A Dynamic Model of Subprime Mortgage Default: Estimation and Policy Implications

Patrick Bajari, University of Minnesota and NBER
Sean Chu, Federal Reserve Board of Governors
Denis Nekipelov, University of California, Berkeley
Minjung Park, University of California, Berkeley

Very Preliminary

This version: May 4, 2012

Abstract

The increase in defaults in the subprime mortgage market is widely held to be one of the causes behind the recent financial turmoil. Key issues of policy concern include identifying the main drivers behind the wave of defaults and predicting the effects of various policy instruments designed to mitigate default. To address these questions, we estimate a dynamic structural model of subprime borrowers’ default behavior. We propose a simple and intuitive estimation method, and use our model estimates to simulate how borrowers’ default behavior would change under various scenarios. The simulation exercises allow us to quantify the importance of various factors, such as home price declines and loosened underwriting standards, in explaining the recent increase in subprime defaults. Furthermore, we use simulations to assess the effects of principal write-downs and other foreclosure mitigation policies on the behavior of various subsets of borrowers.

JEL Classifications: C5, G21
1 Introduction

In this paper, we estimate and simulate a dynamic structural model of subprime borrowers’ default behavior. The collapse of the subprime mortgage market and its subsequent role in triggering the current recession lends special importance to understanding the key drivers behind the increase in defaults. We use our model to quantify the relative importance of various potential drivers of default, such as deep home price declines and loosened underwriting standards. Another key object of interest is how subprime borrowers’ default behavior would respond to the implementation of various policy proposals such as principal write-downs. We use our model to assess the effectiveness and welfare consequences of foreclosure mitigation policies by simulating borrowers’ default decisions under various scenarios.

We model borrowers’ default decisions as a dynamic programming problem. In each period, a borrower takes one of three possible actions: defaulting, prepaying the loan, or continuing to make just the regularly scheduled payments (which we call “paying” throughout the rest of the paper). Because mortgages have a fixed maturity, commonly 30 years, we model borrowers’ problem as a finite-horizon dynamic programming problem. Also, once a borrower defaults on a loan, there is no further decision to be made and no further flow of utility starting from the next period. We propose an estimation method for borrowers’ dynamic problem that takes advantage of the presence of this terminal action.

Estimation of borrowers’ dynamic programming problem using a full solution nested fixed point algorithm (Rust, 1994) would be computationally burdensome given the high-dimensional state space and the large number of borrowers in our data. We instead estimate our model using a variant of the multi-step estimation methods such as Hotz and Miller (1993), Bajari, Benkard, and Levin (2007) and Pesendorfer and Schmidt-Dengler (2008). In the first step, we recover decision rules as a flexible function of state variables, which include home prices, the amount of net equity in the house, monthly payments due, and borrower and loan characteristics such as credit scores. Because the problem has a finite horizon, the optimal policy function should depend on time to maturity and our estimation allows for such dependence. The richness and large size of our dataset allow us to be very flexible in our specification of the decision rules. As part of the first step, we also use standard time-series econometric techniques to estimate the laws of motion governing the evolution of these state variables. In our modeling, we take a partial-equilibrium approach by treating the evolution of macroeconomic state variables, such as the change in home prices and unemployment rate in each geographic market, as an exogenous process.

The second step of the estimation exploits our ability to directly recover the choice-specific value functions from the observed choice probabilities by means of the Hotz-Miller inversion (1993). The presence of a terminal action with a known continuation payoff (namely, zero) allows us to recover the
actual level of choice-specific value associated with each action.\footnote{In a typical problem, the Hotz-Miller inversion only identifies the difference in choice-specific values between two alternatives, and one needs to solve for a functional equation in order to recover the levels of choice-specific values (e.g., see Bajari, Chernozhukov, Hong and Nekipelov, 2009). In our application, the existence of a terminal action allows us to recover the levels of choice-specific values directly from the Hotz-Miller inversion.} We then employ a series estimator to construct the \textit{ex ante} value function conditional on prepaying and paying, respectively, as a flexible function of current state variables.

In the third step, our proposed estimator treats the constructed \textit{ex ante} values as well as state variables in the period utility function as regressors and allows us to recover structural parameters using simple OLS or SUR.

This estimation method is intuitive and extremely easy to implement. It also addresses one of the key data challenges we face. Because subprime mortgages are a relatively new product that was only introduced in recent years, we lack observations for loans close to maturity. Thus, we cannot recover decision rules for loans close to maturity in our first step, which would pose a significant problem if we were to use the typical forward simulation approach of Bajari, Benkard and Levin. However, our proposed estimation method does not suffer from this problem as it requires only one period ahead forward simulation. Another advantage of our estimation method is that it makes identification of the discount factor very clear. In our model, time to maturity influences \textit{ex ante} value functions but does not have any impact on period utility. This allows the discount factor to be identified in our model (see Magnac and Thesmar, 2002, for discussions on the topic). We prove identification of the discount factor in our model and present our estimate of the discount factor along with estimates of other structural parameters.

Once we estimate our model, we use our model to understand importance of various drivers of default as well as potential effects of mortgage policies. In doing so, we use the first-step policy function estimates to simulate borrowers’ behavior under various “counterfactual” regimes. This approach differs from the more common approach, which is to compute the counterfactual outcomes by re-solving for the optimal behavior using estimates of the structural parameters from the last step. The usual argument for re-solving for the optimal behavior is to address the Lucas critique, which applies whenever a proposed shock would change the equilibrium behavior such that there is a new reduced-form relationship between state variables and the policy function. We argue that the panel structure of our data allows us to avoid the Lucas critique as long as we judiciously choose which scenarios to consider. Because the panel structure allows us to identify the policy function over a very wide range of state variables, so long as we limit ourselves to studying policy interventions that do not completely go “out of sample” and do not change the state transition function, the new optimal behavior is correctly captured by the reduced-form policy function. In other words, if a scenario involves state-variable realizations that are actually observed...
for a subset of borrowers in the data, then the reduced-form policy function is still a valid description of how the borrowers behave in that scenario. This approach is similar to within-sample comparative statics and is computationally much lighter than re-solving the dynamic programming problem for each borrower. The downside of this approach is that, as we might expect, there are certain counterfactuals that cannot be analyzed under this approach, namely any scenarios involving transitions to states not spanned by the estimation sample or resulting in changes to the state transitions themselves. In this paper, we illustrate that there are many interesting scenarios that can be fruitfully explored without re-solving for the new optimal behavior.

Our estimation exploits a rich dataset from LoanPerformance, which covers the majority of subprime and Alt-A mortgages\(^3\) that were securitized between 2000 and 2007. The unit of observation is an individual mortgage observed at a point in time. For each loan, we observe information from the borrower’s loan application, including the terms of the contract, the appraised value of the property, the loan-to-value (LTV) ratio, the level of documentation, and the borrower’s credit 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 or is prepaid. To track movements in home prices, we merge the LoanPerformance data with zip code-level home price indices, also from LoanPerformance.

This paper contributes to the literature by estimating a fully dynamic model of borrower behavior. It is clear that the decision to default is a dynamic decision given the irreversibility of default and the resulting option value of delaying default. Not correctly capturing such dynamic features may lead to inconsistent estimates and generate misleading welfare implications. Most of existing empirical work on loan default uses a duration framework (Deng, Quigley and van Order, 2000; Foster and Van Order, 1984), which does not explicitly address the dynamic features of borrowers’ decisions as expectations do not play a role in the duration framework. The paper also makes a contribution by demonstrating that a certain set of interesting counterfactuals can be successfully performed using first-step reduced-form policy function estimates. This is an important point that may apply to many empirical problems, because re-computing the equilibrium is often computationally burdensome while performing simulations using the first-step estimates is much easier. We also make a methodological contribution by proving identification of the discount factor in our model. Finally, our paper informs our understanding of borrowers’ default incentives and evaluates welfare effects of key policy tools, a topic of immense interest to policymakers.

The rest of this paper proceeds as follows. In Section 2, we present a dynamic model of borrower default on mortgage loans. Section 3 discusses identification of the model’s primitives including the discount factor. Section 4 discusses our estimation methodology. In Section 5, we describe the data.

\(^3\)Alt-A’s are a type of mortgage that is riskier than prime but less risky than subprime. In this paper, we casually use the term “subprime market” to refer to both subprime and Alt-A mortgages.
Section 6 presents estimation results and Section 7 discusses our simulation exercises. Section 8 concludes the paper.

2 Model

We formulate borrowers’ decisions using a dynamic, discrete-time model. Each agent enters a mortgage contract lasting \( T \) time periods, and solves a dynamic programming problem with a finite time horizon ending at \( T \). We assume that there is no interaction among borrowers, which implies that our setup is a single-agent model. The components of the model are as follows.

2.1 Actions

At each time period \( t \) over the life of borrower \( i \)’s loan, the borrower chooses an action \( a_{i,t} \) from the finite set \( A = \{0, 1, 2\} \).\(^4\) The possible actions in \( A \) are to default \( (a_{i,t} = 0) \), to prepay the mortgage \( (a_{i,t} = 1) \), or to make just the regularly scheduled payment for the current time period \( (a_{i,t} = 2) \). Default is a terminal action: once a borrower defaults, there is no further decision to be made and no further flow of utility starting from the next period.

2.2 Period Utility and State Transition

We assume that each borrower observes a vector of state variables \( s_{i,t} \in S \) in each period. The support of the state variable \( S \) is product space which is a subset of \( k \)-dimensional Euclidean space (we allow the subspaces of this product space to be either continuous or discrete). The state vector \( s_{i,t} \), fully observed by the econometrician, includes various factors that directly enter the period utility of paying or prepaying relative defaulting, such as borrower \( i \)’s FICO score, current home value, monthly payments, etc. We also assume that the borrower is characterized by a time-invariant vector of characteristics \( c_i \in C \) (borrower type, observed by the econometrician) and a time-dependent vector of idiosyncratic shocks \( \varepsilon_{i,t} = (\varepsilon_{i,0,t}, \varepsilon_{i,1,t}, \varepsilon_{i,2,t}) \) (unobserved by the econometrician). The set \( C \) is assumed to be finite. In our empirical application, \( c_i \) will capture whether the borrower is an investor or an owner-occupier of the house. Each element of \( \varepsilon_{i,t} \) is assumed to have a continuous support on the real line. We make the following assumption regarding the marginal distributions of random variables.

\(^4\)Note that we use \( t \) to denote the loan’s age, not calendar time. A 36-month old loan will have \( t = 36 \) whether the loan was originated in January 2003 or October 2007. In our estimation, we limit our attention to all loans with the same maturity (30 years) and can therefore think of \( t \) as also representing the time to maturity.
ASSUMPTION 1

1. **Conditional independence of the idiosyncratic shocks:**

   \[ s_{i,t} \perp \varepsilon_{i,t} | c_i \]

2. **Decomposition of the borrower’s heterogeneity:**

   \[ \varepsilon_{i,t} | \varepsilon_{i,t-1}, a_{i,t-1}, c_i \sim \varepsilon_{i,t} | a_{i,t-1}, c_i. \]

3. **Exclusion restriction (c_i cannot be represented as a linear combination of the elements of s_{i,t}):** \( C \) does not belong to any proper linear subspace of \( S \).

4. **Markov transition of the state variables:** \( s_{i,t} \) follows a first-order reversible Markov process, conditional on \( a_{i,t} \).

In our empirical analysis we use a conventional normalization of the distribution of the idiosyncratic shocks, assuming that components of \( \varepsilon_{i,t} \) are mutually independent, have a type I extreme value distribution, and are iid across borrowers and over time. However, this normalization is not essential to establish the existence and identification results in this paper.

The Markov process driving the state variables is assumed to have a transition density \( g(s_{i,t+1}|a_{i,t}, s_{i,t}) \). Most state variables evolve according to an exogenous process, and their distribution at time \( t+1 \) does not depend upon the borrower’s action, \( a_{i,t} \). For instance, the credit score and location of each borrower is constant over time, while home price growth rates are assumed to follow AR(2) process. The only state variable whose transition is influenced by \( a_{i,t} \) is the monthly payment. When borrower \( i \) prepays in period \( t \), we assume that the borrower refinances into a new loan that matures at the same time as the old loan\(^5\) and whose interest rate is equal to the current market interest rate.\(^6\) Thus, the payment level will depend upon the borrower’s choice. For the rest of this paper, we will use \( g(s_{i,t+1}|s_{i,t}, a_{i,t}) \) to denote the state transitions, with the implicit understanding that transition of most state variables is not affected by \( a_{i,t} \).

---

\(^5\)For example, if the borrower refinances when the loan is 50 months old, we assume that the new loan will mature in \( T - 50 \) months.

\(^6\)We assume that the market interest rate available to borrower \( i \) at time \( t \) is equal to \( r^{m}_{i,t} = r^{m}_{i}(z_{i,t}) + \xi_{i} \), where \( r^{m}_{i}(z_{i,t}) \) is the prevailing rate available for loans with observable characteristics \( z_{i,t} \), and \( \xi_{i} \) is a borrower-specific spread that is constant over time. For a given borrower, we can identify \( \xi_{i} \) as the residual from regressing the observed interest rate on the observed characteristics of the original loan.
We assume that the per-period utility of the borrower is separable in the idiosyncratic component. We can characterize the borrower’s utility as:

\[ U(a_{i,t}, s_{i,t}; c_i) = u(a_{i,t}, s_{i,t}; c_i) + \varepsilon_{i,a_{i,t},t} \]

We assume that the per-period utility of the borrower has a deterministic component that is a time-invariant function of the state, action, and the borrower’s “type” \( u(a_{i,t}, s_{i,t}; c_i) \). Provided the cardinality of the utility function, the observable actions can only identify the relative utility level. As a result, we can normalize the utility from one of the actions to zero. We chose this action to be the default decision: \( u(a_{i,t} = 0, s_{i,t}; c_i) = 0 \). If there are state variables that increase the costs of default (i.e., decrease the utility of default), they will be interpreted in our model as increasing the utility of pay or prepay relative to default, so this normalization is innocuous.

### 2.3 Decision Rule and Value Function

We consider the borrower’s problem as an optimal stopping problem. We assume that the default decision is not reversible and the borrower cannot “re-start” the borrowing within his foreseeable future.\(^7\) Provided that the default (“stopping”) decision is non-reversible, and given that we normalize the utility from the default decision to zero, the borrower receives a perpetual payoff of zero after defaulting. If the borrower pays or prepays his mortgage, he receives a payoff corresponding to the expected discounted stream of utility \( U(a_{i,t}, s_{i,t}; c_i) \) towards the mortgage maturity (that characterizes the combination of utility from consumption of housing services and disutility from payments for the mortgage) and then receives a perpetual stream of “housing services” once the mortgage was paid off. Parameter \( \beta \) is the discount factor that characterizes the time-impatience of the borrower.

Borrower’s decision rule \( D \) is a mapping from the vector of observable payoff-relevant variables into actions, \( D : S \times C \times \mathbb{R}^3 \mapsto A \). Provided that the draws of the payoff shock \( \varepsilon_{i,t} \) are idiosyncratic (and, thus, non-anticipating), the borrower will be concerned with their future expected decisions. We denote the expected decision \( \sigma_{k,t}(s_{i,t}, c_i) = E \left[ 1\{D_t(s_{i,t}, c_i, \varepsilon_{i,t}) = k\} \middle| s_{i,t}, c_i \right] \) for \( k \in A \) and call \( \sigma \) the policy function where \( \sigma = (\sigma_1(s_{i,1}, c_i), \ldots, \sigma_T(s_{i,T}, c_i)) \) and \( \sigma_t(s_{i,t}, c_i) = (\sigma_{0,t}(s_{i,t}, c_i), \sigma_{1,t}(s_{i,t}, c_i), \sigma_{2,t}(s_{i,t}, c_i)) \).

Considering the expected discounted utility of the borrower who has not defaulted prior to period

\(^7\)We explicitly control for that in our empirical analysis by controlling for the mortgages that return to the “normal” state after a period of delinquency.
$t < T$, we introduce the \textit{ex ante} value function:

$$V_{t, \sigma}(s_{i,t}; c_i) = E_{\sigma, g(s)} \left[ \sum_{\tau=t}^{T} \left( \beta^{\tau-t} U(a_{i,\tau}, s_{i,\tau}; c_i) \prod_{\tau_1=1}^{\tau-1} 1(a_{i,\tau_1} > 0) \right) \right] | s_{i,t}$$

In the above equation, $\sigma$ is the policy function. The term $\prod_{\tau_1=1}^{\tau-1} 1(a_{i,\tau_1} > 0)$ captures the fact that once a borrower defaults, there is no further flow of utility starting from the next period.

The choice-specific value function by $V_{t, \sigma}(a_{i,t} = k, s_{i,t}; c_i)$ signifies the deterministic component of the discounted payoff in period $t$ that the borrower expects to receive when choosing action $k$:

$$V_{t, \sigma}(a_{i,t} = k, s_{i,t}; c_i) = u(a_{i,t} = k, s_{i,t}; c_i) + \beta E[V_{t+1, \sigma}(s_{i,t+1}; c_i) | s_{i,t}, a_{i,t} = k].$$

Because the period utility of default is normalized to zero ($u(a_{i,t} = 0, s_{i,t}; c_i) = 0$) and default is a terminal action, the choice-specific value of default is zero, i.e., $V_{t, \sigma}(a_{i,t} = 0, s_{i,t}; c_i) = 0$ for all $t \leq T$.

For the final period $T$, the choice-specific values for the remaining actions (prepay or pay) are equal to the deterministic component of the period payoff $u(a_{i,T}, s_{i,T}; c_i)$ plus the house value in period $T$, as the borrower obtains full ownership of the house once the loan matures.\(^8\)

### 2.4 Optimal Policy Functions

We set up the decision problem of the borrower as the optimal stopping problem with the finite time horizon with discrete time. In the following theorem we establish a formal existence and uniqueness result characterizing the borrower’s optimal decision in our model.

**Theorem 1** Under Assumption 1 there exists a unique decision rule in the cutoff form $D^*_t(s_{i,t}, c_i, \varepsilon_{i,t})$ supported on $A$ for $t = 1, 2, \ldots, T$ that solves the maximization problem

$$\sup_{(D_1, D_2, \ldots, D_T) \in A^T} V_{1, \sigma}(s_{i,1}; c_i).$$

**Proof.** Our argument will follow directly by the backward induction. In fact, we note that in the last period (at the mortgage maturity) the borrower faces a static optimization problem to choose between $\varepsilon_{i,0,T}$, $V_T(1, s_{i,T}; c_i) + \varepsilon_{i,1,T}$, and $V_T(2, s_{i,T}; c_i) + \varepsilon_{i,2,T}$, and $V_T(3, s_{i,T}; c_i) + \varepsilon_{i,3,T}$. The optimal decision is to make the choice that delivers the highest payoff. As a result, the decision rule corresponds to $D^*_T(s_{i,T}, \varepsilon_{i,T}; c_i) = \ldots$

\(^8\)We assume that the house value in period $T$ is equivalent to perpetual stream of housing services from $T + 1$ onward.
arg \max_{k \in A} \{ V_T(1, s_{i,T}; c_i) + \varepsilon_{i,k,T} \}. Provided that the payoff shocks are idiosyncratic and have a continuous distribution that we normalize, the optimal choice probability is characterized by a continuous function of \( (V_T(k, s_{i,T}; c_i), k \in A) \). Knowing the probabilities of actions in period \( T \), we can obtain the choice-specific value function in the period \( T - 1 \) as

\[
V_{T-1}(k, s_{i,T-1}; c_i) = u(k, s_{i,T-1}; c_i) + \beta E \left[ \sum_{k' \in A} 1\{D^*_T = k'\} (V_T(k', s_{i,T}; c_i) + \varepsilon_{i,k',T} \mid s_{i,T-1}, c_i) \right].
\]

Provided that the optimal decision in the final period \( T \) is already found, the optimal decision problem in period \( T - 1 \) becomes a static problem of choice between three alternatives. Its solution, again, trivially exists and is (almost surely) unique because the distribution of \( \varepsilon_{i,T-1} \) is continuous. \( \blacksquare \)

Q.E.D.

If one prefers to normalize the distribution of idiosyncratic shocks to be the standard type I extreme value distribution, then it becomes possible to provide a closed-form expression for the probabilities of default, prepay, and pay at a given state and any period \( t \leq T \) as:

\[
\begin{align*}
\sigma_{0,t}(s_{i,t}, c_i) &= \Pr(D^*_t = 0 \mid s_{i,t}, c_i) = \frac{\exp(V_t(a_{i,t} = 0, s_{i,t}; c_i))}{\sum_{k=0}^{2} \exp(V_t(a_{i,t} = k, s_{i,t}; c_i))} \\
\sigma_{1,t}(s_{i,t}, c_i) &= \Pr(D^*_t = 1 \mid s_{i,t}, c_i) = \frac{\exp(V_t(a_{i,t} = 1, s_{i,t}; c_i))}{\sum_{k=0}^{2} \exp(V_t(a_{i,t} = k, s_{i,t}; c_i))} \\
\sigma_{2,t}(s_{i,t}, c_i) &= \Pr(D^*_t = 2 \mid s_{i,t}, c_i) = \frac{\exp(V_t(a_{i,t} = 2, s_{i,t}; c_i))}{\sum_{k=0}^{2} \exp(V_t(a_{i,t} = k, s_{i,t}; c_i))}
\end{align*}
\]

Thus, we explicitly derived the optimal cutoff decision rule and characterized the choice-specific value functions corresponding to the optimal decision rule.

The normalization of the distribution of the idiosyncratic shocks to type I extreme value distribution allows us to provide explicit expressions for the gaps in choice-specific values between different options in terms of the optimal choice probabilities. Assuming that from now on we consider the value functions and choice probabilities that correspond to the optimal decision rule, we omit the subscript \( \sigma \). The Hotz-Miller inversion represents the \textit{difference} in the choice-specific values as follows.

\[
\begin{align*}
\log \left( \frac{\sigma_{1,t}(a_{i,t}, c_i)}{\sigma_{0,t}(a_{i,t}, c_i)} \right) &= V_t(a_{i,t} = 1, s_{i,t}; c_i) - V_t(a_{i,t} = 0, s_{i,t}; c_i) \\
\log \left( \frac{\sigma_{2,t}(a_{i,t}, c_i)}{\sigma_{0,t}(a_{i,t}, c_i)} \right) &= V_t(a_{i,t} = 2, s_{i,t}; c_i) - V_t(a_{i,t} = 0, s_{i,t}; c_i)
\end{align*}
\]

In our model, we can identify the choice-specific values themselves because default is a terminal action, pegging the choice-specific value of default to 0: that is, \( V_t(a_{i,t} = 0, s_{i,t}; c_i) = 0 \). Thus we can recover the
**ex ante** value function, \( V_t(s_i;c_i) \), directly from the data. Our estimation method, discussed further in this paper, exploits this feature of our model.

### 3 (Semiparametric) Identification

In this section we demonstrate that the borrower’s optimal decision model is identified. Our model contains two structural elements that determine the payoff of the borrower. First is the deterministic component of the per period payoff function \( u(\cdot,\cdot,\cdot) \). Second is the time preference parameter \( \beta \). We will aim at providing the argument for the nonparametric identification of \( u(\cdot,\cdot,\cdot) \) as well as the identification of the time preference parameter. We emphasize that our identification result *does not rely on the extreme value assumption regarding the distribution of idiosyncratic shocks*. We denote the heterogeneity-specific joint cdf of the distribution of payoff shocks \( F_\varepsilon(\cdot|c_i) \).

Our identification argument proceeds in the conventional way by providing the mapping between the distribution of the data and the structural components of the model. We assume that the observed data distribution characterizes the joint probability distribution of actions, states, and borrower-level heterogeneity components.

**Theorem 2** Suppose that the distribution of idiosyncratic shocks conditional on the borrower-specific heterogeneity variables \( c_i \), \( F_\varepsilon(\cdot|c) \), has a full support with the density strictly positive on \( \mathbb{R}^3 \). Also suppose that for at least two periods \( t \) and \( t' \sigma_{k,t}(\cdot;\cdot) \neq \sigma_{k,t'}(\cdot;\cdot) \) for \( k \in A \).

1. If the data distribution contains the information on at least two periods and the discount factor \( \beta \) is fixed, then the per period utility is nonparametrically identified.
2. If the data distribution contains the information on at least three periods, then both the discount factor and the per period utility function are identified.

**Proof.** Using the joint cdf of the idiosyncratic payoff shocks, we introduce the following functions:

\[
\sigma_0(z_1,z_2; c_i) = \int \left\{ z_1 + \varepsilon_{i,0,t} \geq z_2 + \varepsilon_{i,1,t}, z_1 + \varepsilon_{i,0,t} \geq z_2 + \varepsilon_{i,2,t} \right\} F_\varepsilon(\varepsilon|c_i)
\]

\[
\sigma_1(z_1,z_2; c_i) = \int \left\{ z_1 + \varepsilon_{i,1,t} \geq z_2 + \varepsilon_{i,0,t}, z_1 + \varepsilon_{i,1,t} \geq z_2 + \varepsilon_{i,2,t} \right\} F_\varepsilon(\varepsilon|c_i)
\]

\[
\sigma_2(z_1,z_2; c_i) = \int \left\{ z_2 + \varepsilon_{i,2,t} \geq z_1 + \varepsilon_{i,0,t}, z_2 + \varepsilon_{i,2,t} \geq z_1 + \varepsilon_{i,1,t} \right\} F_\varepsilon(\varepsilon|c_i)
\]
and function

\[ \nu(z_1, z_2; c_i) = \int \left( z_1 \mathbf{1}_{z_1 + \varepsilon_1,0,t \geq z_1 + \varepsilon_i,1,t} + z_2 \mathbf{1}_{z_2 + \varepsilon_i,2,t \geq z_2 + \varepsilon_i,1,t} \right) F_e(\varepsilon | c_i). \]

We note that the introduced functions are known (given that we can normalize the distribution of the idiosyncratic shocks) and are monotone and differentiable in their arguments. Our result will be based on the following technical lemma.

**Lemma 1**

Under our assumptions, the system of equations

\[ \sigma_0(z_1, z_2; c_i) = \bar{\sigma}_0(c_i), \]
\[ \sigma_1(z_1, z_2; c_i) = \bar{\sigma}_1(c_i), \]

has a unique solution for each \( c_i \) if and only if \( \bar{\sigma}_0(c_i) + \bar{\sigma}_1(c_i) < 1 \).

**Proof:**

Consider partial derivatives

\[ \frac{\partial \sigma_0(z_1, z_2)}{\partial z_1} = -\int_{-\infty}^{\infty} \frac{\partial^2 F_e}{\partial \varepsilon_0 \partial \varepsilon_1} (\varepsilon_0, \varepsilon_0 - z_1, \varepsilon_0 - z_2) d\varepsilon_0, \quad \frac{\partial \sigma_0(z_1, z_2)}{\partial z_2} = -\int_{-\infty}^{\infty} \frac{\partial^2 F_e}{\partial \varepsilon_0 \partial \varepsilon_2} (\varepsilon_0, \varepsilon_0 - z_1, \varepsilon_0 - z_2) d\varepsilon_0. \]

Similarly, we can find that

\[ \frac{\partial \sigma_1(z_1, z_2)}{\partial z_1} = \int_{-\infty}^{\infty} \frac{\partial^2 F_e}{\partial \varepsilon_1 \partial \varepsilon_1} (z_1 + \varepsilon_1, \varepsilon_1, z_1 + \varepsilon_1, z_2) d\varepsilon_1 + \int_{-\infty}^{\infty} \frac{\partial^2 F_e}{\partial \varepsilon_1 \partial \varepsilon_2} (z_1 + \varepsilon_1, \varepsilon_1, z_1 + \varepsilon_1, z_2 + \varepsilon_1) d\varepsilon_1 = \int_{-\infty}^{\infty} \left( \frac{\partial^2 F_e}{\partial \varepsilon_0 \partial \varepsilon_1} + \frac{\partial^2 F_e}{\partial \varepsilon_1 \partial \varepsilon_2} \right) (\varepsilon_0, \varepsilon_0 - z_1, \varepsilon_0 - z_2) d\varepsilon_0, \]

and

\[ \frac{\partial \sigma_1(z_1, z_2)}{\partial z_2} = -\int_{-\infty}^{\infty} \frac{\partial^2 F_e}{\partial \varepsilon_1 \partial \varepsilon_2} (\varepsilon_0, \varepsilon_0 - z_1, \varepsilon_0 - z_2) d\varepsilon_0. \]

We assumed that the joint distribution of errors has a continuous density with a full support on \( \mathbb{R}^3 \). Provided that \( \frac{\partial \sigma_0(z_1, z_2)}{\partial z_1} \frac{\partial \sigma_0(z_1, z_2)}{\partial z_2} > 0 \) the mapping \( z_1 \mapsto z_2 \) implicitly defined by equation \( \sigma_0(z_1, z_2; c_i) = \bar{\sigma}_0(c_i) \) is invertible for any \( c_i \). Moreover, if we denote this mapping \( z_2 = m_0(z_1, \bar{\sigma}_0(c_i); c_i) \), then using the
result regarding the derivative of the inverse function, we can conclude that

\[
\frac{\partial m_0(z_1, \bar{\sigma}_0(c_i); c_i)}{\partial z_1} \leq 0.
\]

Similarly, we can define a map \( z_1 = m_1(z_2, \bar{\sigma}_1(c_i); c_i) \), then using the result regarding the derivative of the inverse function, we can conclude that

\[
\frac{\partial m_1(z_2, \bar{\sigma}_0(c_i); c_i)}{\partial z_1} \geq 0.
\]

We can explore the asymptotic behavior of both maps. Consider \( m_0 \) first. Suppose that \( z_1 \to -\infty \). Then

\[
\lim_{z_1 \to -\infty} m_0(z_1, \bar{\sigma}_0(c_i); c_i) = z_2^*, \text{ where } z_2^* \text{ solves } \int 1\{\varepsilon_{i,0,t} \geq z_2^* + \varepsilon_{i,2,t}\} F_{c}(dc \mid c_i) = \bar{\sigma}_0(c_i). \text{ Also let }
\]

\[
z_1^* \text{ solve } \int 1\{\varepsilon_{i,0,t} \geq z_1^* + \varepsilon_{i,1,t}\} F_{c}(dc \mid c_i) = \bar{\sigma}_0(c_i). \text{ Then } \lim_{z_1 \to z_1^*-0} m_0(z_1, \bar{\sigma}_0(c_i); c_i) = -\infty.
\]

Next consider \( m_1 \). Suppose that \( z_1^{**} \) is the solution of \( \int 1\{\varepsilon_{i,1,t} \geq z_1^{**} + \varepsilon_{i,2,t}\} F_{c}(dc \mid c_i) = \bar{\sigma}_1(c_i) \). Then as \( z_1 \to +\infty \), the map approaches asymptotically to the line: \( m_1(z_2, \bar{\sigma}_1(c_i); c_i) \to z_1 + z_2^* \). Suppose that \( z_1^{**} \) is the solution of \( \int 1\{z_1^{**} + \varepsilon_{i,1,t} \geq \varepsilon_{i,0,t}\} F_{c}(dc \mid c_i) = \bar{\sigma}_1(c_i) \). Then \( \lim_{z_2 \to -\infty} m_1(z_2, \bar{\sigma}_1(c_i); c_i) = z_1^{**} \).

Thus \( m_1 \) is a continuous strictly decreasing mapping from \((-\infty, z_1^{**})\) into \((-\infty, z_2^*)\) and \( m_1 \) is a continuous strictly increasing mapping from \([z_1^{**}, +\infty)\) into the real line.

Provided that both curves are continuous and monotone, they intersect if and only if their projections on \( z_1 \) and \( z_2 \) axes overlap. The projections on the \( z_2 \) axis are guaranteed to overlap ((\(-\infty, z_2^*) \subset \mathbb{R})). The projections on the \( z_1 \) axis will overlap if and only if \( z_1^{**} < z_1^* \). Given that function \( \sigma(z) = \int 1\{\varepsilon_{i,0,t} - \varepsilon_{i,1,t} \leq z\} F_{c}(dc \mid c_i) \) is strictly monotone in \( z \), then \( z_1^{**} < z_1^* \) if and only if \( \sigma_0(c_i) + \sigma_1(c_i) < 1 \).

This proves the statement of Lemma 1.

Next we can show that the model is nonparametrically identified. The observed probability distribution characterizes the conditional choice probabilities \( \{\sigma_{k,t}(\cdot; \cdot), k \in A\} \). Given the structure of the optimal solution, there is a direct link between the choice-specific value functions in period \( t \) and the choice probability which is expressed through the distribution of the idiosyncratic payoff shocks. In particular, for each \( s \in S \) and \( c \in C \) and each \( t \leq T \), we can write the system of identifying equations

\[
\begin{align*}
\sigma_0(V_t(1, s; c), V_t(2, s; c) ; c) &= \sigma_{0,t}(s; c) \\
\sigma_1(V_t(1, s; c), V_t(2, s; c) ; c) &= \sigma_{1,t}(s; c)
\end{align*}
\]

Given the result of Lemma 1, we can solve for the functions \( V_t(1, s; c) \) and \( V_t(2, s; c) \) over \( S \) and \( C \).

The conditional distribution \( s_{i,t+1} \mid s_{i,t} a_{i,t}, c_i \) is observable. As a result, for each \( k \in \{1, 2\} \) and each
We can consider the system of equations in periods $t$ and $t' \leq T$:

$$V_t(1, s; c) = u(1, s; c) + \beta E \left[ \nu(V_{t+1}(1, s'; c), V_{t+1}(2, s'; c)) \mid a_{i,t} = 1, s, c \right]$$

$$V_t(2, s; c) = u(2, s; c) + \beta E \left[ \nu(V_{t+1}(1, s'; c), V_{t+1}(2, s'; c)) \mid a_{i,t} = 2, s, c \right]$$

$$V_{t'}(1, s; c) = u(1, s; c) + \beta E \left[ \nu(V_{t'+1}(1, s'; c), V_{t'+1}(2, s'; c)) \mid a_{i,t'} = 1, s, c \right]$$

This is the system of three linear equations with three unknowns $u(1, s; c)$, $u(2, s; c)$, and $\beta$. We note that to set up this system of equations, we need to have observations for at least three periods.

Then simple differencing solves for the discount factor:

$$\beta = \frac{V_t(1, s; c) - V_{t'}(1, s; c)}{E \left[ \nu(V_{t+1}(1, s'; c), V_{t+1}(2, s'; c)) \mid a_{i,t} = 1, s, c \right] - E \left[ \nu(V_{t'+1}(1, s'; c), V_{t'+1}(2, s'; c)) \mid a_{i,t} = 1, s, c \right]}.$$ 

The denominator in this expression is not equal to zero because of the assumption of the theorem that $\sigma_{k,t}(s; c) \neq \sigma_{k,t'}(s; c)$ for at least two periods. We also can recover the per period utility function as

$$u(1, s; c) = \left( E \left[ \nu(V_{t+1}(1, s'; c), V_{t+1}(2, s'; c)) \mid a_{i,t} = 1, s, c \right] - E \left[ \nu(V_{t'+1}(1, s'; c), V_{t'+1}(2, s'; c)) \mid a_{i,t} = 1, s, c \right] \right)^{-1} \times \left( V_t(1, s; c)E \left[ \nu(V_{t'+1}(1, s'; c), V_{t'+1}(2, s'; c)) \mid a_{i,t} = 1, s, c \right] - V_{t'}(1, s; c)E \left[ \nu(V_{t'+1}(1, s'; c), V_{t'+1}(2, s'; c)) \mid a_{i,t} = 1, s, c \right] \right).$$

Similarly, we can explicitly recover the utility $u(2, s; c)$ for any $s$ and $c$ in their support.

We note that if the discount factor is fixed, then we can identify the utility function from just two periods.

This theorem establishes a general result that the considered model is identified (which includes the identification of the time preference parameter). It relies on the testable assumption that for two (possibly consecutive) periods the observed choice probability function is different. The theoretical justification for this comes from the intrinsically finite horizon decision where towards the mortgage maturity the borrowers would be less inclined to choose the default option. The validity of this assumption can be directly verified from the data and we do that in our empirical analysis by formally testing the differences in the estimated policy functions across the time periods.

We can provide very simple expressions that show how one can recover the structural parameters of the model from the observed policy functions using the example where the payoff shocks are iid with
discount factor requires the presence of a variable that does not affect the period utility but affects of time preferences (Magnac and Thesmar, 2002; Fang and Wang, 2010) shows that identification of the proposition. Since there is a big reward at the time of loan maturity (house ownership), the probability of a diminishing incentive to default as the loan approaches maturity, satisfying the assumption of the discount factor is identified.

By the assumption of Theorem 2, the denominator of this expression is not equal to zero. As a result, using the transition function recovered from the data. However, the series estimator is computationally lighter, and given the large number of observations in our data set, the series estimation should be very precise as well.

We can nonparametrically recover $\sigma_{0,t}(s_{i,t})$, $\sigma_{1,t}(s_{i,t})$ and $\sigma_{2,t}(s_{i,t})$ from the data and we can also construct $E \left[ \log \left( \sigma_{0,t+1}(s_{i,t+1})^{-1} \right) | s_{i,t}, a_{i,t} = k \right]$ by projecting the recovered ex ante values $V_{t+1}(s_{i,t+1})$ onto basis functions of today’s state variables $s_{i,t}$. Thus, if the discount factor is known, we can directly recover the period utility nonparametrically from the data. However, we can make an even stronger statement as Theorem 2 states. We can also provide an explicit expression for the discount factor that can be obtained using a three-period panel of borrowers. For each action we can write

\[
\begin{align*}
u(a_{i,t} = k, s_{i,t} = \bar{s}) &= \log \left( \frac{\sigma_{k,t}(\bar{s})}{\sigma_{0,t}(\bar{s})} \right) - \beta E \left[ \log \left( \sigma_{0,t+1}(s_{i,t+1})^{-1} \right) | s_{i,t} = \bar{s}, a_{i,t} = k \right], \\
u(a_{i,t+1} = k, s_{i,t+1} = \bar{s}) &= \log \left( \frac{\sigma_{k,t+1}(\bar{s})}{\sigma_{0,t+1}(\bar{s})} \right) - \beta E \left[ \log \left( \sigma_{0,t+2}(s_{i,t+2})^{-1} \right) | s_{i,t+1} = \bar{s}, a_{i,t+1} = k \right]
\end{align*}
\]

Recalling that the period utility function does not depend on time, we can take the difference between the two equations and express

\[
\beta = \frac{\log \left( \frac{\sigma_{k,t+1}(\bar{s})}{\sigma_{0,t+1}(\bar{s})} \right) \sigma_{0,t}(\bar{s})}{E \left[ \log \left( \sigma_{0,t+1}(s_{i,t+1})^{-1} \right) | s_{i,t} = \bar{s}, a_{i,t} = k \right] - E \left[ \log \left( \sigma_{0,t+2}(s_{i,t+2})^{-1} \right) | s_{i,t+1} = \bar{s}, a_{i,t+1} = k \right]}
\]

By the assumption of Theorem 2, the denominator of this expression is not equal to zero. As a result, the discount factor is identified.

In general, it can be shown that for reasonable values of the remaining parameters, our model implies a diminishing incentive to default as the loan approaches maturity, satisfying the assumption of the proposition. Since there is a big reward at the time of loan maturity (house ownership), the probability of default decreases over time even if the state variables $s$ do not change. The prior literature on identification of time preferences (Magnac and Thesmar, 2002; Fang and Wang, 2010) shows that identification of the discount factor requires the presence of a variable that does not affect the period utility but affects.

\[\text{Conceptually, this series estimator is similar to simulating one period forward to compute } E \left[ V_{t+1}(s_{i,t+1}) | s_{i,t}, a_{i,t} = k \right] \text{ using the transition function recovered from the data. However, the series estimator is computationally lighter, and given the large number of observations in our data set, the series estimation should be very precise as well.}\]
state transitions or *ex ante* values. In our model, we can identify the discount factor even if all state variables are constant over time because time to maturity plays the role of excluded variable in our model due to finite horizon: Time to maturity influences *ex ante* value but does not have any impact on period utility. The two features that are important for identification of the discount factor in our setup are the finite horizon and the presence of a terminal choice.

4 Empirical Methodology

4.1 Estimation

We consider a semiparametric plug-in estimator for the decision model. The plug-in estimator will be constructed in three steps. In the first step, we will recover choice probabilities nonparametrically using a series estimator or logit sieve. In the second step, we will construct *ex ante* value functions using a series estimator. In the third step, we will recover structural parameters along with the discount factor by regressing the recovered choice specific values on the constructed *ex ante* value and state variables in the per period payoff.

**Step 1** First, we use the observed data on the mortgage decisions to estimate the choice probabilities nonparametrically. Suppose that the dataset has $I(t)$ consumers whose loans have $t$ months remaining until maturity. Then, if $d^a_{i,t}$ is a binary indicator of action $a \in \{0, 1, 2\}$, the choice probability can be estimated using the vector of orthogonal polynomials $q^K(\cdot)$ as follows:

$$\hat{\sigma}_{k,t}(s) = q^K(\cdot) \left( \frac{1}{I(t)} \sum_{t=1}^{I(t)} q^K(s_{i,t}) q^{K'}(s_{i,t}) \right)^{-1} \left( \frac{1}{I(t)} \sum_{t=1}^{I(t)} d^a_{i,t} q^K(s_{i,t}) \right)$$

The estimation is done separately for each $t$ since the choice probabilities vary across $t$ due to nonstationarity introduced by finite horizon of the model. The estimation will be also done separately for prepay ($k = 1$) and prepay ($k = 2$) options. The number of series terms will be a function of the total sample size with $K \to \infty$ as $I \to \infty$.\footnote{We provide the detailed conditions further in this section.} As we will see, for our distribution results to be valid (and, thus, the first-step estimation error to have no impact on the convergence rate for the estimated parameters), it is sufficient to find an estimator for the choice probabilities with a uniform convergence rate of at least $(I)^{1/4}$. Such estimators will exist if the choice probability is a sufficiently smooth function of the state.

Using the estimated choice probabilities, we can use the assumption that the unobserved error terms have a logistic distribution and explicitly express the choice-specific value function for the pay and prepay
options as

\[ \hat{V}_t(a_{i,t} = k, s_{i,t}) = \log \left( \frac{\hat{\sigma}_{k,t}(s_{i,t})}{\hat{\sigma}_{0,t}(s_{i,t})} \right). \]

Note that we don’t have to perform any iterations to recover the choice-specific value function in our model. It can be recovered directly from the estimated choice probabilities. Using the estimated probability of default, we can also estimate the ex ante value function as

\[ \hat{V}_t(s_{i,t}) = \log \left( \frac{1}{\hat{\sigma}_{0,t}(s_{i,t})} \right). \]

(4)

The recovered value functions will have the rate of convergence that is equal to the convergence rate of the estimated choice probabilities, under specific support conditions on the state variables.

**Step 2** In the second step, we estimate the expected ex ante value associated with prepay and pay, respectively. This estimate is based on the integration of the ex ante value function with respect to the conditional distribution of the future state draw. This can be done, for instance, by using a series estimator:

\[
\hat{\mathbb{E}} [V_{t+1}(s_{i,t+1}) \mid s_{i,t} = s, a_{i,t} = k] = q^{K'}(s) \left( \frac{1}{I(t)} \sum_{i=1}^{I(t)} q^K(s_{i,t}) q^{K'}(s_{i,t}) \right)^{-1} \frac{1}{I(t)} \sum_{i=1}^{I(t)} \hat{V}_{t+1}(s_{i,t+1})q^K(s_{i,t})
\]

As in the first step, the estimation is done separately for each \( t \).

**Step 3** Considering the parametric specification of the consumer’s per period payoff defined by \( u(a_{i,t} = k, s_{i,t}; \gamma) \), we estimate parameter vector \( \gamma \) and the discount factor. In case where the per period payoffs are defined by the linear indices of the state variables, we estimate the parameters by running a regression of the estimated choice-specific payoff of \( \hat{V}_t(a_{i,t} = k, s_{i,t}) \) on \( \hat{\mathbb{E}} [V_{t+1}(s_{i,t+1}) \mid s_{i,t}, a_{i,t} = k] \) and the state variables in the per period payoff. The coefficient estimates for the state variables will correspond to the parameters \( \gamma \) in the linear index characterizing the per period payoff, and the coefficient of the expected ex ante value function will correspond to the estimated discount factor.

### 4.2 Asymptotic Theory for the Plug-In Estimator

We consider a multi-step sieve estimation procedure. Given that we are analyzing the single-agent dynamic optimization problem, the goal is to characterize the distribution and provide a multi-step estimator such that in the first and second steps we can estimate the policy function and ex ante value function at a sufficiently fast rate and in the third step we can estimate the parameters of the payoff function at a parametric rate. In the semiparametric third step estimation, we denote the parameter
Assumption 2

1. Parameter space \( \Gamma \) is a convex compact set. Profit function \( u_i(a_i, s; \gamma) \) is continuous in \( \gamma \) for each \( a_i \in A \). Moreover, for each \( \gamma \in \Gamma \) there exists an envelope function \( F(\cdot) \):

\[
|u_i(a_i, s; \gamma)| \leq F(a_i, s),
\]

such that \( E[F^2] < \infty \).

2. The data \( \{(a_{i,t}, s_{i,t}, s_{i,t+1})_{t=1}^{T-1}\} \) are generated by the stationary distribution determined by Markov transition kernel for the state variable. The Markov transition kernel \( K_t(\cdot, s) \), corresponding to the state transition at time \( t \) is continuous with full support on \( S \) for each \( s \in S \) and is differentiable in \( s \) for each \( t \geq 1 \). Moreover, for each \( s \in S \) \( \int K_t(s^2) \, ds < \infty \).

3. The approximating series expansion \( \{q^{k(m)}\} \) forms a basis in \( C^{k(m)}(S) \), such that the eigenvalues of \( E \left[ q^{k(m)}(s_{t+1}) \, q^{k(m)}(s_t) \big| s_t = s \right] \) are bounded away from zero for all \( s \in S \).

4. The components of the basis \( |q^{k(m)}| \leq C \) for some finite \( C \).

5. For any convex compact set \( T \subset S \), sieves provide sufficiently good approximation:

\[
\inf_{\mu \in \mathcal{R}^{k(m)}} \left\| \int s' \in S K_t(s', s) \, ds' - \mu^{k(m)} \right\| = O(k(m)^{-\alpha}),
\]

for some \( \alpha < \frac{1}{2} \) and all \( t \geq 1 \).

Assumption 2 delivers the conditions that assure that the first-step estimator will be consistent. Since \textit{ex ante} value function is a simple transformation of choice probabilities, it follows that the second-step estimator of \textit{ex ante} value function is also consistent. Moreover, provided that the estimator is differentiable, we will be able to deliver the convergence rate for its estimation.

**Theorem 3** Under Assumption 2.2-2.5, denoting the overall data size \( M \), conditions \( k(M)/M \to \infty \) and \( M/k(M)^{1+2\alpha} \to 0 \), we obtain that

\[
\| \hat{\sigma}_{k,t}(s) - \sigma_{k,t}(s) \| = O_p \left( \sqrt{\frac{k(M)}{M}} \right)
\]

This theorem establishes the convergence rate for the first- and second-step estimator. Provided that we assumed differentiability of the Markov kernel, we can use the existing statistical results to establish
the optimal convergence rate. In particular, if the state variable has \(d\) dimensions, then the optimal (fastest) convergence rate is \(M^{\frac{2}{d+2}}\). The convergence rate of the first- and second-step estimators is important because it allows us to establish the parametric convergence rate for the estimator of our structural parameters \(\gamma\).

**Theorem 4** Under Assumption 2, provided that \(M/k(M) = o(M^{1/4})\) and \(M/k(M)^{1+2\alpha} \to 0\), we obtain that for the third-step estimator \(\hat{\gamma}\):

\[
\|\hat{\gamma} - \gamma\| = O(M^{-1/2}).
\]

We note that the crucial condition to attain the parametric convergence rate for our estimator of structural parameters is a sufficiently fast convergence rate for the first- and second-step estimators. This implicitly imposes the restriction on the dimensionality of the vector of state variables. In order to obtain an estimator of structural parameters that converges at a parametric rate, we need the dimensionality of the state variable vector not to exceed 4. If we use a larger state space, then we need to impose additional smoothness assumptions on the Markov transition kernel. In particular, if we allow the Markov transition kernel to have 2 continuous derivatives, then the allowed dimensionality of the state variable vector will not exceed 6.

This means that one needs to be careful in coordinating the dimensionality of state space and smoothness restrictions on the transition of the state variable. If the relative rate condition of Theorem 4 is satisfied then we can establish the following result.

**Theorem 5** Suppose that \(E \left[ \left( \frac{\partial u_i(a_j, \gamma)}{\partial \gamma} \bigg|_{\gamma = \gamma_0} \right)^2 \right] < \infty\) and random variable \(\sigma_k, t(s_i)\) for a fixed \(t\) satisfies the Lindeberg condition. Moreover, suppose that Assumption 2 is satisfied and the first-step estimator satisfies \(M/k(M) = o(M^{1/4})\) and \(M/k(M)^{1+2\alpha} \to 0\). Then

\[
\sqrt{M}(\hat{\gamma} - \gamma_0) \Rightarrow \mathcal{N}(0, \Omega),
\]

for some covariance matrix \(\Omega\). Moreover, the bootstrap is valid.

5 Data

We use data from LoanPerformance on subprime and Alt-A mortgages that were originated between January 2000 and September 2007 and securitized in the private-label market. The coverage of the LoanPerformance dataset is extensive, with more than 85% of all securitized subprime and Alt-A mortgages included in the dataset.
For each loan, we observe the loan terms and borrower characteristics reported at the time of origination, such as the type of mortgage (fixed rate, adjustable rate, etc.), the initial contract interest rate, the level of documentation (full, low, or nonexistent\textsuperscript{11}), the appraisal value of the property, the LTV ratio, the location of the property (by zip code), and the borrower’s FICO score. We focus on 30-year fixed-rate mortgages, the most common mortgage type. We further restrict our sample to loans that are first liens and are for properties located in 20 major Metropolitan Statistical Areas (MSAs)\textsuperscript{12}.

The data also track each loan over the course of its life, reporting the outstanding balance, delinquency status, and scheduled payment in each month. We define default as occurring if the loan is delinquent for more than 90 days, a common definition of default in the mortgage literature. Default is a terminal event, so if a loan defaults in month $t$, the loan is no longer in the sample starting from month $t + 1$. We define prepayment as occurring if the loan balance goes to zero before maturity because the borrower pays the loan in full (likely through refinancing). We track the status of each loan in our sample through December 2009. This means that we have data on only up to the first 10 years of subprime loans, although the loans have maturity of 30 years. However, our proposed estimation methodology does not suffer from this data constraint as we don’t have to forward simulate for all future periods. One-period ahead forward simulation is sufficient for identification of the parameters under our estimation methodology.

We do not directly observe the borrower’s income at the time of origination, our proxy for current income. Instead, we impute it based on the reported front-end debt-to-income ratio\textsuperscript{13}. The front-end debt-to-income ratio is available only for a very small fraction (3.5\%) of all loans, significantly reducing our sample. In our earlier work (Bajari, Chu and Park, 2011) we found that this sample restriction did not affect our main findings on borrowers’ default behavior. Furthermore, even with this restriction, we still have more than half a million borrowers in the sample. Hence, we use this sample throughout this paper. For more detailed discussions of the LoanPerformance data, see Demyanyk and van Hemert (2009) and Keys et al. (2010).

Data on monthly county-level unemployment rates, our proxy for individual-level unemployment, come from the Bureau of Labor Statistics. To track movements in home prices, we use housing price indices (HPI) at the zip code level, also from LoanPerformance. The home price indices are reported at a monthly frequency, and are determined using the transaction prices of the properties that undergo

\textsuperscript{11}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.
\textsuperscript{12}The MSAs included in our sample 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.
\textsuperscript{13}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.
repeat sales at different points in time in a given geographic area. We impute the current value of a home by adjusting its appraised value at the time of origination by the index. Because home-price declines are thought to be one of the main drivers behind the recent surge in mortgage defaults, and because there is a high degree of variation across locations in home-price movements, it is important to have home-price data at a fine geographic level. Hence, we believe that the use of the zip-code level HPI from LoanPerformance enhances the robustness of our results. By contrast, most previous studies on mortgages and on housing markets in general have used the HPI from Case-Shiller, which is only at the MSA level. Table 1 describes variables that are included in the state vector.

6 Estimation Results

6.1 Results on First-Step Estimates

We start by discussing first-step estimates of the policy function. Because the policy function estimates are reduced-form in nature, the estimates themselves do not have well-defined economic interpretations. Thus, we focus on the goodness of fit of the policy function estimates, instead of discussing the coefficients. Having policy function estimates that do a reasonable job of matching empirical probabilities is crucial for the plausibility of the simulation results to follow. We investigate the performance of our policy function estimates in three ways. First, we report within-sample fit of our estimates, where we do the first-step estimation using the full sample and compare the predicted probabilities of default, prepay and pay in each period to the empirical counterparts. Second, we report out-of-sample fit of our estimates, where we use a half of the sample for estimation and the other half for validation, and compare the predicted probabilities in each period of the validation sample to the empirical counterparts. Third, for each loan in the data, we start with its first observation and forward simulate the borrowers’ decisions until the end of 2009 using the first-step estimates of policy function and state transitions. We then compare the predicted probability of eventual default or prepay by the end of 2009 to the empirical counterpart. The fit in the previous two methods depends on the precision of policy function estimates only (since we use the realized values of state variables in each period in computing the predicted probabilities of default, prepay and pay), while the fit in this third method depends on the precision of both policy function estimates and state transition estimates. More noise is introduced in the third method, so the fit is necessarily worse.

Table 2 shows within-sample fit, reporting the overall fit as well as fit by various subgroups. The table
clearly shows that the within-sample fit of the first-step policy function estimates are excellent.

Because we included very flexible splines of the state variables in estimation of the policy function, one might worry about over-fitting and potentially poor performance of out-of-sample predictions. To check this possibility, we split our sample into two and use one half for estimation and the other half for validation. The fit for the validation sample is reported in Table 3.

Table 3 shows that the fit is excellent even in the validation sample, although, not surprisingly, it is slightly worse than the within-sample fit in Table 2. Although the fit is great in Tables 2 and 3, they only reflect accuracy of the first-step policy function estimates. Another critical piece that will play an important role in counterfactual simulations is the accuracy of the estimated state transitions. To evaluate the combined fit of estimated policy functions and transition functions, we start with the first observation of each borrower, simulate the path using the estimated policy functions and transition functions, and then compare the simulated path to the actual data. Table 4 reports comparison of the predicted paths against the actual paths. In particular, we compare the probability of eventual default or prepay by the end of 2009 (which corresponds to the end of the estimation sample so that we can make meaningful comparison between predictions and data) as well as the duration until default or prepay.

The table again shows comparisons for the overall sample as well as for various slices of the sample. It is clear from Table 4 that the fit is not as good as in previous tables due to the additional noise introduced by estimation error in state transitions. However, we still find that the first-step estimates explain the data fairly well. We also plan to try different specifications for state transitions to see how the fit is influenced by the choice of state transition function.

6.2 Results on Structural Estimates

Following the estimation procedure outlined in Section 4.1, we estimate structural parameters of the per period utility as well as the discount factor. We use seemingly unrelated regression (SUR) for the system of two equations (one whose dependent variable is $\log(\sigma_{1,t}(s_{1,t})/\sigma_{0,t}(s_{1,t}))$ and another whose dependent variable
is \( \log(\frac{\tau_{t,c}(s_{t-1})}{\tau_{0,t}(s_{t-1})}) \). We impose cross-equation restrictions that the discount factor should be the same in the prepay and pay equations and that the degree of disutility from payment should be the same in both equations. Table 5 reports our estimates of the structural parameters.

Table 5 about here

Our results are overall very intuitive. Higher home value makes pay option more attractive than default option, borrowers with high credit quality tend to prefer pay option compared to default option, and higher monthly payment makes pay option less attractive than default option. Furthermore, our estimate of monthly discount factor is reasonable. There is nothing in the model that restricts the estimate of the discount factor, so it is very reassuring to see that our estimated discount factor is a plausible number. It is interesting to note that the monthly interest rate implied by the estimated discount factor is higher than the average monthly interest rate these borrowers have on their mortgages in the data. In other words, these borrowers seem to discount the future more than what the market interest rate suggests. This could be partly due to behavioral biases of consumers or due to the fact that liquidity shocks are not fully captured by our model.

add results on investors vs. owner-occupiers

7 Simulations and Welfare Consequences

In this section, we consider how borrowers’ default behavior would change under various scenarios and compute the corresponding change in consumer welfare. Our analysis uses the first-step estimates to simulate borrowers’ behavior under various “counterfactual” regimes. This approach is an unconventional approach and deserves discussion. When we conduct counterfactual analysis using the first-step policy function estimates, which are reduced-form in nature, it is very natural to worry about the Lucas critique. However, given the panel structure of our data, our setup is not subject to this critique so long as two requirements are met. First, for each scenario, the forward-simulated distribution of state variables must remain within the empirical support of the policy function. Note that in our panel data, we have much contemporaneous variation in home prices, credit quality, net equity levels, and so on. Our reduced-form policy function is a valid description of equilibrium behavior for any realization of state variables in the empirical support. Therefore, we can use the reduced-form policy function to determine how aggregate outcomes would change so long as the scenario under consideration does not lead to simulated values of state variables lying outside the empirical support. For example, we could predict the trajectory of the aggregate default rate for a scenario in which all housing markets experience the same evolution of
home prices as a particular MSA observed in the data. More generally, we can study any scenario that involves situations that actually occurred for some subset of borrowers in the data. A second requirement is that the transitions of the state variables must be unchanged under the counterfactuals, because the first-step reduced-form policy function estimates are implicitly conditioned on the transition functions. If our scenarios involve changes in the transition functions of the state variables, the first-step policy function estimates would no longer be valid. We judiciously choose our simulation exercises that meet these requirements and conduct them using the first approach.

Throughout our simulation exercises, we maintain a couple of key assumptions. First, we assume that the evolution of the macro state variables follows an exogenous process. This partial equilibrium approach makes the problem more tractable. A general equilibrium model where home prices and interest rates are endogenously determined is beyond the scope of this paper. In particular, such a general approach would require modeling how default decisions would feed into the determination of home prices, which is not an easy task.

Furthermore, our model addresses default decisions of borrowers conditional on their having already obtained a mortgage. As such, our model cannot be used to predict how loan originations would change under the various scenarios. All simulation exercises in this paper implicitly assume that mortgage originations remain unaffected by the proposed changes, and solely focus on the default behavior of loans that have been originated.

We perform the following set of simulation exercises using the first-step policy function estimates. Some of them are intended to shed light on the relative importance of the major factors that contributed to higher default rates in recent years. Others are intended to assess the impact of various foreclosure mitigation policies.

We know that the subprime mortgage crisis of 2007 was partly characterized by an unusually large fraction of subprime mortgages originated in 2006 becoming delinquent or going into foreclosure only months later. For instance, the cumulative empirical probability of default by the end of 2007 is 10.81% for mortgages originated in 2006, compared to 6.48% for mortgages originated in 2004, even though the older loans have had more time over which to default. The first set of simulations focuses on explaining this difference in performance between older and newer loans:

1. It is well known that falling home prices played a key role in the recent increase in defaults. To quantify its importance, we ask what the aggregate default rate among subprime borrowers would have been under alternative evolutionary paths for home prices. Specifically, we ask how all loans in the data that were alive as of January 2004 would have fared up through the censoring date
(December 2009) if in that year homes in all markets had experienced the same precipitous decline in value as the average Las Vegas house three years later in 2007 (Scenario 1). By comparing the predicted default rates given actual home price changes to predicted default rates under the counterfactual scenario, we can determine how home price declines affect borrowers’ default behavior. Another exercise simulates the default behavior of all subprime loans that were alive as of January 2006 under the counterfactual scenario in which all homes nationwide experience an increase in value in the year 2006 equalling that experienced by the average Las Vegas house two years earlier, in 2004 (Scenario 2). This exercise tells us how the mortgage market would have performed in 2007-2009 if the housing bubble of the earlier years had continued.

2. In our earlier work (Bajari, Chu and Park, 2011), we found that deterioration over time in the credit quality of subprime borrowers was another major factor behind the recent increase in subprime defaults. To investigate this issue, we examine how much lower aggregate default rates would be if the borrowers who took out loans in the later years had the same overall credit quality as the borrowers from the earlier years. Specifically, we shift the distribution of FICO scores for new borrowers in 2006 upward to match the mean of FICO scores among borrowers in 2004. We then simulate out default decisions for the new loans in 2006 until the censoring date of December 2009. (Scenario 3).

The second set of simulations are intended to evaluate the effects of foreclosure mitigation policies:

1. How effective would mortgage principal write-downs be? Scenario 4 considers the effect of a 10% principal write-down on all outstanding loans; Scenario 5 considers a 20% write-down.

2. Another policy we consider is a cap on the loan-to-value ratio. It is widely believed that loosened underwriting standards, such as the relaxation of downpayment requirements, paved the way for the mortgage crisis. Scenario 6 considers what would happen if LTVs at origination were capped at 0.8 (20% downpayment) for all borrowers whose actual LTVs at origination exceeded 0.8; Scenario 7 caps the original LTV at 0.9 (10% downpayment). Such a stricter requirement reduces the chance of borrowers going underwater even if home prices decline, thereby reducing the incentive to default.

In Table 6, we report simulation results for the counterfactual cases, which we show alongside the baseline model predictions for comparison. The results for Scenarios 1 and 2 indicate that more housing price appreciation causes a substantial reduction in default. Our model explains this effect in part by giving borrowers more net equity at each future point in time, relative to the baseline case. As well, autocorrelation over time in housing prices implies that higher current appreciation leads to higher
expectations for future appreciation. The reverse is true when we subject borrowers to a large price decline: borrowers are much more likely to default because the price decline pushes some borrowers deep into negative net equity, reducing the loss from walking away from the loan, and also creates expectations of future price declines.

[Table 6 about here]

Increasing the overall credit quality of the borrower pool significantly reduces the aggregate default probability. This suggests that loosened underwriting standards, which permitted consumers with low credit quality to obtain mortgages, was a significant contributor to the higher default rates among subprime mortgages in recent years. Principal write-downs have the intended effect of reducing default. LTV caps also have the same qualitative effect, but have a much smaller impact than principal write-downs. This difference in magnitude is primarily driven by the fact that the LTV caps are binding only for a small fraction of borrowers, whereas the principal write-downs are applied to the entire population.

In Table 7, we report how mortgage borrowers’ welfare changes under the various scenarios. Using (4), we compute the ex ante value function of each borrower in each time period under the baseline scenario and counterfactual scenario. As a summary measure of how borrowers’ welfare changes between the two scenarios, we compare the mean of the ex ante values across borrowers and over time between the two scenarios. The table shows that, not surprisingly, borrowers’ welfare goes up when the aggregate default rate goes down and borrowers’ welfare goes down when the aggregate default rate goes up.

[Table 7 about here]

In order to understand the overall welfare impact on borrowers, lenders and the economy, one needs to formulate lenders’ value function and understand the spillover effects of borrowers’ default decisions on the overall economy, which is beyond the scope of this paper. However, we believe that understanding borrowers’ welfare change under various policy levers is a useful first step.

8 Conclusion

We have estimated and simulated a dynamic structural model of mortgage default. Using our model, we have quantified the importance of home price declines as well as looser underwriting standards in

\footnote{Due to nonstationarity of the value function, the average of the ex ante values over time does not have a clear meaning. However, we think the average is still a useful summary measure for comparison of welfare between two scenarios since the degree of nonstationarity is the same between the two scenarios.}
creating the conditions that led to the recent wave of mortgage defaults. We have also used the model to investigate the impact of various mortgage policies that have been proposed by regulators in response to the mortgage crisis.

In addition to answering these timely economic questions, our paper makes a few methodological contributions. First, we propose an estimation method that is intuitive and easy to implement. Second, we prove that we can identify the discount factor by exploiting the features of the empirical setup. Finally, we show that it is possible to conduct some interesting counterfactual analyses via simulations based on the first-step policy function estimates. This approach allows us to avoid having to re-solve the dynamic programming problem for more than half a million borrowers, making the analyses much more computationally feasible. The same idea could be usefully applied to other empirical problems where the structure of data makes certain counterfactual analyses immune to the Lucas critique.
References


Table 1: State Variables in First-Step Policy Function Estimation

<table>
<thead>
<tr>
<th>State Variable</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low Doc</td>
<td>= 1 if the loan was done with no or low documentation, = 0 otherwise.</td>
</tr>
<tr>
<td></td>
<td>Fixed over time.</td>
</tr>
<tr>
<td>Multiple Liens</td>
<td>= 1 if the borrower has other, junior mortgages, = 0 otherwise.</td>
</tr>
<tr>
<td></td>
<td>Fixed over time.</td>
</tr>
<tr>
<td>FICO</td>
<td>FICO score, a credit score developed by Fair Issac &amp; Co.</td>
</tr>
<tr>
<td></td>
<td>Scores range between 300 and 850, with higher scores indicating</td>
</tr>
<tr>
<td></td>
<td>higher credit quality. 5 splines of FICO are used in estimation.</td>
</tr>
<tr>
<td></td>
<td>Fixed over time.</td>
</tr>
<tr>
<td>Income</td>
<td>Monthly income reported at origination. 3 splines are used.</td>
</tr>
<tr>
<td></td>
<td>Fixed over time.</td>
</tr>
<tr>
<td>Interest Rate</td>
<td>Contractual interest rate. 3 splines are used in estimation.</td>
</tr>
<tr>
<td></td>
<td>Fixed over time.</td>
</tr>
<tr>
<td>Market Rate</td>
<td>Market interest rate. 3 splines are used in estimation. Lagged</td>
</tr>
<tr>
<td></td>
<td>value of market rate is also included since it affects state transition.</td>
</tr>
<tr>
<td>H</td>
<td>Current housing value. 3 splines are used in estimation.</td>
</tr>
<tr>
<td></td>
<td>Lagged value of H is also included.</td>
</tr>
<tr>
<td>Net Equity</td>
<td>Current housing value - outstanding loan balance. 3 splines are used.</td>
</tr>
<tr>
<td>Payment</td>
<td>Monthly payment. 3 splines are used. Fixed over time.</td>
</tr>
<tr>
<td>Unemployment Rate</td>
<td>Monthly unemployment rate at the county level. 3 splines are used.</td>
</tr>
<tr>
<td></td>
<td>Lagged value of unemployment rate is also included.</td>
</tr>
<tr>
<td>Original LTV</td>
<td>Loan to value ratio at origination. 3 splines are used.</td>
</tr>
<tr>
<td>MSA dummies</td>
<td></td>
</tr>
</tbody>
</table>

We also include in estimation interactions among the state variables in first stage estimation.
Table 2: Within-Sample Fit of First-Step Estimates

<table>
<thead>
<tr>
<th></th>
<th>Default Probability Prediction</th>
<th>Data</th>
<th>Prepay Probability Prediction</th>
<th>Data</th>
<th>Pay Probability Prediction</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>0.530%</td>
<td>0.530%</td>
<td>1.847%</td>
<td>1.847%</td>
<td>97.622%</td>
<td>97.622%</td>
</tr>
<tr>
<td>FICO G1</td>
<td>1.136%</td>
<td>1.139%</td>
<td>2.357%</td>
<td>2.378%</td>
<td>96.507%</td>
<td>96.483%</td>
</tr>
<tr>
<td>FICO G2</td>
<td>0.805%</td>
<td>0.806%</td>
<td>2.012%</td>
<td>2.001%</td>
<td>97.183%</td>
<td>97.193%</td>
</tr>
<tr>
<td>FICO G3</td>
<td>0.648%</td>
<td>0.647%</td>
<td>1.878%</td>
<td>1.875%</td>
<td>97.474%</td>
<td>97.478%</td>
</tr>
<tr>
<td>FICO G4</td>
<td>0.417%</td>
<td>0.415%</td>
<td>1.802%</td>
<td>1.807%</td>
<td>97.781%</td>
<td>97.777%</td>
</tr>
<tr>
<td>FICO G5</td>
<td>0.169%</td>
<td>0.173%</td>
<td>1.704%</td>
<td>1.687%</td>
<td>98.127%</td>
<td>98.140%</td>
</tr>
<tr>
<td>FICO G6</td>
<td>0.069%</td>
<td>0.066%</td>
<td>1.367%</td>
<td>1.383%</td>
<td>98.564%</td>
<td>98.551%</td>
</tr>
<tr>
<td>Orig LTV G1</td>
<td>0.211%</td>
<td>0.214%</td>
<td>1.618%</td>
<td>1.589%</td>
<td>98.172%</td>
<td>98.197%</td>
</tr>
<tr>
<td>Orig LTV G2</td>
<td>0.334%</td>
<td>0.334%</td>
<td>1.760%</td>
<td>1.758%</td>
<td>97.906%</td>
<td>97.908%</td>
</tr>
<tr>
<td>Orig LTV G3</td>
<td>0.476%</td>
<td>0.478%</td>
<td>1.840%</td>
<td>1.861%</td>
<td>97.685%</td>
<td>97.661%</td>
</tr>
<tr>
<td>Orig LTV G4</td>
<td>0.592%</td>
<td>0.591%</td>
<td>1.816%</td>
<td>1.817%</td>
<td>97.592%</td>
<td>97.592%</td>
</tr>
<tr>
<td>Orig LTV G5</td>
<td>0.797%</td>
<td>0.813%</td>
<td>2.066%</td>
<td>2.074%</td>
<td>97.137%</td>
<td>97.113%</td>
</tr>
<tr>
<td>Orig LTV G6</td>
<td>0.748%</td>
<td>0.733%</td>
<td>2.005%</td>
<td>1.982%</td>
<td>97.247%</td>
<td>97.286%</td>
</tr>
<tr>
<td>Low Doc = 0</td>
<td>0.529%</td>
<td>0.529%</td>
<td>1.822%</td>
<td>1.822%</td>
<td>97.649%</td>
<td>97.649%</td>
</tr>
<tr>
<td>Low Doc = 1</td>
<td>0.534%</td>
<td>0.534%</td>
<td>1.898%</td>
<td>1.898%</td>
<td>97.569%</td>
<td>97.569%</td>
</tr>
<tr>
<td>Multi Liens = 0</td>
<td>0.514%</td>
<td>0.514%</td>
<td>1.899%</td>
<td>1.899%</td>
<td>97.587%</td>
<td>97.587%</td>
</tr>
<tr>
<td>Multi Liens = 1</td>
<td>0.745%</td>
<td>0.745%</td>
<td>1.190%</td>
<td>1.190%</td>
<td>98.065%</td>
<td>98.065%</td>
</tr>
<tr>
<td>Net Equity G1</td>
<td>1.654%</td>
<td>1.679%</td>
<td>0.304%</td>
<td>0.263%</td>
<td>98.043%</td>
<td>98.058%</td>
</tr>
<tr>
<td>Net Equity G2</td>
<td>0.711%</td>
<td>0.697%</td>
<td>1.131%</td>
<td>1.151%</td>
<td>98.158%</td>
<td>98.152%</td>
</tr>
<tr>
<td>Net Equity G3</td>
<td>0.549%</td>
<td>0.542%</td>
<td>1.803%</td>
<td>1.800%</td>
<td>97.649%</td>
<td>97.658%</td>
</tr>
<tr>
<td>Net Equity G4</td>
<td>0.330%</td>
<td>0.335%</td>
<td>2.447%</td>
<td>2.470%</td>
<td>97.223%</td>
<td>97.195%</td>
</tr>
<tr>
<td>Net Equity G5</td>
<td>0.185%</td>
<td>0.185%</td>
<td>2.626%</td>
<td>2.592%</td>
<td>97.189%</td>
<td>97.222%</td>
</tr>
<tr>
<td>Net Equity G6</td>
<td>0.110%</td>
<td>0.109%</td>
<td>1.910%</td>
<td>1.922%</td>
<td>97.980%</td>
<td>97.970%</td>
</tr>
</tbody>
</table>

This table examines probability of default/prepay/pay in each period.

G1: bottom 10% in the specified variable; G2: 10%-25%; G3: 25%-50%; G4: 50%-75%; G5: 75%-90%; G6: top 10% in the specified variable.
Table 3: Out-of-Sample Fit of First-Step Estimates

<table>
<thead>
<tr>
<th>Prediction Data</th>
<th>Prepay Probability</th>
<th>Pay Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Default Probability</td>
<td>Prediction</td>
<td>Data</td>
</tr>
<tr>
<td>All</td>
<td>0.536%</td>
<td>0.524%</td>
</tr>
<tr>
<td>FICO G1</td>
<td>1.212%</td>
<td>1.062%</td>
</tr>
<tr>
<td>FICO G2</td>
<td>0.849%</td>
<td>0.758%</td>
</tr>
<tr>
<td>FICO G3</td>
<td>0.644%</td>
<td>0.652%</td>
</tr>
<tr>
<td>FICO G4</td>
<td>0.395%</td>
<td>0.435%</td>
</tr>
<tr>
<td>FICO G5</td>
<td>0.152%</td>
<td>0.191%</td>
</tr>
<tr>
<td>FICO G6</td>
<td>0.073%</td>
<td>0.060%</td>
</tr>
<tr>
<td>Orig LTV G1</td>
<td>0.226%</td>
<td>0.194%</td>
</tr>
<tr>
<td>Orig LTV G2</td>
<td>0.348%</td>
<td>0.319%</td>
</tr>
<tr>
<td>Orig LTV G3</td>
<td>0.481%</td>
<td>0.470%</td>
</tr>
<tr>
<td>Orig LTV G4</td>
<td>0.585%</td>
<td>0.597%</td>
</tr>
<tr>
<td>Orig LTV G5</td>
<td>0.809%</td>
<td>0.812%</td>
</tr>
<tr>
<td>Orig LTV G6</td>
<td>0.754%</td>
<td>0.719%</td>
</tr>
<tr>
<td>Low Doc = 0</td>
<td>0.542%</td>
<td>0.515%</td>
</tr>
<tr>
<td>Low Doc = 1</td>
<td>0.523%</td>
<td>0.542%</td>
</tr>
<tr>
<td>Multi Liens = 0</td>
<td>0.523%</td>
<td>0.504%</td>
</tr>
<tr>
<td>Multi Liens = 1</td>
<td>0.699%</td>
<td>0.780%</td>
</tr>
<tr>
<td>Net Equity G1</td>
<td>1.672%</td>
<td>1.659%</td>
</tr>
<tr>
<td>Net Equity G2</td>
<td>0.732%</td>
<td>0.661%</td>
</tr>
<tr>
<td>Net Equity G3</td>
<td>0.556%</td>
<td>0.550%</td>
</tr>
<tr>
<td>Net Equity G4</td>
<td>0.329%</td>
<td>0.323%</td>
</tr>
<tr>
<td>Net Equity G5</td>
<td>0.181%</td>
<td>0.203%</td>
</tr>
<tr>
<td>Net Equity G6</td>
<td>0.107%</td>
<td>0.103%</td>
</tr>
</tbody>
</table>

This table examines probability of default/prepay/pay in each period.

G1: bottom 10% in the specified variable; G2: 10%-25%; G3: 25%-50%;
G4: 50%-75%; G5: 75%-90%; G6: top 10% in the specified variable
Table 4: Simulated Probability of Eventual Default or Prepay by End of 2009

<table>
<thead>
<tr>
<th></th>
<th>Prob. Default</th>
<th>Duration to Default</th>
<th>Prob. Prepay</th>
<th>Duration to Prepay</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Prediction</td>
<td>18.632%</td>
<td>35.11155</td>
<td>54.970%</td>
<td>26.508</td>
</tr>
<tr>
<td>Data</td>
<td>16.005%</td>
<td>31.94577</td>
<td>55.742%</td>
<td>24.307</td>
</tr>
<tr>
<td>FICO G1 Prediction</td>
<td>32.909%</td>
<td>30.426</td>
<td>54.724%</td>
<td>26.139</td>
</tr>
<tr>
<td>Data</td>
<td>27.965%</td>
<td>28.043</td>
<td>58.381%</td>
<td>24.128</td>
</tr>
<tr>
<td>FICO G2 Prediction</td>
<td>24.856%</td>
<td>32.021</td>
<td>53.655%</td>
<td>25.661</td>
</tr>
<tr>
<td>Data</td>
<td>21.913%</td>
<td>30.522</td>
<td>54.363%</td>
<td>23.304</td>
</tr>
<tr>
<td>FICO G3 Prediction</td>
<td>20.564%</td>
<td>33.831</td>
<td>53.358%</td>
<td>25.735</td>
</tr>
<tr>
<td>Data</td>
<td>18.693%</td>
<td>32.449</td>
<td>54.191%</td>
<td>23.835</td>
</tr>
<tr>
<td>FICO G4 Prediction</td>
<td>15.283%</td>
<td>37.401</td>
<td>55.744%</td>
<td>26.723</td>
</tr>
<tr>
<td>Data</td>
<td>12.975%</td>
<td>35.143</td>
<td>56.495%</td>
<td>24.799</td>
</tr>
<tr>
<td>FICO G5 Prediction</td>
<td>8.464%</td>
<td>41.187</td>
<td>58.698%</td>
<td>27.535</td>
</tr>
<tr>
<td>Data</td>
<td>5.973%</td>
<td>36.931</td>
<td>58.106%</td>
<td>24.687</td>
</tr>
<tr>
<td>FICO G6 Prediction</td>
<td>4.671%</td>
<td>43.736</td>
<td>54.884%</td>
<td>28.982</td>
</tr>
<tr>
<td>Data</td>
<td>2.550%</td>
<td>41.772</td>
<td>53.557%</td>
<td>26.028</td>
</tr>
<tr>
<td>Net Equity G1 Prediction</td>
<td>30.198%</td>
<td>32.749</td>
<td>38.914%</td>
<td>28.649</td>
</tr>
<tr>
<td>Data</td>
<td>32.535%</td>
<td>28.843</td>
<td>35.012%</td>
<td>29.623</td>
</tr>
<tr>
<td>Net Equity G2 Prediction</td>
<td>27.683%</td>
<td>35.045</td>
<td>49.413%</td>
<td>29.195</td>
</tr>
<tr>
<td>Data</td>
<td>22.749%</td>
<td>32.007</td>
<td>52.230%</td>
<td>27.297</td>
</tr>
<tr>
<td>Net Equity G3 Prediction</td>
<td>20.420%</td>
<td>34.934</td>
<td>54.458%</td>
<td>26.528</td>
</tr>
<tr>
<td>Data</td>
<td>16.854%</td>
<td>31.599</td>
<td>57.175%</td>
<td>24.199</td>
</tr>
<tr>
<td>Net Equity G4 Prediction</td>
<td>13.751%</td>
<td>34.616</td>
<td>60.344%</td>
<td>24.580</td>
</tr>
<tr>
<td>Data</td>
<td>12.223%</td>
<td>32.707</td>
<td>60.249%</td>
<td>22.047</td>
</tr>
<tr>
<td>Net Equity G5 Prediction</td>
<td>9.153%</td>
<td>36.573</td>
<td>60.937%</td>
<td>24.650</td>
</tr>
<tr>
<td>Data</td>
<td>8.534%</td>
<td>35.357</td>
<td>57.825%</td>
<td>23.442</td>
</tr>
<tr>
<td>Net Equity G6 Prediction</td>
<td>6.716%</td>
<td>40.695</td>
<td>53.686%</td>
<td>27.550</td>
</tr>
<tr>
<td>Data</td>
<td>4.944%</td>
<td>31.943</td>
<td>51.554%</td>
<td>25.052</td>
</tr>
</tbody>
</table>

This table examines probability of eventual default/prepay by the end of 2009. Duration is measured in months.

G1: bottom 10% in the specified variable; G2: 10%-25%; G3: 25%-50%;
G4: 50%-75%; G5: 75%-90%; G6: top 10% in the specified variable
Table 5: Structural Estimates

<table>
<thead>
<tr>
<th></th>
<th>Period Payoff of Prepay</th>
<th>Period Payoff of Pay</th>
</tr>
</thead>
<tbody>
<tr>
<td>Housing Value</td>
<td>0.080 (0.002) ***</td>
<td>0.080 (0.002) ***</td>
</tr>
<tr>
<td>Payment</td>
<td>-0.037 (0.0006) ***</td>
<td>-0.037 (0.0006) ***</td>
</tr>
<tr>
<td>Income</td>
<td>-0.010 (0.0001) ***</td>
<td>0.001 (0.00004) ***</td>
</tr>
<tr>
<td>FICO</td>
<td>-0.233 (0.001) ***</td>
<td>0.021 (0.0005) ***</td>
</tr>
<tr>
<td>Unemployment Rate</td>
<td>-0.240 (0.0004) ***</td>
<td>-0.010 (0.0001) ***</td>
</tr>
<tr>
<td>Low Doc</td>
<td>0.046 (0.002) ***</td>
<td>-0.003 (0.0005) ***</td>
</tr>
<tr>
<td>Multiple Liens</td>
<td>-0.315 (0.003) ***</td>
<td>-0.004 (0.001) ***</td>
</tr>
<tr>
<td>$\beta$ (coeff on $\hat{E}[V_{t+1}(s_{i,t+1})</td>
<td>s_{i,t},a_{i,t}]$)</td>
<td>0.988 (0.0002) ***</td>
</tr>
<tr>
<td>MSA dummies</td>
<td>Included</td>
<td>Included</td>
</tr>
<tr>
<td>No. of Obs</td>
<td>1,355,544</td>
<td>1,355,544</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.8418</td>
<td>0.9893</td>
</tr>
</tbody>
</table>

SUR with constraints that coefficients on payment and $\hat{E}[V_{t+1}(s_{i,t+1})|s_{i,t},a_{i,t}]$ are the same between the two equations. We plan to report bootstrapped standard errors to account for estimation errors in first policy function functions and \textit{ex ante} value functions.

The numbers currently reported are not bootstrapped.
Table 6: Simulations using First-Step Policy Function Estimates

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Counterfactual</th>
<th>Baseline</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Default</td>
<td>Prepay</td>
</tr>
<tr>
<td>Scenario 1: Home Price Decline</td>
<td>16.311%</td>
<td>44.804%</td>
</tr>
<tr>
<td>Scenario 2: Home Price Increase</td>
<td>13.974%</td>
<td>50.540%</td>
</tr>
<tr>
<td>Scenario 3: Higher Credit Quality</td>
<td>16.031%</td>
<td>38.458%</td>
</tr>
<tr>
<td>Scenario 4: 10% Principal Write-Down</td>
<td>13.390%</td>
<td>62.600%</td>
</tr>
<tr>
<td>Scenario 5: 20% Principal Write-Down</td>
<td>9.359%</td>
<td>69.071%</td>
</tr>
<tr>
<td>Scenario 6: LTV Cap at 0.8</td>
<td>17.830%</td>
<td>55.525%</td>
</tr>
<tr>
<td>Scenario 7: LTV Cap at 0.9</td>
<td>18.532%</td>
<td>54.952%</td>
</tr>
</tbody>
</table>

This table examines probability of eventual default/prepay by the end of 2009.

The first (second) column reports predicted probability of eventual default or prepay by December 2009 under the specified counterfactual (baseline) scenario.

Table 7: Borrowers’ Welfare

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Counterfactual</th>
<th>Baseline</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scenario 1: Home Price Decline</td>
<td>7.784</td>
<td>7.956</td>
</tr>
<tr>
<td>Scenario 2: Home Price Increase</td>
<td>6.558</td>
<td>6.174</td>
</tr>
<tr>
<td>Scenario 3: Higher Credit Quality</td>
<td>6.821</td>
<td>6.174</td>
</tr>
<tr>
<td>Scenario 4: 10% Principal Write-Down</td>
<td>8.239</td>
<td>8.150</td>
</tr>
<tr>
<td>Scenario 5: 20% Principal Write-Down</td>
<td>8.378</td>
<td>8.150</td>
</tr>
<tr>
<td>Scenario 6: LTV Cap at 0.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scenario 7: LTV Cap at 0.9</td>
<td>8.155</td>
<td>8.150</td>
</tr>
</tbody>
</table>

This table compares borrowers’ welfare under the specified counterfactual scenarios against welfare under the baseline.

Our measure of welfare is the average ex ante values of borrowers.