default in period k, $F_p(k \mid X, \eta_d, \eta_p)$ as the probability of termination by prepayment in period k, and $F_c(k \mid X, \eta_d, \eta_p)$ as the probability of censoring in period k by the end of the sample. Following Deng, Quigley, and van Order (2000), and McCall (1996), we can write the probabilities as follows:

$$\begin{aligned} F_d(k \mid X, \eta_d, \eta_p) &= S(k, k \mid X, \eta_d, \eta_p) - S(k+1, k \mid X, \eta_d, \eta_p) - 0.5(S(k, k \mid X, \eta_d, \eta_p) \\ &\quad + S(k+1, k+1 \mid X, \eta_d, \eta_p) - S(k+1, k \mid X, \eta_d, \eta_p) - S(k, k+1 \mid X, \eta_d, \eta_p)) \\ F_p(k \mid X, \eta_d, \eta_p) &= S(k, k \mid X, \eta_d, \eta_p) - S(k, k+1 \mid X, \eta_d, \eta_p) - 0.5(S(k, k \mid X, \eta_d, \eta_p) \\ &\quad + S(k+1, k+1 \mid X, \eta_d, \eta_p) - S(k+1, k \mid X, \eta_d, \eta_p) - S(k, k+1 \mid X, \eta_d, \eta_p)) \\ F_c(k \mid X, \eta_d, \eta_p) &= S(k, k \mid X, \eta_d, \eta_p) \end{aligned} \tag{22}$$

The term $0.5(S(k, k \mid \eta_d, \eta_p) + S(k+1, k+1 \mid \eta_d, \eta_p) - S(k+1, k \mid \eta_d, \eta_p) - S(k, k+1 \mid \eta_d, \eta_p))$ is an adjustment that is necessary because the durations are discrete random variables. Because we do not observe η_d or η_p in the data, we must form the likelihood function using unconditional probabilities, obtained by mixing over the type distribution:

$$\begin{aligned}
F_d(k \mid X) &= \rho F_d(k \mid X, \eta_{d1}, \eta_{p1}) + (1-\rho) F_d(k \mid X, \eta_{d2}, \eta_{p2}) \\
F_p(k \mid X) &= \rho F_p(k \mid X, \eta_{d1}, \eta_{p1}) + (1-\rho) F_p(k \mid X, \eta_{d2}, \eta_{p2}) \\
F_c(k \mid X) &= \rho F_c(k \mid X, \eta_{d1}, \eta_{p1}) + (1-\rho) F_c(k \mid X, \eta_{d2}, \eta_{p2})
\end{aligned} \tag{23}$$

The log likelihood function of this model is then given by

$$\log L = \sum_{i=1}^{N} [I(y_i = d) \log(F_d(K_i)) + I(y_i = p) \log(F_p(K_i)) + I(y_i = c) \log(F_c(K_i))] \tag{24}$$

where $I(y_i = j)$ is equal to one if borrower i's mortgage ends by termination mode j, and equals zero otherwise.

Table 1: Variable Definitions

Variable	Definition
ND_{it}	= 1 if loan *i* does not default in period *t*, = 0 if defaults (in foreclosure or Real Estate Owned)
V/L	Market value of the property / Outstanding principal balance
Exp_Bwd	Home price growth rate over the previous twelve months
Exp_HMS	Expected home price appreciation based on user costs. This is equal to "Imputed Rent / Actual Rent" reported in Himmelberg, Mayer, and Sinai (2005). This measure is normalized to an MSA-specific 24-yr average. See Himmelberg, Mayer, and Sinai (2005) for a detailed description.
Exp_Fwd	Realized home price growth rate over the next twelve months
Vol	Standard deviation of home prices for the past twelve months, divided by 10
IR	Difference in the present values of the payment stream at the mortgage note rate and the current interest rate. Described in the text.
MR	The number of months until the next reset for an ARM
FRM	= 1 if the mortgage is a fixed rate mortgage, = 0 otherwise
Low Doc	= 1 if the loan was done with no or low documentation, = 0 otherwise
FICO	FICO score, a credit score developed by Fair Isaac & Co. Scores range between 300 and 850, with higher scores indicating higher credit quality.
Low FICO	= 1 if FICO is less than 620, = 0 otherwise
Medium FICO	= 1 if FICO is between 620 and 700, = 0 otherwise
High FICO	= 1 if FICO is above 700, = 0 otherwise
Original LTV	Loan to value at origination
Multiple Liens	= 1 if the borrower has other, junior mortgages, = 0 otherwise
Unemployment	Monthly unemployment rate at the country level from BLS
PI ratio	Monthly Payments / Monthly Income. We impute income using debt-to-income ratios. Income stays constant over time. Monthly Payments may vary over time.
Loan Age	The age of the loan in months
Local demographics	Log population, mobility, per-capita income, the median age of householder, average household size, % college educated, % black, % Hispanic

Table 2: Summary Statistics for Estimation Sample

	Prepaid	Defaulted	Censored or paid to maturity	All loans	
		Loan-level variables			
	Mean	Mean	Mean	Mean	Std dev
FRM	.268	.184	.463	.315	.465
Multiple Liens	.093	.150	.176	.123	.328
Original L/V	.766	.798	.747	.764	.139
(FICO score)/100	6.188	5.839	6.462	6.230	.736
Low FICO dummy	.526	.717	.377	.504	.500
Med. FICO dummy	.320	.238	.375	.327	.469
High FICO dummy	.154	.045	.248	.169	.375
Low documentation	.397	.354	.445	.407	.491
No. obs.	81,758	14,459	38,735	134,952	
		Time-dependent loan-level variables over all periods			
$log(V/L)$	.459	.338	.548	.481	.351
Mo. payment/income	.316	.336	.271	.301	.124
Loan age in months	15.381	16.009	21.644	17.794	12.789
Mo. until next reset*	10.880	11.551	8.977	10.537	9.300
Exp_Bwd	.120	.068	.061	.093	.100
Exp_HMS	1.069	1.059	1.145	1.096	.179
Exp_Fwd	-.078	-.033	.049	-.0261	.130
Recent volatility of V	1.741	.985	1.442	1.556	1.458
IR	.036	.044	-.004	.0212	.103
No. obs.	1,392,737	253,796	990,777	2,637,310	

Table 2: Summary Statistics for Estimation Sample, Continued

	Prepaid	Defaulted	Censored or paid to maturity	All loans	
Time-dependent loan-level variables, as last observed for each loan					
	Mean	Mean	Mean	Mean	Std dev
$log(V/L)$	.524	.361	.457	.487	.393
Mo. payment/income	.320	.345	.281	.312	.135
Loan age in months	19.653	20.070	28.837	22.334	14.091
Mo. until next reset*	13.952	14.234	13.060	13.743	10.533
Exp_Bwd	.118	.055	-.0465	.064	.116
Exp_HMS	1.119	1.086	1.206	1.140	.187
Exp_Fwd	-.045	.0006	.257	.047	.189
Recent volatility of V	1.888	1.033	1.235	1.609	1.404
IR	.036	.048	.035	.037	.115
No. obs.	81,758	14,459	38,735	134,952	
Zip-code–level demographics for each loan					
Unemployment (%)	5.460	5.372	5.117	5.352	1.423
log(Population)	10.382	10.357	10.327	10.364	.692
Mobility	.038	-.281	.025	-.0002	2.361
Per-cap. inc. ($K)	22.478	20.768	22.382	22.267	9.667
Pct. college grad.	.350	.326	.347	.347	.129
Pct. black	.184	.289	.201	.200	.272
Pct. Hispanic	.222	.160	.199	.209	.224
Med. Age of HHer	33.916	33.543	34.080	33.923	4.926
No. obs.	81,758	14,459	38,735	134,952	

* Conditional on being an ARM. "Mobility" is the year in which the average resident moved into her current house, minus the nationwide average (1994.697, or August 1994).

Table 3a: Bivariate Probit with Partial Observability using *Exp_Bwd*

	Specification 1		Specification 2		Specification 3		Specification 4	
	Eqn 1	Eqn 2	Eqn 1	Eqn 2	Eqn 1	Eqn 2	Eqn 1	Eqn 2
$\text{Log}(\frac{V}{L})$	1.968 (0.293) 0.005		0.563 (0.081) 0.001		0.613 (0.103) 0.002		0.659 (0.173) 0.0003	
$\log(\frac{V}{L})\times$ Exp_Bwd			11.550 (2.255) 0.024		7.416 (1.109) 0.023		10.459 (2.298) 0.005	
Vol			-0.047 (0.032) -0.0001		0.012 (0.028) 0.00004		-0.061 (0.037) -0.00003	
$\log(\frac{V}{L})\times$ Vol			0.686 (0.148) 0.001		0.541 (0.111) 0.002		0.598 (0.155) 0.0003	
IR					-0.171 (0.076) -0.0005		0.067 (0.112) 0.00003	
MR					0.012 (0.0008) 0.00004		0.015 (0.001) 7.3e-06	
FRM					0.693 (0.037) 0.001		2.049 (0.998) 0.0007	
Low Doc		-0.146 (0.008) -0.003		-0.151 (0.009) -0.003		-0.153 (0.010) -0.002		-0.160 (0.009) -0.003
FICO/100		0.412 (0.007) 0.007		0.453 (0.014) 0.007		0.464 (0.014) 0.006		0.389 (0.012) 0.006
Original LTV		-0.610 (0.036) -0.011		-0.536 (0.036) -0.008		-0.498 (0.042) -0.006		-0.615 (0.038) -0.010

Table 3a Continued								
	Specification 1		Specification 2		Specification 3		Specification 4	
	Eqn 1	Eqn 2	Eqn 1	Eqn 2	Eqn 1	Eqn 2	Eqn 1	Eqn 2
Multiple Liens		-0.348 (0.010) -0.008		-0.342 (0.013) -0.007		-0.380 (0.014) -0.007		-0.328 (0.012) -0.007
Unemployment		-0.028 (0.002) -0.0005		-0.024 (0.003) -0.0004		-0.019 (0.003) -0.0002		-0.020 (0.004) -0.0003
PI ratio		-0.544 (0.076) -0.009						
PI ratio× Low FICO				-0.565 (0.096) -0.009		-0.536 (0.093) -0.007		-0.420 (0.064) -0.007
PI ratio× Medium FICO				-0.673 (0.068) -0.010		-0.657 (0.069) -0.008		-0.558 (0.051) -0.009
PI ratio× High FICO				-0.439 (0.107) -0.007		-0.381 (0.121) -0.005		-0.361 (0.086) -0.006
Loan Age						-0.029 (0.001) -0.0004		-0.030 (0.001) -0.0005
Loan Age^2/100						0.038 (0.002) 0.0005		0.039 (0.002) 0.0006
Demographics								Yes
MSA FE								Yes
Servicer FE								Yes
Year FE								Yes
Corr (ε_1,ε_2)	-0.105 (0.046)		-0.199 (0.054)		-0.711 (0.091)		-0.513 (0.061)	
No. Obs	2625791		2623066		2472282		2439607	
Log Likelihood	-84645.18		-84233.20		-75917.13		-74276.22	

Dependent Variable = No Default. No Default includes both continued payments and prepayments. No random effects.
Contents of each cell: estimated coefficient, standard error, marginal effects. Standard errors are clustered by loan.

Table 3a: Bivariate Probit with Partial Observability using *Exp_Bwd*, Continued

	Specification 5 Owner-Occupiers		Specification 6 Investors		Specification 7 Subprime Only	
	Eqn 1	Eqn 2	Eqn 1	Eqn 2	Eqn 1	Eqn 2
Log($\frac{V}{L}$)	0.492 (0.173) 0.00009		2.065 (0.633) 0.008		0.758 (0.186) 0.001	
log($\frac{V}{L}$)$\times$ Exp_Bwd	9.981 (2.386) 0.002		-10.553 (4.107) -0.039		9.576 (1.493) 0.018	
Vol	-0.074 (0.042) -0.00001		0.417 (0.055) 0.002		-0.136 (0.031) -0.0003	
log($\frac{V}{L}$)$\times$ Vol	0.598 (0.155) 0.0001		0.212 (0.257) 0.0008		0.410 (0.152) 0.0008	
IR	0.060 (0.124) 0.00001		-1.403 (0.377) -0.005		-0.451 (0.100) -0.0009	
MR	0.016 (0.001) 3.0e-06		0.018 (0.005) 0.00007		0.029 (0.002) 0.00006	
FRM	3.145 (1.186) 0.0007		1.146 (0.336) 0.003		1.191 (0.155) 0.001	
Low Doc		-0.150 (0.009) -0.003		-0.246 (0.036) -0.004		-0.162 (0.009) -0.003
FICO/100		0.391 (0.013) 0.006		0.382 (0.040) 0.005		0.352 (0.013) 0.006
Original LTV		-0.573 (0.040) -0.009		-1.304 (0.198) -0.018		-0.439 (0.042) -0.007

Table 3a Continued						
	Specification 5 Owner-Occupiers		Specification 6 Investors		Specification 7 Subprime Only	
	Eqn 1	Eqn 2	Eqn 1	Eqn 2	Eqn 1	Eqn 2
Multiple Liens		-0.330 (0.013) -0.008		-0.255 (0.056) -0.005		-0.314 (0.014) -0.007
Unemployment		-0.016 (0.005) -0.0003		-0.083 (0.016) -0.001		-0.014 (0.005) -0.0002
PI ratio× Low FICO		-0.451 (0.076) -0.007		-0.334 (0.160) -0.005		-0.363 (0.056) -0.006
PI ratio× Medium FICO		-0.610 (0.065) -0.010		-0.170 (0.147) -0.002		-0.431 (0.057) -0.007
PI ratio× High FICO		-0.583 (0.093) -0.010		0.658 (0.365) 0.009		-0.656 (0.090) -0.011
Loan Age		-0.030 (0.001) -0.0005		-0.027 (0.007) -0.0004		-0.029 (0.001) -0.0005
Loan Age^2/100		0.040 (0.002) 0.0007		0.043 (0.015) 0.0006		0.039 (0.002) 0.0006
Demographics		Yes		Yes		Yes
MSA FE		Yes		Yes		Yes
Servicer FE		Yes		Yes		Yes
Year FE		Yes		Yes		Yes
Corr (ε_1,ε_2)	-0.562 (0.069)		-0.414 (0.159)		0.040 (0.066)	
No. Obs	2100019		339588		1884019	
Log Likelihood	-68226.25		-5848.67		-70945.14	

Table 3b: Bivariate Probit with Partial Observability using *Exp_HMS* and *Exp_Fwd*

	Spec 3: Exp_HMS		Spec 4: Exp_HMS		Spec 3: Exp_Fwd		Spec 4: Exp_Fwd	
	Eqn 1	Eqn 2	Eqn 1	Eqn 2	Eqn 1	Eqn 2	Eqn 1	Eqn 2
$\text{Log}(\frac{V}{L})$	1.747 (0.222) 0.007		-0.0797 (0.569) -0.0002		0.780 (0.111) 0.003		0.968 (0.269) 0.0008	
$\log(\frac{V}{L})\times$ Exp_HMS	-0.961 (0.148) -0.004		0.589 (0.461) 0.001					
$\log(\frac{V}{L})\times$ Exp_Fwd					-1.854 (0.278) -0.007		-2.387 (0.942) -0.002	
Vol	0.034 (0.028) 0.0001		0.023 (0.063) 0.00005		0.034 (0.027) 0.0001		-0.002 (0.044) -1.8e-06	
$\log(\frac{V}{L})\times$ Vol	0.781 (0.120) 0.003		0.575 (0.265) 0.001		0.699 (0.115) 0.003		0.647 (0.202) 0.0005	
IR	-0.295 (0.067) -0.001		-0.413 (0.140) -0.0008		-0.242 (0.071) -0.0009		-0.053 (0.149) -0.00004	
MR	0.013 (0.001) 0.0001		0.019 (0.002) 0.00004		0.012 (0.0008) 0.00004		0.016 (0.001) 0.00001	
FRM	0.665 (0.004) 0.002		1.252 (0.328) 0.002		0.671 (0.034) 0.002		1.919 (1.037) 0.001	
Low Doc		-0.158 (0.010) -0.002		-0.168 (0.009) -0.003		-0.156 (0.010) -0.002		-0.162 (0.009) -0.003
FICO/100		0.466 (0.014) 0.006		0.392 (0.013) 0.006		0.467 (0.014) 0.006		0.390 (0.012) 0.006

Table 3b Continued								
	Spec 3: Exp_HMS		Spec 4: Exp_HMS		Spec 3: Exp_Fwd		Spec 4: Exp_Fwd	
	Eqn 1	Eqn 2	Eqn 1	Eqn 2	Eqn 1	Eqn 2	Eqn 1	Eqn 2
Original LTV		-0.472 (0.043) -0.006		-0.651 (0.049) -0.010		-0.472 (0.041) -0.006		-0.600 (0.040) -0.010
Multiple Liens		-0.397 (0.014) -0.007		-0.327 (0.013) -0.007		-0.389 (0.014) -0.007		-0.329 (0.012) -0.007
Unemployment		-0.016 (0.003) -0.0002		-0.020 (0.005) -0.0003		-0.020 (0.003) -0.0002		-0.022 (0.004) -0.0003
PI ratio× Low FICO		-0.525 (0.095) -0.006		-0.416 (0.066) -0.007		-0.532 (0.093) -0.007		-0.422 (0.064) -0.007
PI ratio× Medium FICO		-0.641 (0.077) -0.008		-0.545 (0.057) -0.009		-0.648 (0.069) -0.008		-0.557 (0.052) -0.009
PI ratio× High FICO		-0.355 (0.131) -0.004		-0.387 (0.090) -0.006		-0.348 (0.126) -0.004		-0.355 (0.088) -0.006
Loan Age		-0.029 (0.001) -0.0004		-0.029 (0.001) -0.0005		-0.030 (0.001) -0.0004		-0.030 (0.001) -0.0005
Loan Age^2/100		0.037 (0.002) 0.0005		0.039 (0.002) 0.0006		0.038 (0.002) 0.0005		0.039 (0.002) 0.0006
Demographics		No		Yes				Yes
MSA FE		No		Yes				Yes
Servicer FE		No		Yes				Yes
Year FE		No		Yes				Yes
Corr (ε_1,ε_2)	-0.523		-0.328 (0.047)		-0.563 (0.034)		-0.449 (0.055)	
No. Obs	2390450		2361564		2472282		2439607	
Log Likelihood	-73850.89		-72308.71		-75935.03		-74292.88	

Dependent Variable = No Default. No Default includes both continued payments and prepayments. No random effects.
Contents of each cell: estimated coefficient, standard error, marginal effects. Standard errors are clustered by loan.

Table 4: Bivariate Probit with Partial Observability with Random Effects

	Spec 3: Exp_Bwd		Spec 4: Exp_Bwd		Spec 4: Exp_HMS		Spec 4: Exp_Fwd	
	Eqn 1	Eqn 2	Eqn 1	Eqn 2	Eqn 1	Eqn 2	Eqn 1	Eqn 2
$\log(\frac{V}{L})$	0.581 (0.188)		0.417 (0.256)		0.305 (0.545)		0.505 (0.245)	
$\log(\frac{V}{L})\times$ Exp_Bwd	6.712 (1.057)		6.051 (1.485)					
$\log(\frac{V}{L})\times$ Exp_HMS					0.066 (0.412)			
$\log(\frac{V}{L})\times$ Exp_Fwd							-1.475 (0.518)	
Vol	0.043 (0.048)		-0.016 (0.064)		0.009 (0.073)		0.004 (0.063)	
$\log(\frac{V}{L})\times$Vol	0.342 (0.182)		0.389 (0.232)		0.616 (0.268)		0.562 (0.227)	
IR	-0.207 (0.140)		-0.216 (0.210)		-0.523 (0.219)		-0.300 (0.204)	
MR	0.013 (0.002)		0.016 (0.004)		0.019 (0.004)		0.017 (0.003)	
FRM	0.624 (0.057)		1.018 (0.232)		0.893 (0.153)		0.942 (0.178)	
Low Doc		-0.160 (0.020)		-0.177 (0.018)		-0.186 (0.019)		-0.181 (0.019)
FICO		0.490 (0.025)		0.415 (0.023)		0.418 (0.024)		0.418 (0.023)
Original LTV		-0.445 (0.075)		-0.602 (0.072)		-0.621 (0.075)		-0.594 (0.072)
Multiple Liens		-0.370 (0.031)		-0.305 (0.028)		-0.306 (0.028)		-0.308 (0.028)
Unemployment		-0.020 (0.006)		-0.025 (0.010)		-0.023 (0.010)		-0.206 (0.010)

Table 4 Continued								
	Spec 3: Exp_Bwd		Spec 4: Exp_Bwd		Spec 4: Exp_HMS		Spec 4: Exp_Fwd	
	Eqn 1	Eqn 2	Eqn 1	Eqn 2	Eqn 1	Eqn 2	Eqn 1	Eqn 2
PI ratio× Low FICO		-0.653 (0.028)		-0.524 (0.040)		-0.521 (0.041)		-0.527 (0.040)
PI ratio× Medium FICO		-0.692 (0.080)		-0.606 (0.081)		-0.601 (0.086)		-0.604 (0.082)
PI ratio× High FICO		-0.111 (0.253)		-0.226 (0.183)		-0.267 (0.187)		-0.198 (0.188)
Loan Age		-0.032 (0.002)		-0.034 (0.002)		-0.033 (0.002)		-0.034 (0.002)
Loan Age^2/100		0.041 (0.004)		0.044 (0.003)		0.044 (0.004)		0.045 (0.003)
Demographics				Yes		Yes		Yes
MSA FE				Yes		Yes		Yes
Servicer FE				Yes		Yes		Yes
Year FE				Yes		Yes		Yes
Corr (ε_1,ε_2)	-0.664 (10792.12)		-0.491 (212.08)		-0.372 (44.61)		-0.700 (34375.93)	
RE Scale Parameter	0.020 (0.014)	0.223 (0.010)	0.041 (0.021)	0.246 (0.010)	1.166 (0.022)	0.238 (0.010)	0.039 (0.021)	0.252 (0.010)
No. Obs	651770		651770		651770		651770	
Log Likelihood	-18694.60		-18310.03		-17757.77		-18312.08	

Dependent Variable = No Default. No Default includes both continued payments and prepayments. Contents of each cell: estimated coefficient, standard error. Standard errors are clustered by loan. Due to the computational burden, we use a ¼ random sample of loans for estimation of random effects models.

Table 5: Marginal Effects (based on Table 3a)

		Specification 1	Specification 2	Specification 3	Specification 4
	1 std. dev.	Marginal Effects	Marginal Effects	Marginal Effects	Marginal Effects
Log($\frac{V}{L}$)	0.351	24.22%	32.88%	44.63%	7.55%
Exp_Bwd	0.100		19.24%	21.16%	4.22%
Vol	1.458		14.29%	23.52%	2.77%
IR	0.103			-1.04%	0.06%
MR	10.607			7.25%	1.34%
FRM	1			26.98%	12.00%
Low Doc	1	-39.81%	-42.26%	-40.37%	-48.54%
FICO/100	0.736	77.09%	86.19%	82.83%	79.90%
Original LTV	0.139	-21.52%	-19.25%	-16.76%	-23.82%
Multiple Liens	1	-125.37%	-125.35%	-137.40%	-127.25%
Unemployment	1.423	-10.09%	-8.79%	-6.43%	-7.75%
PI ratio	0.124	-17.15%			
PI ratio× Low FICO	0.184		-26.93%	-23.96%	-21.62%
PI ratio× Medium FICO	0.152		-26.43%	-24.19%	-23.61%
PI ratio× High FICO	0.113		-12.86%	-10.47%	-11.41%
Demographics					Yes
MSA FE					Yes
Servicer FE					Yes
Year FE					Yes

This table reports marginal effects (relative to the hazard of default computed at the sample means conditional on default) associated with a one-standard-deviation increase in each regressor. For binary variables, it is a unit change instead of a one-standard-deviation change.

Table 6: Comparison of 2004- and 2006 Vintage Loans (based on Table 3a Specification 3)

	2004 Mean (1)	2006 Mean (2)	Δ in RHS Variable (2) – (1)	Δ in Default Probability
$\text{Log}(\frac{V}{L})$	0.539	0.307	-0.233	55.84%
Exp_Bwd	0.123	0.023	-0.100	39.92%
Vol	2.111	1.118	-0.993	30.26%
IR	-0.039	-0.003	0.036	0.68%
MR*	13.255	16.417	3.162	-4.09%
FRM	0.288	0.242	-0.046	2.33%
Low Doc	0.515	0.434	-0.081	-6.21%
FICO/100	6.507	6.181	-0.327	69.57%
Original LTV	0.758	0.771	0.013	2.94%
Multiple Liens	0.169	0.242	0.073	19.01%
Unemployment	5.667	4.612	-1.055	-9.01%
PI ratio× Low FICO	0.347	0.529	0.182	44.68%
PI ratio× Medium FICO	0.369	0.362	-0.007	-2.14%
PI ratio× High FICO	0.284	0.109	-0.175	-30.51%

This table reports how the difference in each regressor between 2004- and 2006 vintage loans affects the probability of default, relative to the empirical probability of default in a given month for 2004 vintage loans.
* conditional on being an ARM.

Table 7: Univariate Probit

	Spec 1	Spec 2	Spec 3	Spec 4	Spec 5	Spec 6
$\log(\frac{V}{L})$	0.458 (0.025) 0.008		0.464 (0.076) 0.004	-0.077 (0.133) -0.0007	0.436 (0.077) 0.004	0.192 (0.061) 0.003
$\log(\frac{V}{L})\times$ Exp_Bwd	0.049 (0.148) 0.0009		-0.569 (0.193) -0.005			-2.426 (0.234) -0.037
$\log(\frac{V}{L})\times$ Exp_HMS				0.422 (0.075) 0.004		
$\log(\frac{V}{L})\times$ Exp_Fwd					0.389 (0.084) 0.003	
Vol	0.145 (0.006) 0.003		0.057 (0.009) 0.0005	0.056 (0.009) 0.0005	0.057 (0.009) 0.0005	-0.025 (0.009) -0.0004
$\log(\frac{V}{L})\times$ Vol	-0.102 (0.013) -0.002		-0.014 (0.019) -0.0001	-0.050 (0.015) -0.0004	-0.039 (0.015) -0.0003	0.182 (0.021) 0.003
IR	-0.522 (0.029) -0.009		-0.397 (0.044) -0.003	-0.388 (0.044) -0.003	-0.390 (0.044) -0.003	-0.336 (0.049) -0.005
MR	0.009 (0.0004) 0.0002		0.003 (0.0005) 0.00003	0.003 (0.0005) 0.00003	0.003 (0.0005) 0.00003	0.008 (0.0006) 0.0001
FRM	0.427 (0.009) 0.006		0.237 (0.011) 0.002	0.242 (0.011) 0.002	0.237 (0.011) 0.002	0.442 (0.012) 0.007
Low Doc		-0.131 (0.007) -0.003	-0.138 (0.008) -0.001	-0.139 (0.008) -0.001	-0.138 (0.008) -0.0016	-0.128 (0.009) -0.002
FICO/100		0.365 (0.010) 0.007	0.300 (0.009) 0.003	0.299 (0.010) 0.003	0.301 (0.009) 0.003	0.287 (0.011) 0.004

Table 7 Continued						
	Spec 1	Spec 2	Spec 3	Spec 4	Spec 5	Spec 6
Original LTV		-0.911 (0.031) -0.018	-0.053 (0.104) -0.0005	-0.180 (0.106) -0.002	-0.082 (0.104) -0.0007	-0.407 (0.088) -0.006
Multiple Liens		-0.316 (0.009) -0.008	-0.255 (0.011) -0.003	-0.251 (0.011) -0.003	-0.254 (0.011) -0.003	-0.236 (0.011) -0.005
Unemployment		-0.028 (0.002) -0.0006	-0.013 (0.004) -0.0001	-0.006 (0.004) -0.00005	-0.012 (0.004) -0.0001	-0.038 (0.005) -0.0006
PI ratio× Low FICO		-0.468 (0.062) -0.009	-0.400 (0.031) -0.004	-0.395 (0.032) -0.003	-0.398 (0.031) -0.004	-0.398 (0.037) -0.006
PI ratio×Medium FICO		-0.549 (0.046) -0.011	-0.490 (0.036) -0.004	-0.481 (0.036) -0.004	-0.491 (0.036) -0.004	-0.525 (0.037) -0.008
PI ratio× High FICO		-0.465 (0.063) -0.009	-0.429 (0.063) -0.004	-0.419 (0.064) -0.004	-0.432 (0.063) -0.004	-0.355 (0.066) -0.005
Loan Age			-0.028 (0.001) -0.0002	-0.028 (0.001) -0.0002	-0.028 (0.001) -0.0002	-0.029 (0.001) -0.0004
Loan $Age^2/100$			0.037 (0.002) 0.0003	0.037 (0.002) 0.0003	0.037 (0.002) 0.0003	0.038 (0.002) 0.0006
Demographics			Yes	Yes	Yes	Yes
MSA & Year FE			Yes	Yes	Yes	Yes
Servicer FE			Yes	Yes	Yes	Yes
No. Obs	2472282	2472282	2439607	2361564	2439607	1155760
Pseudo R^2	0.039	0.049	0.083	0.083	0.083	0.116
Log Likelihood	-78436.70	-77574.53	-74001.62	-72001.58	-73993.16	-62708.54

Dependent Variable = No Default. Contents of each cell: estimated coefficient, standard error, marginal effects. Standard errors are clustered by loan. For Specifications 1-5, No Default includes both continued payments and prepayments. For Specification 6, No Default includes continued payments only.

Table 8: Competing Hazards Model

	Specification 1 No Unobserved Heterogeneity		Specification 2 2 Unobserved Types		Specification 3 2 Unobserved Types		Specification 4 2 Unobserved Types	
	Default	Prepay	Default	Prepay	Default	Prepay	Default	Prepay
$\text{Log}(\frac{V}{L})$	-2.258 (0.044) 0.105	0.184 (0.010) 1.202	-2.371 (0.078) 0.093	0.838 0.010 2.312	-3.292 (0.226) 0.037	0.196 (0.049) 1.217	-3.788 (0.232) 0.023	0.060 (0.051) 1.062
$\log(\frac{V}{L})\times$ Exp_Bwd					-1.847 (0.291) 0.158	1.862 (0.053) 6.440	-1.497 (0.299) 0.224	1.626 (0.054) 5.084
Vol					0.200 (0.011) 1.222	0.0007 (0.004) 1.001	0.177 (0.011) 1.193	0.009 (0.005) 1.009
$\log(\frac{V}{L})\times$ Vol					0.207 (0.020) 1.230	0.035 (0.005) 1.036	0.248 (0.020) 1.282	0.049 (0.005) 1.050
IR			0.784 (0.101) 2.189	0.871 (0.045) 2.388	-1.205 (0.111) 0.300	0.690 (0.052) 1.994	-1.311 (0.115) 0.270	0.536 (0.052) 1.709
MR			-0.061 (0.018) 0.941	-0.212 (0.007) 0.809	-0.029 (0.018) 0.972	-0.194 (0.007) 0.823	-0.0008 (0.018) 0.999	-0.179 (0.007) 0.836
FRM			-0.874 (0.034) 0.417	-0.948 (0.015) 0.388	-0.902 (0.034) 0.406	-0.917 (0.015) 0.400	-0.830 (0.034) 0.436	-0.881 (0.015) 0.414
Low Doc			0.376 (0.023) 1.456	0.105 (0.010) 1.111	0.316 (0.023) 1.371	0.097 (0.010) 1.102	0.381 (0.023) 1.464	0.072 (0.010) 1.075
FICO/100			-1.192 (0.026) 0.304	-0.262 (0.010) 0.769	-0.737 (0.028) 0.479	-0.222 (0.012) 0.801	-0.588 (0.029) 0.555	-0.169 (0.012) 0.844

Table 8 Continued								
	Specification 1		Specification 2		Specification 3		Specification 4	
	Default	Prepay	Default	Prepay	Default	Prepay	Default	Prepay
Original LTV			-0.635 (0.157) 0.530	1.670 (0.039) 5.313	0.383 (0.171) 1.467	1.346 (0.043) 3.842	0.231 (0.176) 1.260	1.344 (0.044) 3.836
Multiple Liens			0.881 (0.032) 2.413	-0.118 (0.017) 0.889	0.890 (0.031) 2.434	-0.133 (0.017) 0.876	0.864 (0.032) 2.374	-0.114 (0.017) 0.892
Unemploy-ment			-0.009 (0.007) 0.991	-0.033 (0.003) 0.968	0.007 (0.007) 1.007	-0.030 (0.003) 0.970	-0.016 (0.007) 0.984	-0.031 (0.004) 0.969
PI ratio× Low FICO			1.071 (0.094) 2.920	0.806 (0.036) 2.238	0.922 (0.094) 2.513	0.802 (0.037) 2.230	1.975 (0.104) 7.207	0.911 (0.037) 2.487
PI ratio× Medium FICO			0.842 (0.203) 2.321	1.014 (0.067) 2.756	0.341 (0.203) 1.406	1.006 (0.068) 2.736	1.0447 (0.205) 2.843	0.964 (0.067) 2.622
PI ratio× High FICO			0.163 (0.019) 1.178	-0.158 (0.009) 0.854	0.083 (0.019) 1.086	-0.172 (0.009) 0.842	0.094 (0.020) 1.099	-0.150 (0.009) 0.861
Loan Age							1.479 (0.108) 4.389	1.005 (0.049) 2.732
Loan Age^2/100							-0.192 (0.016) 0.826	-0.172 (0.008) 0.842
Demographics							Yes	Yes
(η_{p1}, η_{d1}, η_{p2}, η_{d2})			(-2.5, 2.9, -4.8, -0.3) SE (0.07, 0.21, 0.08, 0.21)		(-2.5, -2.5, -4.7, -5.5) SE (0.10, 0.26, 0.11, 0.27)		(-4.1, -3.7, -6.4, -6.7) SE (0.13, 0.32, 0.14, 0.33)	
Pr. of Type 1			0.277 (0.004)		0.268 (0.004)		0.269 (0.004)	
No. Loans	120394		120394		120394		120394	
Log Likelihood	-395027.48		- 385070.70		- 383670.32		- 382698.65	

Contents of each cell: estimated coefficient, standard error, hazard ratio. Hazard ratios are exponentiated coefficients and have the interpretation of hazard ratios for a one-unit change in X.

Table A1: Univariate Probit with Quarterly Observations

	Spec 1	Spec 2	Spec 3	Spec 4	Spec 5	Spec 6
$\log(\frac{V}{L})$	0.645 (0.071) 0.016	0.507 (0.079) 0.012	0.882 (0.115) 0.022	-0.129 (0.134) -0.003	0.642 (0.070) 0.016	0.468 (0.079) 0.011
$\log(\frac{V}{L})\times$ Exp_Bwd	-0.084 (0.174) -0.002	-0.963 (0.215) -0.022				
$\log(\frac{V}{L})\times$ Exp_HMS			-0.216 (0.056) -0.005	0.497 (0.081) 0.012		
$\log(\frac{V}{L})\times$ Exp_Fwd					-0.091 (0.067) -0.002	0.547 (0.092) 0.013
Vol	0.124 (0.008) 0.003	0.083 (0.010) 0.002	0.117 (0.008) 0.003	0.081 (0.010) 0.002	0.123 (0.008) 0.003	0.083 (0.010) 0.002
$\log(\frac{V}{L})\times$ Vol	-0.087 (0.018) -0.002	-0.009 (0.021) -0.0002	-0.075 (0.016) -0.002	-0.069 (0.017) -0.002	-0.094 (0.015) -0.002	-0.053 (0.017) -0.001
IR	-0.119 (0.038) -0.003	-0.448 (0.050) -0.010	-0.155 (0.039) -0.004	-0.445 (0.051) -0.010	-0.137 (0.039) -0.003	-0.441 (0.050) -0.010
MR	0.003 (0.0005) 0.00008	0.005 (0.0006) 0.0001	0.003 (0.0005) 0.00008	0.005 (0.0006) 0.0001	0.003 (0.0005) 0.00008	0.005 (0.0006) 0.0001
FRM	0.260 (0.011) 0.006	0.287 (0.013) 0.006	0.255 (0.012) 0.006	0.292 (0.013) 0.006	0.258 (0.011) 0.006	0.287 (0.013) 0.006
Low Doc	-0.154 (0.009) -0.004	-0.154 (0.009) -0.004	-0.152 (0.009) -0.004	-0.155 (0.009) -0.004	-0.154 (0.009) -0.004	-0.154 (0.009) -0.004
FICO/100	0.373 (0.010) 0.009	0.343 (0.011) 0.008	0.373 (0.010) 0.009	0.342 (0.011) 0.008	0.374 (0.010) 0.009	0.344 (0.011) 0.008

Table A1 Continued						
	Spec 1	Spec 2	Spec 3	Spec 4	Spec 5	Spec 6
Original LTV	0.0005 (0.099) 0.00001	-0.100 (0.107) -0.002	0.010 (0.104) 0.0002	-0.253 (0.108) -0.006	-0.011 (0.098) -0.0003	-0.142 (0.106) -0.003
Multiple Liens	-0.325 (0.011) -0.011	-0.286 (0.012) -0.009	-0.321 (0.011) -0.011	-0.282 (0.012) -0.009	-0.324 (0.011) -0.011	-0.285 (0.012) -0.009
Unemployment	-0.005 (0.003) -0.0001	-0.011 (0.005) -0.0003	-0.006 (0.003) -0.0001	-0.003 (0.005) -0.00008	-0.006 (0.003) -0.0001	-0.010 (0.005) -0.0002
PI ratio× Low FICO	-0.580 (0.039) -0.014	-0.444 (0.036) -0.010	-0.578 (0.038) -0.014	-0.439 (0.038) -0.010	-0.581 (0.039) -0.014	-0.440 (0.037) -0.010
PI ratio×Medium FICO	-0.684 (0.041) -0.017	-0.549 (0.041) -0.013	-0.676 (0.041) -0.017	-0.540 (0.042) -0.013	-0.687 (0.041) -0.017	-0.548 (0.041) -0.013
PI ratio× High FICO	-0.618 (0.070) -0.015	-0.514 (0.070) -0.012	-0.600 (0.072) -0.015	-0.504 (0.072) -0.012	-0.621 (0.071) -0.015	-0.517 (0.070) -0.012
Loan Age	-0.025 (0.001) -0.0006	-0.025 (0.001) -0.0006	-0.025 (0.001) -0.0006	-0.024 (0.001) -0.0006	-0.025 (0.001) -0.0006	-0.024 (0.001) - 0.0006
Loan Age^2/100	0.029 (0.002) 0.0007	0.033 (0.002) 0.0008	0.029 (0.002) 0.0007	0.033 (0.002) 0.0008	0.029 (0.002) 0.0007	0.032 (0.002) 0.0007
Demographics		Yes		Yes		Yes
MSA & Year FE		Yes		Yes		Yes
Servicer FE		Yes		Yes		Yes
No. Obs	867402	855887	838616	828452	867402	855887
Pseudo R^2	0.084	0.095	0.083	0.095	0.084	0.095
Log Likelihood	-62174.55	-60675.65	-60444.25	-59023.47	-62173.76	-60666.19
Dependent Variable = No Default. Contents of each cell: estimated coefficient, standard error, marginal effects. Quarters computed starting from month of initial observation for each loan (e.g., for a loan first appearing in the data in 11/2007, the first quarter is 11/2007 – 01/2008.) Right-hand-side variables are averages over quarters.						