A Dynamic Model of Subprime Mortgage Default: Estimation and Policy Implications[1]

Patrick Bajari, University of Minnesota and NBER

Sean Chu, Federal Reserve Board of Governors

Denis Nekipelov, UC Berkeley

Minjung Park, UC Berkeley

This version: November 23, 2011

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 counterfactual scenarios. The counterfactual 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] The views expressed are those of the authors and do not necessarily reflect the official positions of the Federal Reserve System. Correspondence: patbajari@gmail.com; Sean.Chu@frb.gov; nekipelov@econ.berkeley.edu; mpark@haas.berkeley.edu.

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 especial 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 of foreclosure mitigation policies by simulating borrowers' default decisions under various counterfactual 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 a simple estimation method that takes advantage of the presence of this terminal action.

Estimation of borrowers' dynamic programming problem using the 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 two-step estimation method proposed by Bajari, Benkard, and Levin (2007; BBL henceforth). In the first step, we recover decision rules as a flexible function of state variables, which include home prices, the remaining balance on the loan, monthly payments due, and borrower and loan characteristics such as credit scores. Because the problem has a finite horizon, the time to maturity itself is a state variable. Thus, we estimate decision rules separately for each time period. The richness and large size of our dataset allow us to be very flexible in our specification of the decision rules. Also as part of the first step, we 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). In contrast to the usual case in which only the *differences* in choice-specific values between alternative actions are recoverable, the presence of a terminal action with a known continuation payoff (namely, zero) allows us to recover the actual level of the ex ante expected continuation value associated with each action. Using the estimated transition function of the state variables and an assumption about the distribution of the errors, we construct the expected continuation value for prepaying and paying, respectively, via one-period forward simulation. Our proposed estimator treats the constructed continuation value and other state variables in the period utility as regressors and allows us to recover structural parameters using simple OLS or SUR.

This estimation method is intuitive and easy to implement. It also addresses one of the key data challenges we face. Because subprime mortgages typically prepay or default after only a short amount of time, and in any case 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 BBL. However, our proposed estimation method does not suffer from this problem as it requires forward simulation for one period ahead only. Another advantage of our estimation method is that it makes identification of the discount factor very clear. We prove that the discount factor is identified in our model and present our estimate of the discount factor, along with estimates of other structural parameters.

Once we estimate our model, we consider various counterfactual scenarios. In particular, we conduct counterfactual analysis in two different ways. First, 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 the structural parameters from the second 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. However, the panel structure of our data allows us to address the Lucas critique in a novel way. Because the panel structure allows us to identify the policy function over a wide range of state variables *at each point in time*, so long as we limit ourselves to studying policy interventions that do not go "out of sample" or change the state transition function, the new equilibrium behavior is correctly captured by the reduced-form policy function. In other words, if a counterfactual 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 the new equilibrium. This approach is computationally much lighter than re-solving the dynamic programming problem for each borrower, which is a considerable advantage given the large size of our data.

Second, we also conduct counterfactual analysis in the more typical way, using structural estimates recovered in the second step. Doing so allows us to compare predictions based on the two different methods and to show that counterfactuals based on the first-step estimates are valid under certain conditions. Furthermore, as we might expect, there are certain counterfactuals that can be analyzed only using the structural estimates, namely any scenarios involving transitions to states not spanned by the estimation sample or resulting in changes to the state transitions themselves. In particular, we explore the effect that a key government program called HAMP (Home Affordable Modification Program) would have had on borrower behavior under the hypothetical scenario in which the program had been introduced sooner and made available to borrowers during our sample period.[2] Unlike the previous counterfactuals, this one requires re-simulation of borrowers' optimal behavior using the structural estimates, because the modifications prescribed by HAMP often result in mortgages with longer maturities (up to 40 years) and lower interest rates (as low as 2 percent) than those seen in our sample.

Our estimation exploits a unique 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. The decision to default clearly has dynamic implications given the presence of fixed costs associated with default, such as the lasting impact of a damaged credit history. Not correctly capturing such dynamic features leads to inconsistent estimates and may generate misleading welfare implications. For the most part, 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. Finally, our paper informs our understanding of borrowers' default incentives and evaluates welfare effects of key policy tools, a topic of

[2] In reality, HAMP was introduced in 2009.

[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.

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 and discuss identification of the discount factor. Section 3 discusses our estimation methodology. In Section 4, we describe the data. Section 5 presents estimation results and Section 6 discusses our counterfactual analysis. Section 7 concludes the paper.

2 Model

We formulate borrowers' decisions using a dynamic, discrete-time, single-agent 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. The components of the model are as follows.

2.1 Decision

At each time period t over the life of borrower i's loan, the borrower chooses a decision $a_{i,t}$ from the finite set $D = \{0, 1, 2\}$.[4] The possible actions in D 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 ($a_{i,t} = 2$). We assume that there is no interaction among borrowers, which implies that our setup is a single-agent model. 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

The per period utility is as follows:

$$U(a_{i,t}, s_{i,t}) = u(a_{i,t}, s_{i,t}) + \varepsilon_{i,a_{i,t},t}$$

We assume that the per-period utility of the borrower has a deterministic component that is a time-invariant function of the state and action ($u(a_{i,t}, s_{i,t})$). As is typical in discrete-choice models, we make the following normalization: $u(a_{i,t} = 0, s_{i,t}) = 0$. The state vector $s_{i,t}$ includes borrower i's characteristics, current home value $V_{i,t}$, monthly payments $p_{i,t}$, etc. Following the literature, we assume that agents also

[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.

receive a stochastic, choice-specific shock to payoffs $\varepsilon_{i,a_{i,t},t}$, in order to allow for some degree of randomness in the action observed for each state. We assume that the shocks have a type I extreme value distribution, and are iid across borrowers, across actions, and over time. Agents are impatient and discount the future payoffs with the discount factor β.

The state variables in our model evolve according to the transition function $g(s_{i,t+1}|a_{i,t}, s_{i,t})$. Most state variables evolve according to an exogenous process, and their distribution at time $t+1$ does not depend upon the borrower's action, $a_{i,t}$. The only state variable whose transition is influenced by $a_{i,t}$ is the monthly payment, $p_{i,t}$. 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 $a_{i,t}$.

2.3 Value Function

The value function of the borrower's problem for $t < T$ is as follows:

$$\begin{aligned} V_t(s_{i,t}) &= E_{\sigma,g(s)}\left[\sum_{\tau=t}^{T}\left(\beta^{\tau-t}U(a_{i,\tau}, s_{i,\tau})\prod_{\tau_1=1}^{\tau-1}1(a_{i,\tau_1}>0)\right)|s_{i,t}\right] \\ &= \max_{k=0,1,2}\left\{u(a_{i,t}=k, s_{i,t}) + \beta E\left[V_{t+1}(s_{i,t+1})|s_{i,t}, a_{i,t}=k\right] + \varepsilon_{i,k,t}\right\} \\ &= \max_{k=0,1,2}\left\{V_t(a_{i,t}=k, s_{i,t}) + \varepsilon_{i,k,t}\right\} \end{aligned} \quad (1)$$

In the above equation, σ represents the optimal 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.

We denote the choice-specific value function by $V_t(a_{i,t}=k, s_{i,t})$. Because the period utility of default is normalized to zero ($u(a_{i,t}=0, s_{i,t})=0$) and default is a terminal action, the choice-specific value of default is zero, i.e., $V_t(a_{i,t}=0, s_{i,t})=0$.

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 $u(a_{i,T}, s_{i,T})$ plus the house value in period T, as the

[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 $r^m_{i,t} = r^m_t(z_{i,t}) + \xi_i$, where $r^m_t(z_{i,t})$ is the prevailing rate available for loans with observable characteristics $z_{i,t}$, and ξ_i is a borrower-specific spread that is constant over time. For a given borrower, we can identify ξ_i as the residual from regressing the observed interest rate on the observed characteristics of the original loan.

borrower obtains full ownership of the house once the loan matures. The choice-specific value of default in period T is equal to 0 since the borrower does not get to keep the house if he defaults. The optimal solution exists and it is unique by construction based on backward induction.

2.4 Equilibrium Probabilities

An advantage of working with extreme-value–distributed errors is that they generate rather simple equilibrium probabilities. In our model, the probabilities of default, prepay, and pay at a given state are:

$$\begin{aligned}
\sigma_{0,t}(s_{i,t}) &= \Pr(a_{i,t}=0|s_{i,t}) = \frac{\exp(V_t(a_{i,t}=0,s_{i,t}))}{\sum\limits_{k=0}^{2}\exp(V_t(a_{i,t}=k,s_{i,t}))}\\
\sigma_{1,t}(s_{i,t}) &= \Pr(a_{i,t}=1|s_{i,t}) = \frac{\exp(V_t(a_{i,t}=1,s_{i,t}))}{\sum\limits_{k=0}^{2}\exp(V_t(a_{i,t}=k,s_{i,t}))}\\
\sigma_{2,t}(s_{i,t}) &= \Pr(a_{i,t}=2|s_{i,t}) = \frac{\exp(V_t(a_{i,t}=2,s_{i,t}))}{\sum\limits_{k=0}^{2}\exp(V_t(a_{i,t}=k,s_{i,t}))}
\end{aligned} \tag{2}$$

The Hotz-Miller inversion typically allows identification of the *difference* in choice-specific values, as follows.

$$\begin{aligned}
\log\left(\frac{\sigma_{1,t}(s_{i,t})}{\sigma_{0,t}(s_{i,t})}\right) &= V_t(a_{i,t}=1,s_{i,t}) - V_t(a_{i,t}=0,s_{i,t})\\
\log\left(\frac{\sigma_{2,t}(s_{i,t})}{\sigma_{0,t}(s_{i,t})}\right) &= V_t(a_{i,t}=2,s_{i,t}) - V_t(a_{i,t}=0,s_{i,t})
\end{aligned} \tag{3}$$

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, $V_t(a_{i,t}=0,s_{i,t})=0$. Thus we can recover the *ex ante* value function, $V_t(s_{i,t})$, directly from the data. Our estimation method, discussed in the next section, exploits this feature of the model.

2.5 Identification of Discount Factor

Given the extreme value assumption, we can express the *ex ante* value function as

$$V_t(s_{i,t}) = \log\left(\sum_{k=0}^{2}\exp\left(V_t(a_{i,t}=k,s_{i,t})\right)\right) = \log\left(\frac{1}{\sigma_{0,t}(s_{i,t})}\right)$$

Then, combining the expression of the Hotz-Miller inversion (3) and the Bellman equation (1), we obtain,

$$\begin{array}{c} \log\left(\frac{\sigma_{1,t}(s_{i,t})}{\sigma_{0,t}(s_{i,t})}\right) = u(a_{i,t}=1,s_{i,t}) + \beta E\left[\log\left(\frac{1}{\sigma_{0,t+1}(s_{i,t+1})}\right)|s_{i,t},a_{i,t}=1\right] \\ \log\left(\frac{\sigma_{2,t}(s_{i,t})}{\sigma_{0,t}(s_{i,t})}\right) = u(a_{i,t}=2,s_{i,t}) + \beta E\left[\log\left(\frac{1}{\sigma_{0,t+1}(s_{i,t+1})}\right)|s_{i,t},a_{i,t}=2\right] \end{array} \tag{4}$$

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 $E\left[\log\left(\frac{1}{\sigma_{0,t+1}(s_{i,t+1})}\right)|s_{i,t},a_{i,t}=k\right]$ because we can estimate the state transitions directly from the data. Thus, if the discount factor is known, we can directly recover the per period utility nonparametrically from the data. However, we can make an even stronger statement.

Proposition 1 *Suppose that we observe the panel of borrowers over two time periods. If the discount factor β is known, then the per period utility for the pay and prepay options is nonparametrically identified. Moreover, if we observe the panel with at least three periods and there exists $\bar{s} \in S$ such that*

$$E\left[\log\left(\sigma_{0,t+1}(s_{i,t+1})\right)|s_{i,t}=\bar{s}\right] \neq E\left[\log\left(\sigma_{0,t+2}(s_{i,t+2})\right)|s_{i,t+1}=\bar{s}\right]$$

then both the discount factor and per-period utility are identified from this panel.

Proof. Above, we have demonstrated identification of the per-period utility with known β. Now we show that we can identify the discount factor as well, as long as we have a three-period panel of borrowers. For each action we can write

$$\begin{array}{c} u(a_{i,t}=k,s_{i,t}=\bar{s}) = \log\left(\frac{\sigma_{k,t}(\bar{s})}{\sigma_{0,t}(\bar{s})}\right) - \beta E\left[\log\left(\frac{1}{\sigma_{0,t+1}(s_{i,t+1})}\right)|s_{i,t}=\bar{s},a_{i,t}=k\right] \\ \text{and } u(a_{i,t+1}=k,s_{i,t+1}=\bar{s}) = \log\left(\frac{\sigma_{k,t+1}(\bar{s})}{\sigma_{0,t+1}(\bar{s})}\right) - \beta E\left[\log\left(\frac{1}{\sigma_{0,t+2}(s_{i,t+2})}\right)|s_{i,t+1}=\bar{s},a_{i,t+1}=k\right] \end{array}$$

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

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

By the assumption of the proposition, the denominator of this expression is not equal to zero. As a result, the discount factor is identified. Q.E.D. ■

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. In fact, the probability of default decreases over time even if the state variables s do not change, and this decrease is reinforced when the homeowner's equity increases over time as he pays down the loan. Unlike the prior literature on identification of time preferences (Magnac and Thesmar, 2002;

Fang and Wang, 2010), our identification of the discount factor does not require the presence of a variable that affects the state transition but not the per period utility. In fact, we can identify the discount factor even if all state variables are constant over time. 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.

3 Empirical Methodology

3.1 Estimation

Next we consider a semiparametric plug-in estimator for the considered strategic decision model. The plug-in estimator will be constructed in two steps. We consider the series estimator with the vector of orthogonal polynomials $q^K(\cdot)$. We then form the matrix $Q_K = \left(q^K(s_1), \dots, q^K(s_T)\right)$. Then the semiparametric estimation procedure can be constructed by "regression" on the vector of polynomials.

Step 1 First, we use the observed data on the mortgage decisions to estimate the choice probabilities non-parametrically. Suppose that the total number of consumers in the panel is J and the total number of observed time periods is T^*. Then, if $d^a_{i,t}$ is a binary indicator of action $a \in \{0,\, 1\,, 2\}$ and T_i is the total time to maturity for the mortgage issues for consumer i, the choice probability can be estimated, for instance, using a multi-dimensional kernel $K(\cdot)$ such that

$$\widehat{\sigma}_{k,t}(s) = q^{K\prime}\left(s\right) \left(\frac{1}{JT^*} \sum_{t=1}^{T^*} \sum_{j=1}^{J} q^K\left(s_{jt}\right) q^{K\prime}\left(s_{jt}\right) \right)^{-1} \frac{1}{JT^*} \sum_{t=1}^{T^*} \sum_{j=1}^{J} d_{j,t}^{a_{j,t}=k} q^K\left(s_{jt}\right)$$

The number of series terms will be a function of the total sample size with $K \to \infty$ as $JT^* \to \infty$.[7] Alternatively, one can use an orthogonal sieve-based estimation procedure where we would project the indicator $d^a_{j,t}$ on the sieve terms. As we will see, for our distribution results to be valid (and, thus, the first-stage 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 $(JT^*)^{1/4}$. 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 pre-pay options as

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

[7] We provide the detailed conditions further in this section.

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

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

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 To perform the second step, we estimate the expected ex ante value. 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 the series estimator:

$$\widehat{E}\left[V_{t+1}(s_{i,t+1}) \,\middle|\, s_{i,t} = s\right] = q^{K\prime}(s)\left(\frac{1}{JT^*}\sum_{t=1}^{T^*}\sum_{j=1}^{J} q^K(s_{jt})\, q^{K\prime}(s_{jt})\right)^{-1} \frac{1}{JT^*}\sum_{t=1}^{T^*}\sum_{j=1}^{J} V_{t+1}(s_{j,t+1}) q^K(s_{jt})$$

Then considering the parametric specification of the consumer's payoffs defined by $u_k(s_{it};\theta)$, we estimate parameter vector θ and the discount factor. In case where the payoffs are defined by the linear indices of the state variables, we estimate the parameters by running the second-stage regression of the estimated choice-specific payoff of $\widehat{V}_t(a_{i,t} = k, s_{i,t})$ on $\widehat{E}\left[V_{t+1}(s_{i,t+1}) \,\middle|\, s_{i,t}\right]$ and the state variables. The coefficient estimates for the state variables will correspond to the parameters θ 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.

Note that our estimation procedure requires forward simulation for one period ahead only. This is a very useful feature for our setup, since we have data only on the first several years of loans, although loans mature in 30 years. Because the forward simulation approach proposed in BBL requires forward simulation for all future periods, we would need to extrapolate the recovered policy function to periods close to maturity, which are likely to be very imprecise due to nonstationarity. Our approach avoids this issue and still allows us to estimate structural parameters.

3.2 Asymptotic Theory for the Plug-In Estimator

We consider a two-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 two step estimator such that in the first step we can estimate the policy function at a sufficiently fast rate and in the second stage we can estimate the parameters of the payoff function at a parametric rate. In the semiparametric second stage estimation, we denote the parameter vector γ.

ASSUMPTION 1

1. *Parameter space* Γ *is a convex compact set. Profit function* $u_i\left(a_i, s; \gamma\right)$ *is continuous in* γ *for each* $(a_i, a_{-i}) \in \mathcal{A}$. *Moreover, for each* $\gamma \in \Gamma$ *there exists an envelope function* $F(\cdot)$*:*

$$\left|u_i\left(a_i, s; \gamma\right)\right| \leq F(a_i, s),$$

such that $E[F^2] < \infty$

2. *The data* $\left\{\{a_{1t}, \ldots, a_{It}, s_t, s_{t+1}\}_{t=1}^{T-1}\right\}_{i=1}^{I}$ *are generated by the stationary distribution determined by Markov transition kernel for the state variable. The Markov transition kernel* $\mathcal{K}_t(\cdot, s)$*, corresponding to the state transition at time* t *is continuous with full support on* $\mathcal{S}$ *for each* $s \in \mathcal{S}$ *and is differentiable in* s *for each* $t \geq 1$*. Moreover, for each* $s \in \mathcal{S}$ $\int \mathcal{K}_t(s'^2\, ds' < \infty$

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

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

5. *For any convex compact set* $T \subset \mathcal{S}$*, sieves provide sufficiently good approximation:*

$$\inf_{\mu \in \mathbb{R}^{k(m)}} \left\| \int_{s' \in \mathcal{S}} \mathcal{K}_t(s', s)\, ds' - \mu'^{k(m)} \right\| = O(k(m)^{-\alpha}),$$

for some $\alpha < \frac{1}{2}$ *and all* $t \geq 1$*.*

Assumption 1 delivers the conditions that assure that the first-stage estimator will be consistent. Moreover, provided that the estimator is differentiable, we will be able to deliver the convergence rate for its estimation.

Theorem 2 *Under Assumption 1.2-5, 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-stage 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 $M^{\frac{2}{4+d}}$. The convergence rate of the first-step estimator is important because it allows us to establish the parametric convergence rate for the second step estimator.

Theorem 3 *Under Assumption 1, provided that $M/k(M) = o(M^{1/4})$ and $M/k(M)^{1+2\alpha} \to 0$, we obtain that for the second-step estimator $\hat{\gamma}$:*

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

We note that the crucial condition to attain the parametric convergence rate for the second step estimator is a sufficiently fast convergence rate for the first step estimator. This implicitly imposes the restriction on the dimensionality of the vector of state variables. In order to obtain a second stage estimator convergent 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 3 is satisfied then we can establish the following result.

Theorem 4 *Suppose that $E\left[\left(\frac{\partial u_i(a_i,s;\gamma)}{\partial \gamma}\Big|_{\gamma=\gamma_0}\right)^2\right] < \infty$ and random variable $\sigma_{k,t}(s_{it})$ for a fixed t satisfies the Lindeberg condition. Moreover, suppose that Assumption 1 is satisfied and the first stage 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 Ω. Moreover, the bootstrap is valid.

4 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[8]), 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 MSAs.[9]

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.[10] 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 repeat sales at different points in time in a given geographic area. We impute the current value of a home

[8] 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.

[9] 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.

[10] 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.

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.

5 Estimation Results

5.1 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 counterfactual results. 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.

[Table 2 about here]

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 about here]

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 (BTW this discrepancy might disappear when Sean does estimation using the full sample). 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.

[Table 4 about here]

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 very well.

5.2 Results on Second-Step Estimates

Following the estimation procedure outlined in Section 3.1, 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 $\log(\frac{\sigma_{1,t}(s_{i,t})}{\sigma_{0,t}(s_{i,t})})$ and another whose dependent variable is $\log(\frac{\sigma_{2,t}(s_{i,t})}{\sigma_{0,t}(s_{i,t})})$). [We would probably need to change specifications of per period payoff, since the current specifications can't explain why the probability of prepay would decrease as loan gets close to maturity] 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]

[Discuss results]

6 Counterfactual Analysis

In this section, we consider various counterfactual scenarios. In particular, we conduct counterfactual analysis in two different ways. First, we use the first-step estimates to simulate borrowers' behavior under various counterfactual regimes. Second, we use the structural parameter estimates to re-solve for the optimal behavior and compute the counterfactual outcomes. This second approach is the one typically used in the literature. The first 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 counterfactual, 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 counterfactual assumption 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 the counterfactual 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 counterfactual that only involves scenarios 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 counterfactuals involve changes in the transition functions of the state variables, the first-step policy function estimates would no longer be valid. We judiciously choose counterfactual exercises that meet these requirements and conduct them using the first approach.

For the set of scenarios considered using the first approach, we also compute the counterfactual outcomes using the second approach in order to compare predictions of these two different approaches. The comparison will allow us to show that counterfactuals based on the first-step estimates are indeed valid under certain conditions. We also use the second approach for the counterfactuals that cannot be performed using the first approach. In particular, we explore the effect that a key government program called Home Affordable Modification Program would have had on borrower behavior if the program had been introduced during our sample period. This one requires re-simulation of borrowers' optimal behavior using the structural estimates, because the modifications prescribed by HAMP often result in mortgages with longer maturities (up to 40 years) and lower interest rates (as low as 2 percent) than those seen in our sample.

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 counterfactuals. 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.

6.1 Counterfactuals using the First-Step Policy Function Estimates

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.

6.2 Counterfactuals using the Structural Estimates

In this subsection, we compare the simulation results generated using the first-stage estimates with outcomes obtained through the more typical approach of using structural estimates. In addition, we explore an additional counterfactual scenario that studies the effects of one of the key government interventions in primary mortgage markets following the subprime crisis. In an effort to stem the tide of defaults, the government introduced the Home Affordable Modification Program in 2009.[11] The objectives of the program were twofold: to create a process to help lenders identify at-risk loans that could profitably be modified in a way that reduced the borrower's total monthly payment to no more than 31 percent of gross income; and to subsidize such modifications. As of the first quarter of 2011, 1.5 million loans have been modified under HAMP on a trial basis, with 670,000 of these modifications made permanent.[12] We explore the effect that the HAMP program would have had on borrower behavior under the hypothetical scenario in which the program had been introduced sooner and made available to borrowers during our

[11] There have been two other major government interventions in the primary mortgage market. The very unsuccessful "Hope for Homeowners" program, introduced in 2008, allows delinquent borrowers to refinance into government-guaranteed loans if the lender agrees to forgive part of the principal balance. Very few lenders have participated in the program due to their reluctance to write down principal. The Home Affordable Refinance Program (HARP), created at the same time as HAMP, streamlines refinancing for borrowers whose loans are in government-sponsored entity (GSE) mortgage pools and who are current on their mortgages and meet certain other qualifying criteria. The qualifying requirements, including the restriction to GSE mortgages, exclude most if not all of the mortgages in our sample.

[12] Holden et al. (2011).

sample period.

We first identify the set of loans that, but for the timing of the HAMP program, could have qualified for a modification. To qualify for HAMP, a loan must be a first lien originated before 2009, backed by an owner-occupied property, and have an unpaid balance of at most \$729,750. The program also requires that the borrower must be under demonstrable financial hardship, but because we do not observe each borrower's current income, we ignore this requirement.

Then, for the set of qualifying mortgages, we simulate the borrowers' behavior over time assuming that their monthly payments are lowered at some point between January 2006 and the end of the sample period. The HAMP rules contain a large number of contingencies and conditions—including a probationary six-month "trial period" for borrowers of modified loans—but we only take into account the main provisions of the program. In particular, HAMP spells out that monthly payments be reduced through the following "waterfall":

- First, the lender reduces the monthly payment, $p_{i,t}$, to 38 percent of the borrower's gross income, $y_{i,t}$, through any chosen means (including interest rate reductions, extending the loan duration, forbearing on principal, or writing down principal).
- Next, $p_{i,t}$ is further reduced to 31 percent of gross income through the following sequence of steps:
  - The interest rate is reduced to a minimum of 2 percent.[13]
  - If $p_{i,t}/y_{i,t}$ is still greater than 0.31, then the loan duration, T, is extended out to a maximum of 40 years.
  - If $p_{i,t}/y_{i,t}$ is still greater than 0.31, then the lender forbears on however much principal is required to bring $p_{i,t}/y_{i,t}$ down to 0.31. The forborne principal is due with zero interest at either loan maturity or whenever the loan is prepaid.

To make our scenario more concrete, we assume that the initial reduction of $p_{i,t}/y_{i,t}$ down to 38 percent is made through the same sequence of steps as required for reducing $p_{i,t}/y_{i,t}$ from 38 to 31 percent.

In addition to specifying the terms of the loan modification, the HAMP program also contributes \$1000 yearly, for up to five years, toward the principal repayment of borrowers of modified loans who continue to pay on time. We incorporate this subsidy by reducing monthly payments to the borrower by \$83 per month for the first 5 years after modification.

[13] The rate was fixed for five years and allowed to rise thereafter, subject to a maximum rate of increase. Because we limit our attention to hypothetical modifications made in or after January 2006, we ignore the possibility of future rate increases.

Next, by averaging over simulation draws of each consumer's behavior, we compute the lender's expected net present value (NPV) of making a loan modification in each time period.[14] In addition to the parameter estimates, the expected NPV also depends upon the lender's discount rate, the severity of losses on loans that default, and the value of various government subsidies made directly to the lender.[15] We set the lender's discount rate equal to the prevailing Freddie Mac Primary Mortgage Market Survey (PMMS) rate for 30-year mortgages. For loss severities, we assume that when a loan defaults, the lender forecloses on the loan and immediately recovers an MSA-specific percentage of the current market value of the loan, minus a fixed cost of $10,000.[16] Finally, the lenders enjoy two direct subsidies: (1) The government compensates the lender for 50 percent of the reduction in monthly payments from 38 percent to 31 percent; (2) Lenders receive a lumpsum payment of $1000 for modifying a loan that is already delinquent or $1500 for modifying a still-current loan.

Finally, we compute the distribution of NPV changes implied by loan modification over our estimation sample, after rejecting modifications of loans that it never makes sense for the lender to modify. We assume that for loans that can be profitably modified, the modification occurs in the first period in which the NPV implied by the modification is greater than the NPV of the unmodified loan.[17] Additionally, we compute the aggregate change in default and prepayment behavior implied by all profitable loan modifications.

7 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 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 or implemented 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 a new 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

[14] We assume that a loan can only be modified once.

[15] By making default less likely, the $1000 annual contribution toward the borrower's principal repayment constituted an indirect subsidy to lenders.

[16] In reality, a small percentage of loans that become delinquent do eventually cure. However, because we treat default as a terminal event, we abstract from this possibility, and instead choose a conservative specification of the loss given default.

[17] The program actually required that loans pass an NPV test officially designed by the government in order qualify for a HAMP modification. The official NPV test, described at http://www.hmpadmin.com, is of course different from the NPV test that we perform based on our own model estimates. According to Holden et al. (2011), the vast majority of loans submitted for eligibility review for a HAMP modification passed the official NPV test.

setup. Finally, we show that it is possible to conduct 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 counterfactuals 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

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

[3] **Bajari, Patrick, Sean Chu, and Minjung Park**. 2011 "An Empirical Model of Subprime Mortgage Default from 2000 to 2007" Working Paper.

[4] **Demyanyk, Yuliya, and Otto van Hemert.** 2009. "Understanding the Subprime Mortgage Crisis." forthcoming in *Review of Financial Studies*.

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

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

[7] **Fang, Hanming, and Yang Wang**. 2010. "Estimating Dynamic Discrete Choice Models with Hyperbolic Discounting, with an Application to Mammography Decisions." Working Paper.

[8] **Foote, Christopher, Kristopher Gerardi, and Paul Willen**. 2008a. "Negative Equity and Foreclosure: Theory and Evidence." *Journal of Urban Economics*, 64(2): 234-245.

[9] **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.

[10] **Foster, Chester and Robert Van Order**. 1984. "An Option-Based Model of Mortgage Default." *Housing Finance Review*, 3: 351-372.

[11] **Gerardi, Kristopher, Adam Hale Shapiro, and Paul Willen.** 2008. "Subprime Outcomes: Risky Mortgages, Homeownership Experiences, and Foreclosures." Working Paper.

[12] **Gerardi, Kristopher, Andreas Lehnert, Shane Sherlund, and Paul Willen.** 2009. "Making Sense of the Subprime Crisis." Working Paper.

[13] **Holden, Steve, Austin Kelly, Douglas McManus, Therese Scharlemann, Ryan Singer, and John Worth.** 2011. "The HAMP NPV Model: Development and Early Performance." FHFA Working Paper 11-1.

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

[15] **Keys, Benjamin, Tanmoy Mukherjee, Amit Seru, and Vikrant Vig.** 2010. "Did Securitization Lead to Lax Screening? Evidence From Subprime Loans" *Quarterly Journal of Economics*, 125(1): 307-362.

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

[17] **Magnac, Thierry, and David Thesmar**. 2002. "Identifying Dynamic Discrete Decision Processes." Econometrica, 70(2): 801-816.

[18] **Mayer, Christopher and Karen Pence**. 2008. "Subprime Mortgages: What, Where, and to Whom?" NBER Working Paper.

[19] **Mian, Atif, and Amir Sufi**. 2009. "The Consequences of Mortgage Credit Expansion: Evidence from the U.S. Mortgage Default Crisis." *Quarterly Journal of Economics*, 124(4): 1449-1496.

[20] **Quigley, John, 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.

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

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

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

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

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

We include in estimation interactions between each of the above state variables and 5 splines of Tt, as well as interactions among the state variables (other than Tt).

Table 2: Within-Sample Fit of First-Step Estimates

	Default Probability		Prepay Probability		Pay Probability	
	Prediction	Data	Prediction	Data	Prediction	Data
All	0.490%	0.490%	1.870%	1.870%	97.639%	97.639%
FICO G1	0.933%	1.021%	2.2%	2.228%	96.867%	96.751%
FICO G2	0.851%	0.806%	2.129%	2.09%	97.02%	97.100%
FICO G3	0.702%	0.683%	2.053%	2.012%	97.246%	97.305%
FICO G4	0.301%	0.29%	2.036%	2.074%	97.663%	97.636%
FICO G5	0.111%	0.151%	1.193%	1.253%	98.695%	98.596%
FICO G6	0.032%	0.023%	1.328%	1.27%	98.64%	98.707%
Orig LTV G1	0.288%	0.28%	1.574%	1.401%	98.138%	98.319%
Orig LTV G2	0.265%	0.261%	1.727%	1.751%	98.009%	97.988%
Orig LTV G3	0.451%	0.434%	1.842%	1.950%	97.707%	97.615%
Orig LTV G4	0.498%	0.546%	1.779%	1.73%	97.723%	97.725%
Orig LTV G5	0.682%	0.691%	2.257%	2.356%	97.062%	96.954%
Orig LTV G6	0.783%	0.728%	2.143%	2.049%	97.074%	97.222%
Low Doc = 0	0.464%	0.464%	1.948%	1.948%	97.588%	97.588%
Low Doc = 1	0.537%	0.537%	1.733%	1.733%	97.73%	97.73%
Multi Lien = 0	0.463%	0.463%	1.944%	1.944%	97.593%	97.593%
Multi Lien = 1	0.821%	0.821%	0.973%	0.973%	98.207%	98.207%
Net Equity G1	1.566%	1.526%	0.368%	0.231%	98.066%	98.243%
Net Equity G2	0.657%	0.632%	1.161%	1.295%	98.182%	98.073%
Net Equity G3	0.511%	0.555%	1.858%	1.887%	97.631%	97.558%
Net Equity G4	0.286%	0.277%	2.411%	2.331%	97.303%	97.392%
Net Equity G5	0.176%	0.185%	2.683%	2.667%	97.141%	97.148%
Net Equity G6	0.094%	0.069%	1.899%	1.988%	98.008%	97.943%

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

	Default Probability		Prepay Probability		Pay Probability	
	Prediction	Data	Prediction	Data	Prediction	Data
All	0.523%	0.471%	1.951%	1.818%	97.526%	97.711%
FICO G1	1.018%	0.871%	2.571%	1.926%	96.411%	97.203%
FICO G2	0.927%	0.775%	2.158%	1.952%	96.916%	97.273%
FICO G3	0.724%	0.685%	2.054%	2.075%	97.223%	97.24%
FICO G4	0.325%	0.259%	2.069%	2.075%	97.605%	97.665%
FICO G5	0.121%	0.18%	1.276%	1.26%	98.603%	98.560%
FICO G6	0.024%	0.047%	1.503%	1.078%	98.473%	98.875%
Orig LTV G1	0.473%	0.093%	1.931%	1.249%	97.596%	98.659%
Orig LTV G2	0.222%	0.308%	1.756%	1.631%	98.022%	98.061%
Orig LTV G3	0.437%	0.512%	1.784%	1.994%	97.78%	97.494%
Orig LTV G4	0.512%	0.483%	1.818%	1.746%	97.670%	97.771%
Orig LTV G5	0.742%	0.698%	2.42%	2.094%	96.838%	97.207%
Orig LTV G6	0.885%	0.639%	2.345%	2.017%	96.77%	97.344%
Low Doc = 0	0.525%	0.441%	2.082%	1.850%	97.392%	97.709%
Low Doc = 1	0.518%	0.523%	1.72%	1.761%	97.762%	97.716%
Multi Lien = 0	0.484%	0.449%	2.056%	1.861%	97.459%	97.69%
Multi Lien = 1	0.994%	0.736%	0.66%	1.288%	98.347%	97.975%
Net Equity G1	1.562%	1.588%	0.354%	0.227%	98.084%	98.185%
Net Equity G2	0.747%	0.559%	1.202%	1.273%	98.052%	98.168%
Net Equity G3	0.563%	0.486%	1.895%	1.814%	97.542%	97.7%
Net Equity G4	0.31%	0.273%	2.618%	2.184%	97.072%	97.543%
Net Equity G5	0.137%	0.215%	2.97%	2.492%	96.893%	97.292%
Net Equity G6	0.149%	0.046%	1.607%	2.32%	98.244%	97.633%

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

		Prob. Default	Duration to Default	Prob. Prepay	Duration to Prepay
All	Prediction	15.949%	35.92	59.454%	27.39
	Data	14.877%	30.26	56.772%	24.20
FICO G1	Prediction	27.796%	26.74	60.012%	31.29
	Data	26.347%	27.09	57.485%	23.28
FICO G2	Prediction	20.911%	26.95	58.403%	26.76
	Data	20.968%	23.67	54.435%	23.83
FICO G3	Prediction	18.191%	30.45	58.958%	26.24
	Data	19.372%	33.20	57.068%	24.95
FICO G4	Prediction	10.828%	41.15	63.113%	26.52
	Data	8.732%	32.52	62.535%	23.22
FICO G5	Prediction	8.741%	52.13	56.446%	28.46
	Data	6.024%	45.4	50%	24.73
FICO G6	Prediction	6.122%	59.72	55.318%	28.13
	Data	0.935%	73	51.402%	26.95
Net Equity G1	Prediction	28.922%	30.71	50.578%	29.22
	Data	32.813%	25.19	42.188%	26.15
Net Equity G2	Prediction	22.074%	32.40	58.416%	31.17
	Data	20.270%	31.82	54.73%	28.4
Net Equity G3	Prediction	16.623%	35.7	61.935%	27.81
	Data	16.204%	30.36	57.407%	24
Net Equity G4	Prediction	12.622%	37.32	61.902%	24.53
	Data	11.587%	30.54	60.202%	21.56
Net Equity G5	Prediction	8.862%	39.51	59.108%	24.96
	Data	7.186%	32	61.078%	23.75
Net Equity G6	Prediction	9.710%	45.54	43.362%	29.22
	Data	4.348%	21	44.928%	24.06

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

	Period Payoff of Prepay	Period Payoff of Pay
V	0.148 (0.051) ***	0.38 (0.033) ***
Payment	-0.025 (0.009) ***	-0.025 (0.009) ***
Income	-0.035 (0.002) ***	0.012 (0.0008) ***
FICO	-0.157 (0.011) ***	0.119 (0.006) ***
Unemployment Rate	-0.331 (0.004) ***	-0.023 (0.002) ***
Low Doc	-0.15 (0.015) ***	-0.068 (0.007) ***
Multiple Liens	-0.585 (0.027) ***	-0.179 (0.013) ***
Original LTV	0.345 (0.059) ***	-0.419 (0.032) ***
β (coeff on $\hat{E}\left[V_{t+1}(s_{i,t+1})\|s_{i,t},a_{i,t}\right]$)	0.936 (0.002) ***	0.936 (0.002) ***
MSA dummies	Included	Included
No. of Obs	43180	43180
R^2	0.8850	0.9725

SUR with constraints that coefficients on payment and $\hat{E}\left[V_{t+1}(s_{i,t+1})|s_{i,t},a_{i,t}\right]$ are the same between the two equations.

Table 6: Counterfactuals using First-Step Policy Function Estimates

Scenario	Counterfactual		Baseline	
	Default	Prepay	Default	Prepay
Scenario 1: Home Price Decline	15.9%	48.27%	14.11%	51.82%
Scenario 2: Home Price Increase	13.68%	52.36%	18.56%	39.53%
Scenario 3: Higher Credit Quality	15.79%	18.56%	37.69%	18.36%
Scenario 4: 10% Principal Write-Down	12.99%	65.42%	15.95%	59.45%
Scenario 5: 20% Principal Write-Down	10.48%	70.69%	15.95%	59.45%
Scenario 6: LTV Cap at 0.8	15.46%	59.9%	15.95%	59.45%
Scenario 7: LTV Cap at 0.9	15.9%	59.44%	15.95%	59.45%

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.