Estimating a Dynamic Discrete Choice Model with Partial Observability for Household Mortgage Default and Prepayment Behaviors[1]

Chao Ma[2]

Abstract

Households are forward-looking when deciding whether to default on or refinance their mortgages. There are two types of default generated by two mechanisms: illiquidity-triggered default and strategic default. However, researchers can observe only whether households default but not whether the default is illiquidity-triggered or strategic. Moreover, typically researchers can observe only whether households prepay but not whether the prepayment is due to refinancing or moving. This paper extends the conditional choice probability (CCP) method to estimate a dynamic discrete choice model with partially observable outcomes. Exclusion restrictions and identification at infinity arguments provide the identification conditions. Counterfactual analyses for foreclosure-mitigating loan modification policies show that writing down the principal can reduce both illiquidity-triggered default and strategic default, that interest reduction can reduce illiquidity-triggered default but cannot effectively reduce strategic default, and that term extension can reduce illiquidity-triggered default but will increase strategic default.

Keywords: Dynamic Discrete Choice, Partial Observability, Mortgage, Default, Prepayment, Refinance, Illiquidity, Mobility, Foreclosure

---

[1] I am indebted to Jason Blevins, Donald Haurin, Stephen Cosslett, and Maryam Saeedi for many helpful discussions, guidance and support. I also thank Victor Aguirregabiria, Arie Bereastanu, David Blau, Juin Kuan Chong, Chris Conlon, N. Edward Coulson, Morris Davis, Michael Dickstein, Javier Donna, Lucia Dunn, Liran Einav, Marc Francke, Igal Hendel, Teck Hua Ho, Robert De Jong, Daeho Kim, Yuichi Kitamura, Brad Larsen, Robin Lee, Lung-Fei Lee, Matt Lewis, Jiaming Mao, Whitney Newey, Xun Tang, Huanxing Yang, Nan Yang, Jing Zhang, and seminar participants at Ohio State University, University of Oregon, University of Connecticut, National University of Singapore, Kansas State University, Freddie Mac, Montclair State University, the Annual North American Meeting of the Regional Science Association International, the American Real Estate and Urban Economics Association Winter Meeting, the American Real Estate Society Annual Meeting, the Midwest Economics Association Annual Meeting, and the Annual Meeting of Center for Financial Security at University of Wisconsin-Madison for their helpful comments and suggestions. I would especially like to thank Stephanie Moulton for sharing the mortgage data. All errors are my own.

[2] Wang Yanan Institute for Studies in Economics (WISE) & School of Economics, Xiamen University, Xiamen, Fujian, China 361005. Email: ma.271@buckeyemail.osu.edu

JEL Classifications: C35, C5, C63, D12, D14, D91, G21, J61, L85, R2

1 Introduction

During the 2007-2009 financial crisis, we observe two important phenomena in the residential mortgage market: 1) high default rates; 2) low interest rates and a large number of refinance activities. Thoroughly understanding the mechanisms generating defaults and prepayments is valuable in the following three aspects: first, it can help lenders improve the evaluation and control of the risks of default and prepayment; second, it can help investors improve the pricing of mortgage-backed securities; third, it can help policy makers design interventions to mitigate foreclosures.

There are two options embedded in a mortgage. The choice to default is a put option, while the choice to refinance is a call option. If the value of a house is below the unpaid mortgage balance, then the house is underwater and the household has a positive financial incentive to default. If the current interest rate in the mortgage market is lower than the coupon rate of the household's existing fixed-rate mortgage, then the household has a positive financial incentive to refinance. Kau, Keenan, Muller and Epperson (1995) developed a theoretical option-pricing model to evaluate mortgages.

Most previous empirical research on mortgage termination, such as Deng, Quigley and Order (2000), has been based on static models.[3] However, in reality, households make decisions in dynamic settings in which there are tradeoffs both between defaulting right now and defaulting later and between refinancing right now and refinancing later. Meanwhile, households' expectations regarding house price and interest rate movements in the future affect their decisions in the current period.[4] First, suppose that in the current period house

[3] Deng, Quigley and Order (2000) and Deng (1997) used the proportional hazard model (PHM) to analyze households' choices of default and prepayment for their mortgages. Kau, Keenan and Li (2011) used a shared-frailty survival model to analyze mortgage termination risk with a control for unobserved heterogeneity. Calhoun and Deng (2002), Rose (2012), and Ergungor and Moulton (2014) applied a multinomial logit model (MNL). An, Clapp and Deng (2010) added "selling the house" into the set of choices for each period and estimated a nested multinomial logit model (NMNL) with a control for omitted mobility characteristics. Clapp, Goldberg, Harding and LaCour-Little (2001) compared the results of a MNL and a Cox PHM for the choices of refinance, move, default, or continuing to pay. Clapp, Deng and An (2006) compared the results of a MNL, a PHM, a mass-point mixed logit (MML) model, and a mass-point mixed hazard (MMH) model.

[4] Three types of studies on dynamics in mortgage default decisions have been conducted. First, Campbell and Cocco (2015) developed a theoretical dynamic model of households' default decisions. Second, using macro-level data, Corbae and Quintin (2015) calibrated a dynamic general equilibrium model with heterogeneous

prices are low and that many households' houses are underwater. If the forward-looking households forecast that the house prices will rebound in the future, they may not choose to default right now, as default will incur substantial costs, including losing their home, needing to move, ruining their credit rating and reputation, and losing the opportunity to refinance in the future. If they forecast that house prices will not rebound in the future, they will be more likely to choose to default, as defaulting earlier to avoid making additional payments would be better than defaulting later since they will eventually have to default. Merging households' subjective house price expectation data from the Michigan Survey of Consumers and the mortgage loan performance data from Freddie Mac, Ma (2015a) conducted a reduced-form study and found direct empirical evidence that low (high) house price expectations increase (decrease) default probabilities.

Second, suppose that the current interest rate is attractively low and that many households have a positive financial incentive to refinance. If the households forecast that the interest rate will not drop further, they will refinance in the current period because the value of refinancing will shrink in the next period, as fewer times of payments will then remain. If the households forecast that the interest rate will drop further, they may choose to wait rather than to refinance in the current period. It is true that after refinancing in the current period, they still can refinance again in the future if the interest rate drops further; however, each time they refinance, they would need to pay a fixed transaction cost, including an application fee for the new loan, an appraisal fee for their house, searching costs, and a possible prepayment penalty.

In addition to the issue regarding dynamics, partial observability of the outcomes is another important issue. There are two types of default caused by two different reasons. The first type is strategic default (or ruthless default), which is caused by the financial incentive to default. In this case, defaulting provides households with higher utility than refinancing or continuing to pay, and households strategically choose to default. The second type is illiquidity-triggered default, which occurs when households do not have enough money to make their monthly mortgage payment and consequently are forced to default. In this case,

households making decisions on mortgage selection and default. Third, using loan-level data, Carranza and Navarro (2010), Zhang (2010), Laufer (2011), and Bajari, Chu, Nekipelov and Park (2013) estimated dynamic choice models for households' mortgage default behaviors.

defaulting may not necessarily provide the households with higher utility than refinancing or continuing to pay, as the latter two choices are out of the feasible choice set since the liquidity constraint binds. Unfortunately, in mortgage loan performance data sets, researchers can observe only whether households default but not which reason causes they default. This is the partial observability problem with respect to default. Most previous studies have estimated reduced-form models in which both the measure for the financial incentive to default and the variables related to illiquidity that could trigger default are included in one equation, without treating the illiquidity-triggered default outcome and the strategic default outcome separately. For example, Deng, Quigley and Order (2000) included a measure for the financial incentive to default and the local unemployment rates related to illiquidity in a single equation. Because the two types of default outcomes are generated by two different mechanisms, if one estimates a structural dynamic model without treating them separately, the structural parameters will be biased; moreover, the model will be unable to predict the probability of illiquidity-triggered default and the probability of strategic default separately. Poirier (1980) and Meng and Schmidt (1985) developed an econometric methodology to estimate bivariate probit models with partial observability. Following this methodology, Bajari, Chu and Park (2010) included the measure of the financial incentive to default and the measure of illiquidity in two separate equations and modeled default as the outcome of a two-equation system that accounts for the partial observability of default. However, their analysis was still static. Carranza and Navarro (2010) estimated a dynamic model for mortgage default without considering the difference between illiquidity-triggered default and strategic default.

There are also two types of prepayment caused by two different reasons. The first type is refinance, i.e., households prepay their current mortgage and originate a new mortgage for the same house because the current interest rate is low. This type of prepayment is caused by the incentive to refinance. The second type of prepayment is caused by moving, i.e., households sell their houses and prepay their current mortgages without refinancing. Unfortunately, in most mortgage loan performance data sets, researchers can observe only households' prepayment actions but not whether they are caused by refinancing or moving. Most previous studies either did not distinguish between these two types of prepayment or only assumes that the prepayments are refinancing. An, Clapp and Deng (2010) merged loan performance data including 1985 mortgages and residential real estate transaction data in California by street address in order to separately identify prepayment caused by moving and prepayment related to refinancing. They found different results in the estimation pooling moving and refinancing together as one choice (prepayment) and the estimation distinguishing between moving and refinancing as two different choices. However, for most mortgage loan performance data, especially large data sets, it is either costly or infeasible to merge them with residential real estate transaction data to identify moving and refinancing. I am unaware of any financial institution with such a practice in their business.[5]

The paper closest to my work is Bajari, Chu, Nekipelov and Park (2013), which estimated a dynamic model and treated strategic default and illiquidity-triggered default as the unobserved heterogeneity among borrowers. They modeled illiquidity-triggered default as exogenous Poisson processes where each borrower has a fixed hazard of illiquidity-triggered default. The identification of the hazard of illiquidity-triggered default relied on the assumption that the probability of strategic default declines at the geometric rate as the mortgage approaches maturity, whereas the probability of illiquidity-triggered default declines at the exponential rate; therefore, starting at some period after mortgage origination, the probability of default will be dominated by the strategic default probability.

My work differentiates from Bajari, Chu, Nekipelov and Park's (2013) mainly in the following four aspects. First, different from a fixed hazard of illiquidity-triggered default, in my model the probability of illiquidity-triggered default varies according to time-variant state variables and previous actions. For example, the probability of illiquidity-triggered default can be high when the local unemployment rate is high; the probability of illiquidity-triggered default can be lower than before if borrowers refinanced to a new loan with lower monthly payment amount. Second, the identification relies on the assumption that the probability of

[5] In addition to prepayment related to refinancing and moving, there is a third type of prepayment: households simply prepay their mortgages without refinancing or selling their houses. The literature does not treat this type of prepayment specifically, as the measure of the incentive to refinance remains valid for measuring the incentive of this type of prepayment. Although the households do not refinance after prepayment, they should use other assets currently earning lower interest rates to repay their existing fixed-rate mortgage debt with higher interest rates. The current interest rates for other assets are positively correlated with the current interest rates in the mortgage market. Prepayment without refinancing or selling the houses is even less likely for the borrowers in the loan performance data used in this paper because all the borrowers are low-to-moderate income, first-time homebuyers.

default is dominated by the illiquidity-triggered default probability or the strategic default probability when some state variables have extreme realizations for some borrowers in some periods, which is easier to be satisfied than the assumption that the probability of default will be dominated by the strategic default probability starting at some period after mortgage origination.[6] Third, in my model, borrowers have an expectation for the future that is consistent with the mixture of illiquidity-triggered default and strategic default. When borrowers make decisions among strategic default, refinance, and continuing to pay, the value for the future in their mind is not only related to the ex ante value before making those choices in the future, which can be recovered from the conditional choice probability of strategic default obtained from the first-stage estimation, but also related to the probability and value of the event of illiquidity-triggered default in the future. Fourth, the dynamic discrete choice model in this paper also consider the partial observability issue of prepayment outcome, whereas Bajari, Chu, Nekipelov and Park's (2013) assumed that all the prepayments are caused by refinancing.

In the literature, there are three methods to split default into illiquidity-triggered default and strategic default. One method entails defining strategic default based on observables. Based on the observables in the data from the Panel Study of Income Dynamics (PSID), Gerardi, Herkenhoff, Ohanian, and Willen (2013) defined strategic defaulters as those with negative home equity and sufficient liquid assets to make mortgage payments for one additional month. Based on a borrower-level dataset from TransUnion, Tirupattur, Chang and Egan (2010) defined strategic defaulters as those with underwater houses and other meaningful nonmortgage obligations on which they continue performing. Another method involves resorting to survey data. Guiso, Sapienza and Zingales (2013) recovered the portions of strategic default and illiquidity-triggered default by using three survey questions: "How many people do you know who have defaulted on their mortgages?"; "Of the people you know who have defaulted on their mortgages, how many do you think walked away even if they could afford to pay the monthly mortgage?"; and "If the value of your mortgage exceeded the value of your house by 50K [100K/150K], would you walk away from your house?" A third method entails relying on econometric identification techniques in structural

[6] The identification strategy will be discussed in detail in Section 4.3.

estimations. Bajari, Chu and Park (2010) estimated a static bivariate model with partial observability for mortgage default. With the first two methods, the proportions of illiquidity-triggered default and strategic default are sensitive to how the two types of default are defined based on observables or how the questions in the surveys are phrased. Moreover, although each defaulter is assigned to a certain type, in fact there remains a positive probability that the defaulter belongs to the other type. My study follows the spirit of the third method. Thus, for each individual borrower, the econometric model will generate a probability that the individual defaults strategically and a probability that the individual defaults as a result of illiquidity.

Incorporating partial observability of dependent variables into a dynamic structural model will provide three advantages. First, it can reduce the estimation bias of structural parameters and generate a more precise prediction of default rates and prepayment rates. Second, it can yield separate predictions of the illiquidity-triggered default rate, strategic default rate, refinancing rate, and moving rate. Third, it can provide better counterfactual analyses, as it allows researchers to study how the separate rates (in addition to the mixed rates) respond to policy interventions or macroeconomic shocks. In this paper, I conduct counterfactual analyses for several loan modification policies aimed at foreclosure mitigation. The results show that writing down the principal can effectively reduce both illiquidity-triggered default and strategic default, that interest rate reduction can effectively reduce illiquidity-triggered default (as it can reduce the monthly payment amount) but cannot effectively reduce strategic default (as it cannot dramatically reduce the financial incentive to default), and that term extension can effectively reduce illiquidity-triggered default but will increase strategic default (as households will have a longer period with low or even negative home equity after term extension).

The most important innovation of this paper with respect to econometric methodology is that I incorporate the partial observability of dependent variables into the CCP method and use exclusion restrictions and identification at infinity arguments as the identification strategy.[7] Arcidiacono and Miller (2011) extended the CCP method to

[7] There is a rich econometric literature on estimating dynamic models. Pakes (1986) estimated a dynamic discrete choice model for the optimal stopping problems of patent renewal. Rust (1987) developed a nested fixed point algorithm to compute the value function in Bellman's equation and used MLE to estimate the

incorporate unobserved heterogeneity of independent variables. Both the CCP method with partial observability of dependent variables and the CCP method with unobserved heterogeneity of independent variables involve estimating a mixture of distributions in the first stage. However, there are significant differences between the two extensions.

First, in the CCP method with unobserved heterogeneity of independent variables, agents in different unobserved exogenous types are solving the same problems. By contrast, in the CCP method with partial observability of dependent variables, agents are solving different problems, but the decisions within these problems are mixed in the observable actions or outcomes. Different distributions in the mixture estimated in the first step are associated with different problems, and they will be used to recover the structural parameters in the corresponding problems in the second step. For example, whether to choose refinance among the choices to default, refinance, or continue to pay and whether to move are different problems, but the decision to refinance and the decision to move are mixed in the prepayment outcome. In addition, in the CCP method with partial observability of dependent variables, the agents could be solving the same problem, but the outcomes can be generated through different mechanisms (censored dependent variables). For example, illiquidity-triggered default is generated by the event that the liquidity constraint binds, whereas strategic default is generated by the mechanism that agents choose default among the choices to continue to pay, default, and refinance because default can provide higher value than the other two choices. Moreover, the action in solving one problem can affect the probability of another outcome in the mixture in the future. For example, the probability of illiquidity-triggered

structural parameters. Hotz and Miller (1993) developed the CCP method, which dramatically reduces the computation burden of computing the value function. Hotz, Miller, Sanders and Smith (1994) developed a simulation-based CCP method, which is suitable for problems with large state space. Aguirregabiria and Mira (2002) developed a recursive CCP method. Arcidiacono, Bayer, Blevins and Ellickson (2014) extended the CCP method to estimate dynamic choice models in continuous time. Bajari, Benkard, and Levin (2007) developed the BBL method, which works for single agent dynamic problems and dynamic games, finite horizon problems and infinite horizon problems, and discrete choice problems and continuous choice problems. More details on this literature can be found in the survey paper by Aguirregabiria and Mira (2010). The econometric techniques of estimating dynamic models are extensively used in the empirical literature, including Gowrisankaran and Rysman (2012), Schiraldi (2011), Song and Chintagunta (2003), Conlon (2012), and Erdem, Keane, Oncu and Strebel (2005) on durable goods purchases; Hendel and Nevo (2006) and Wang (2015) on storable goods purchases; Lee (2013) on two-sided markets; and Murphy (2013) on the housing supply.

default can be lower than before if borrowers chose to refinance to a new loan with lower monthly payment amount.

Second, for single agent dynamic problems (not dynamic games), in the CCP method with unobserved heterogeneity of independent variables, when agents are solving a problem, each agent needs to consider only the probability distribution associated with his or her own exogenous type in the mixture. In the second step of the estimation, the structural parameters are recovered based on this fact. By contrast, in the CCP method with partial observability of dependent variables, when agents solve a problem, they may also need to consider the probability distribution in the mixture associated with other problems or other mechanisms that can generate the same outcomes. For example, when an agent is making decision about whether to strategically default, refinance, or continue to pay, he or she should take into consider that in the future there are a probability that his or her liquidity constraint will be binding (illiquidity-triggered default) and a probability that he or she will need to move. Theoretically, if the agent forecasts that the probability of illiquidity-triggered default in the future is high, then he or she will be more likely to strategically choose to default in the current period even if the liquidity constraint is not binding in the current period. In the second step of the estimation, the structural parameters in the strategic decision problem of whether to default, refinance or continue to pay will be recovered on the basis of these facts.

The methodology of estimating dynamic discrete choice models with partially observable outcomes in this paper is extendable to other dynamic problems. For example, in a dynamic entry and exit game, if we observe that a firm exits or does not enter a certain market, one possible reason is that the firm does not think that entry is profitable based on its expectation regarding future cash flow from this market, whereas another possible reason is that the firm is financially constrained.[8] As another example, if we observe that a firm sells a property, plant, or equipment it owned, one possible reason is that the firm thinks that selling is more profitable than operating based on its expectation regarding future operating cash flows, while another possible reason is that the firm would like to increase its cash holdings.

[8] Using data on the airline industry, Aguirregabiria and Ho (2012) and Benkard, Bodoh-Creed, and Lazarev (2010) estimated dynamic entry/exit games without firm financial constraints. Benoit (1984) built theoretical models for static entry/exit games with firm financial constraints.

Another innovation of my paper is that, besides using the usual method addressing the state transitions in the estimation of dynamic choice models (i.e., assume that households have rational expectations and that the state transitions follow a specific process and use the realized data to estimate the process in the first step), I follow the spirit of Manski (2004) and use data from the Michigan Survey of Consumers on households' subjective expectations about future house prices. This method does not require an assumption of rational expectations for households and a specific process (e.g., a first-order autoregressive (AR(1)) process) for the state transitions to infer households' preferences. Pantano and Zheng (2013) incorporated subjective expectations of future choice probabilities into dynamic choice estimations. By contrast, Erdem, Keane, Öncü and Strebel (2005) incorporate subjective expectations of future state variables into the estimation of dynamic choice models. They used consumer expectation data on personal computer (PC) prices from a survey for the dependent variable and used current and past realized PC price data for the explanatory variables to estimate the state transition processes. Departing from this approach, I directly use the empirical distribution of the house price forecasts in each month by the Ohio respondents in the Michigan Survey of Consumers as the distribution of households' house price forecasts for the next period, which does not require an assumption of a structure for the expectation formation processes.

The remainder of this paper is organized as follows. In Section 2, I construct the dynamic discrete choice model for household mortgage default and prepayment behaviors. Section 3 discusses the data and descriptive statistics. Section 4 explains the econometric methodology and identification strategy. Section 5 discusses the empirical results. In Section 6, I conduct two sets of counterfactual analyses. The first set involves mortgage loan modification policies aimed at foreclosure mitigation. The second set considers the Federal Reserve Comprehensive Capital Analysis and Review (CCAR) scenarios. Then, I conclude in Section 7.

2 The Model

In previous literature estimating static models for mortgage terminations, a measure of the financial incentive to default has typically been defined as in Equation (1), where

$LoanBalance_{i,t}$ is the unpaid balance for the mortgage of household i in period t and $Homeprice_{i,t}$ is the value of the house of household i in period t. A measure of the financial incentive to refinance has typically been defined as in Equation (2), where r_t^m is the current market mortgage interest rate at period t, r_i^c is the coupon rate of household i's existing mortgage, and $payment_i$ is the monthly mortgage payment amount in the future. The first term in Equation (2) can be viewed as the present value of the future mortgage payment flow discounted by the current market interest rate, while the second term is equal to the unpaid balance.

$$defaultOpt_{i,t} = LoanBalance_{i,t} - Homeprice_{i,t} \tag{1}$$

$$refiOpt_{i,t} = \sum_{\tau=t}^{T} payment_i \left[\frac{1}{1+r_t^m}\right]^{\tau-t} - \sum_{\tau=t}^{T} payment_i \left[\frac{1}{1+r_i^c}\right]^{\tau-t} \tag{2}$$

Figure 1 shows how the processes generating mortgage performance outcomes are modeled in this paper. In each period t, there is an exogenous process that determines whether household i faces illiquidity (denoted as td). Illiquidity-triggered default occurs if and only if $I\{z_{i,t}\varphi + e_{i,t}^1 < 0\} = 1$, where $z_{i,t}$ includes the monthly mortgage payment amount, the monthly income, the household size, the credit score, the county-level unemployment rate, a minority indicator, and an indicator of whether the household took a downpayment assistance (DPA) grant. φ are the parameters to be estimated. The probability of illiquidity-triggered default is given by Equation (3).

$$Pr\{Illiquid_{i,t}\} = Pr\{z_{i,t}\varphi + e_{i,t}^1 < 0\} = p_t^{td}(z_{i,t};\varphi) = \frac{exp\{-z_{i,t}\varphi\}}{1+exp\{-z_{i,t}\varphi\}} \tag{3}$$

$z_{i,t}\varphi + e_{i,t}^1 < 0$ can be viewed as the household's liquidity constraint. The unobservable $e_{i,t}^1$ is assumed to be i.i.d. across time and individuals and to follow a logit distribution. The liquidity constraint is more likely to be binding if the monthly expenditure is high. Mortgage payments are part of the monthly expenditure, and household size is positively correlated with the monthly expenditure. The liquidity constraint is less likely to be binding if the household has a higher income, a higher savings level, and higher accessibility to other loans. The household may have a lower savings level if it chose to take a DPA grant. The credit score is a measure of the household's accessibility to other loans. Further, the household is more likely to encounter a shortfall of income if the local unemployment rate is high.

If illiquidity-triggered default does not occur, then the household will decide whether to move, sell the current house, and prepay the mortgage (denoted as m). The household will have to move and prepay the mortgage if $I\{w_{i,t}\psi + e_{i,t}^2 < 0\} = 1$, where $w_{i,t}$ includes the difference between the county and national employment growth rate, a minority indicator, an indicator for whether the household took a DPA grant, the household size, and the household age. ψ are the parameters to be estimated. I assume that if both $I\{z_{i,t}\varphi + e_{i,t}^1 < 0\} = 1$ and $I\{w_{i,t}\psi + e_{i,t}^2 < 0 < 0\} = 1$, then illiquidity-triggered default occurs rather than move. For simplicity, I also assume that e^1 and e^2 are independent from each other. The probability of move is given by Equation (4)[9].

$$\begin{aligned} Pr\{Move_{i,t}\} &= Pr\{z_{i,t}\varphi + e_{i,t}^1 \geq 0 \cap w_{i,t}\psi + e_{i,t}^2 < 0\} = p_t^m(z_{i,t}, w_{i,t}; \varphi, \psi) \\ &= Pr\{z_{i,t}\varphi + e_{i,t}^1 \geq 0\} Pr\{w_{i,t}\psi + e_{i,t}^2 < 0\} = \frac{exp\{z_{i,t}\varphi\}}{1+exp\{z_{i,t}\varphi\}} \frac{1}{1+exp\{w_{i,t}\psi\}} \end{aligned} \quad (4)$$

Archer, Ling, and McGill (1996) categorized the mobility-driven factors related to mortgage termination into two groups: location-decision factors and response-to-housing-disequilibrium factors. I use the past-12-month moving average of the difference between the county and the national employment growth rate as a location-decision factor. Households may want to move for new opportunities if there are more opportunities in other places than in their current residing location. Household size belongs to both groups of factors. First, larger households have less mobility; second, households of different sizes have different

[9] Similar to the decision to default strategically and the decision to refinance, the decision to move should also be a dynamic choice by forward-looking households. Using data from the National Longitudinal Survey of Youth (NLSY), Kennan and Walker (2011) estimated a dynamic discrete choice model for individual migration decisions. It is impossible to estimate a dynamic discrete choice model for moving with the observables in the loan performance data. Therefore, I model moving as a static choice and assume that it is separate from the choices to strategically default, refinance, or continue to pay on the mortgage.

probabilities of restructuring. The empirical mobility literature, including Quigley (1987) who used the PSID data and Ferreira, Gyourko and Tracy (2010) who used the American Housing Survey (AHS) data, has provided evidence that larger households, older households, and households with minorities have a significantly lower probability of moving.

I also include an indicator for whether the household took a DPA grant as a variable related to the household's moving probability. In the mortgage sample analyzed in this paper, a borrower could take a grant up to 3% of the home's purchase price to pay for the down payment, origination fee, appraisal fee, and other closing costs. If the borrower took a DPA grant, his or her interest rate would be increased by 0.5%. Taking a grant is similar to choosing a mortgage with low points and a high interest rate, while not taking a grant is similar to choosing a mortgage with high points and a low interest rate.[10] A borrower expecting to move in the near future would choose a mortgage with low points and a high interest rate, as he or she would not be able to continue to take advantage of a low interest rate after moving if he or she chose a mortgage with high points and a low interest rate. By contrast, a borrower expecting to stay in the house over the mortgage horizon would choose a mortgage with high points and a low interest rate. Stanton and Wallace (1998) derived a separating equilibrium model in which borrowers self-select the points and interest rate.[11]

If neither illiquidity nor move occurs in the previous two processes, the household will have the opportunity to make a choice among continuing to pay on the mortgage

---

[10] Generally, when a borrower initiates a typical mortgage, the lender will give him a list with different combinations of points and interest rates. Paying high points at the beginning will lower the interest rate over the entire mortgage horizon.

[11] In most mortgage loan performance data sets, points are not observable, and only the interest rates are observable. Because of this data limitation, some empirical papers, such as Pavlov (2001), have used a two-step regression for the prepayment decision accounting the effect of the unobserved point choice. In the first step, they ran the regression of the difference between the individual mortgage interest rate and the market-prevailing rate on the characteristics of the loan and the household, such as the loan-to-value (LTV) ratio, payment-to-income ratio, and credit score. Then, they used the residual as a proxy of the points for the second step (the regression for the prepayment choice). A higher residual in the first step indicates a higher interest rate relative to the market and implies that the borrower may choose low points. However, as noted by Bajari, Chu and Park (2010), the problem with this method in the existing literature is that the first-step estimation is subject to endogeneity bias, which will render the estimation in the second step inconsistent. Borrowers with a high payment-to-income ratio and a high LTV should be offered high interest rates as they have a higher risk of default; however, borrowers offered low interest rates may be more willing to borrow more and consequently have a higher payment-to-income ratio and a higher LTV. One advantage of the loan performance data set used in this paper is that it contains information about whether borrowers take a DPA grant. Therefore, I do not need to use estimated points, which contain endogeneity bias.

(denoted as c), refinance (denoted as r), or strategic default (denoted as sd). The probability that the household has a chance to make a choice on the mortgage is given by Equation (5).

$$\begin{aligned} p_t^3\left(z_{i,t}, w_{i,t}; \varphi, \psi\right) &= Pr\left\{z_{i,t}\varphi + e_{i,t}^1 \geq 0 \cap w_{i,t}\psi + e_{i,t}^2 < 0\right\} \\ &= 1 - p_t^{td}\left(z_{i,t}; \varphi\right) - p_t^m\left(z_{i,t}, w_{i,t}; \varphi, \psi\right) = \frac{exp\{z_{i,t}\varphi\}}{1+exp\{z_{i,t}\varphi\}} \frac{exp\{w_{i,t}\psi\}}{1+exp\{w_{i,t}\psi\}} \end{aligned} \tag{5}$$

The household's intertemporal utility-maximizing problem given that the household has the opportunity to make a choice among continuing to pay, refinance (denoted as r), or strategic default is formulated in Equation (6).[12] $s_{i,t}$ are all the state variables, and $z_{i,t}, w_{i,t} \subset s_{i,t}$. The transition of the state variables is given by Equation (7), where the state variables in the next period $s_{i,t+1}$ are functions of the state variables and actions in the current period, as well as random shocks in the next period.

$$V_t\left(s_{i,t}\right) = E_{\varepsilon|s_{i,t}} \max_{a_{i,t} \in \{sd,r,c\}} \begin{Bmatrix} u_{i,t}\left(s_{i,t}. a_{i,t} = c\right) + \beta E_t\left[W_{t+1}\left(s_{i,t+1}\right)|s_{i,t}, a_{i,t} = c\right], \\ u_{i,t}\left(s_{i,t}. a_{i,t} = r\right) + \beta E_t\left[W_{t+1}\left(s_{i,t+1}\right)|s_{i,t}, a_{i,t} = r\right], \\ v_{i,t}\left(s_{i,t}. a_{i,t} = sd\right) \end{Bmatrix} \tag{6}$$

$$s.t. \ s_{i,t+1} = f_{i,t}(s_{i,t}, a_{i,t}, \xi_{i,t+1}) \tag{7}$$

$$\begin{gathered} u_{i,t}\left(s_{i,t}, a_{i,t} = c\right) = -\alpha_1 \cdot payment_i + \alpha_2 \cdot HouseStock_i + \varepsilon_{i,t}^c \\ u_{i,t}\left(s_{i,t}, a_{i,t} = r\right) = -\alpha_3 - \alpha_1 \cdot payment\_refi_{i,t} + \alpha_2 \cdot HouseStock_i + \varepsilon_{i,t}^r \\ v_{i,t}\left(s_{i,t}, a_{i,t} = sd\right) = -\alpha_4 - \alpha_5 \cdot HomeValue_{i,t} + \varepsilon_{i,t}^{sd} \end{gathered} \tag{8}$$

$$\begin{gathered} W_{t+1}\left(s_{i,t+1}\right) = p_{t+1}^{td}\left(z_{i,t+1}; \varphi\right) V_{t+1}^{td}\left(s_{i,t+1}\right) + p_{t+1}^m\left(z_{i,t+1}, w_{i,t+1}; \varphi, \psi\right) V_{t+1}^m\left(s_{i,t+1}\right) \\ + p_{t+1}^3\left(z_{i,t+1}, w_{i,t+1}; \varphi, \psi\right) V_{t+1}\left(s_{i,t+1}\right) \end{gathered} \tag{9}$$

[12] This is a partial equilibrium model that considers only households' default decisions given the house price level. It does not account for the fact that households' default decisions can also affect the house price level. Because general equilibrium models are typically difficult to estimate, macroeconomists usually use calibration to analyze general equilibrium models. For instance, Corbae and Quintin (2015) calibrated a dynamic general equilibrium model in which heterogeneous households make decisions regarding mortgage selection and default. Although I estimate a partial equilibrium model, I include local economic variables in the estimation to mitigate potential endogeneity.

The choice-specific period utility functions $u_{i,t}(s_{i,t}, a_{i,t} = c)$ and $u_{i,t}(s_{i,t}, a_{i,t} = r)$ are specified in the first two lines in Equations (8). $payment_i$ is the monthly mortgage payment amount of household i 's existing mortgage. $payment\_refi_{i,t}$ is the new monthly payment amount if the household chooses to refinance in this period. I assume that if the household refinances, the new interest rate will be the prevailing interest rate in the current mortgage market, the new loan amount that the household borrows will be equal to the unpaid balance of the old mortgage, and the term of the new mortgage will be equal to the remaining horizon of the old mortgage. If the household refinances or continues to pay, it will obtain disutility from making monthly payments and obtain utility from living in its home. The utility of living in the home is positively related to the housing stock. I use the purchase price of the house divided by the CPI at purchase as a measure of the housing stock.[13] [14] The intercept α_3 in the period utility functions of refinance captures the unobserved fixed transaction costs for refinancing.[15]

The termination value of strategic default $v_{i,t}(s_{i,t}, a_{i,t} = sd)$ is specified in the third line of Equations (8). If the household defaults, it will lose its home; thus, $HomeValue_{i,t}$ is

[13] In the period utility functions, first, I assume that the coefficient of the mortgage payment amount for continuing to pay is the same as that for refinancing, as the household should obtain the same amount of disutility from paying a dollar no matter whether it chooses to refinance or continue to pay. Second, I assume that the coefficient of home quality for continuing to pay is the same as that for refinancing, as regardless of whether the household refinances or continues to pay, it lives in the same house and should thus obtain the same servicing utility from the house.

[14] In the literature on labor economics, using panel data with individuals' consumption levels, such as the PSID and the Health and Retirement Study (HRS), some papers, such as Blau (2008), have estimated dynamic life cycle models in which the period utility is a function of consumption. In the literature on housing economics, Bajari, Chan, Krueger and Miller (2013) used the PSID data and estimated a dynamic model of housing demand where the period utility is a function of nondurable consumption and the housing stock. However, for mortgage default and prepayment, it is difficult to estimate a dynamic life cycle model with a period utility function containing consumption because the course of consumption is not observed in the loan performance data. Bajari, Chu, Nekipelov and Park (2013) and Carranza and Navarro (2010) did not include consumption in their period utility functions. Laufer (2011) and Zhang (2010) used some assumptions to imply individuals' consumption level and included it in their period utility functions. Laufer (2011) matched the consumption levels of households in the PSID to the households in his loan performance data with similar characteristics. Zhang (2010) assumed that the households have zero savings. In the literature on industrial organization, the period utility functions of consumers in dynamic choice models typically do not include consumption.

[15] There is a rich literature estimating dynamic models with an intercept capturing the unobserved switching cost or transaction cost. Shcherbakov (2009) estimated a dynamic model with the switching cost for consumers to change cable television providers. Ho (2010) estimated a dynamic model with the switching cost for consumers to change banks for deposits. Donna (2012) estimated a dynamic model with the switching cost for people to switch their travel modes between private cars and public transportation. Schiraldi (2011) estimated a dynamic model with the transaction cost for automobile replacement.

the current market value of the house, which is calculated as the purchase price divided by the MSA-level house price index at the time of purchase and multiplied by the current MSA-level house price index. The intercept α_4 captures other unobserved default costs, such the costs of moving and being socially stigmatized.[16]

$V_t \ (s_{i,t})$ in Equation (6) is the ex ante expected value function before the choice is made. Illiquidity-triggered default, moving, and strategic default are terminating actions after which the household no longer makes choices in the model. When the household is making the choice to continue to pay the mortgage, refinance, or strategically default, it takes into consideration that in the next period there will still be positive probabilities of illiquidity-triggered default and move; $V_{t+1}^{td}(s_{i,t+1})$ and $V_{t+1}^{m}(s_{i,t+1})$ are the values associated with the illiquidity-triggered default event and the move event, respectively. There will also be a positive probability that neither illiquidity-triggered default nor move occurs in the next period and the household has the opportunity to make decisions of continuing to pay, refinance, or strategic default; $V_{t+1}(s_{i,t+1})$ is the ex ante expected value associated with such a scenario. Therefore, $W_{t+1}(s_{i,t+1})$, the value for the future, is defined as in Equation (9).

3 Data and Descriptive Statistics

3.1 Loan Performance Data

The mortgage loan performance data come from the Mortgage Revenue Bonds data maintained by the Ohio Housing Finance Agency (OHFA).[17] In my sample, there are 20,487 mortgages closed between 2005 and 2008, all borrowed by Ohio low-to-moderate income, first-time homebuyers with 30-year fixed interest rates. The data contain the monthly payment history through March 2011.[18]

[16] Glover (2015) estimated a dynamic capital structure model to recover the unobserved default costs of firms on debt payments.

[17] As defined by Freddie Mac, "Mortgage Revenue Bonds (MRBs) are tax-exempt bonds that state and local governments issue through housing finance agencies (HFAs) to help fund below-market-interest-rate mortgages for first-time qualifying homebuyers". "Eligible borrowers are first-time homebuyers with low to moderate incomes below 115 percent of median family income". "Originating lenders pool the mortgages into securities guaranteed by either Ginnie Mae or Freddie Mac and sell the securities to the issuing HFAs". "As part of its corporate investment program, Freddie Mac purchases MRBs issued by HFAs".

[18] Details about this data set can be found in Ergungor and Moulton (2014) and Zhang (2010).

Following most of the literature, I define default as more than 90 days past due. As shown in Table I, up to March 2011, 11.25% of the borrowers defaulted, and 10.91% prepaid. The data include information on the borrower's credit score and monthly income and the price of the house at origination, but changes in this information are not tracked over time. The data also contain mortgage characteristics such as the coupon interest rate, loan amount, monthly payment amount and initial LTV, as well as demographic characteristics such as household size, location, race, and ethnicity.[19] Descriptive statistics for the important variables are displayed in Table II.

3.2 Macroeconomic Data

The monthly MSA-level house price index data come from the Freddie Mac Housing Price Index (FMHPI). The monthly mortgage market interest rate data come from the Freddie Mac Primary Mortgage Market Survey (PMMS). Using Equation (1) and Equation (2), together with the FMHPI, the PMMS, and the individual mortgage data, I can compute $defaultOpt_{i,t}$ and $refiOpt_{i,t}$; the descriptive statistics are displayed in Table II.[20] I also extract the monthly county-level unemployment rate and employment growth rate from the Bureau of Labor Statistics (BLS).

3.3 Michigan Survey of Consumers

The Michigan Survey of Consumers asks approximately 500 people each month about their expectations regarding some economic variables in the future. Starting in January 2007, in each month, the Michigan Survey of Consumers asked subjects to provide their expectations regarding the house price growth rates over the next year and over the next five years. On

[19] Because this is a very special sample (low-to-moderate income first-time homebuyers) rather than a random sample drawn from the whole population, the result from this sample cannot be extended to the whole population. However, the model and methodology developed in this paper are extendable to other loan performance data sets, such as the data maintained by Corelogic, which contains almost the whole population of mortgages. Moreover, because the mortgages in this sample are pooled into the Mortgage Revenue Bonds Securities, purchased by Freddie Mac, and traded in the secondary market, the results of default and prepayment risks obtained from this sample are nevertheless helpful for pricing the Mortgage Revenue Bonds Securities. The advantage of this data set is that it contains information on some characteristics of households, such as household size, which other loan performance data do not have.

[20] The loan performance data do not track the course of the price for an individual house over time. I use the purchase price divided by the house price index at purchase and multiplied by the current house price index as an approximation of $Homeprice_{i,t}$.

average, each month there are 22 people from Ohio reporting their responses to the house price expectation questions. I use these data to represent the future house price expectations of the households in my mortgage loan performance data, who also reside in Ohio.

4 Econometric Methodology

I extend the CCP method to estimate a dynamic discrete choice model with partial observability of the dependent variables for household mortgage default and prepayment behaviors. Exclusion restrictions and identification at infinity arguments provide the identification conditions. In the first stage, the parameters in the probabilities of illiquidity-triggered default and moving (exogenous events) and the parameters in the conditional choice probabilities of strategic default, refinancing, and continuing to pay (policy functions) are estimated. The parameters of the state transition processes are also estimated. In the second stage, I estimate the structural parameters in the households' intertemporal utility-maximizing problem of whether to default, refinance, or continue to pay.

Denote $\tilde{u}_t^k(x_{i,t};\alpha)$ as $u_{i,t}(x_{i,t}, a_{i,t} = k; \alpha)$ net of current idiosyncratic shocks, where $k \in \{c, r\}$. Denote $\tilde{v}_t^{sd}(x_{i,t};\alpha)$ as $v_{i,t}(x_{i,t}, a_{i,t} = sd; \alpha)$ net of current idiosyncratic shocks. Denote $v_t^k(s_{i,t})$ as the choice-specific value functions net of current idiosyncratic shocks, where $k \in \{c, r, sd\}$. Then,

$$\begin{aligned} v_t^c(s_{i,t}) &= \tilde{u}_t^c(x_{i,t};\alpha) + \beta E_t\left[W_{t+1}(s_{i,t+1}) | s_{i,t}, a_{i,t} = c\right], \\ v_t^r(s_{i,t}) &= \tilde{u}_t^r(x_{i,t};\alpha) + \beta E_t\left[W_{t+1}(s_{i,t+1}) | s_{i,t}, a_{i,t} = r\right], \\ v_t^{sd}(s_{i,t}) &= \tilde{v}_t^{sd}(x_{i,t};\alpha) \end{aligned} \tag{10}$$

Following most literature, I assume that the idiosyncratic shocks follow a Type I extreme value distribution. Given the state variables $s_{i,t}$ and conditional on the household not facing illiquidity or moving, the equilibrium conditional choice probabilities of continuing to pay, refinancing, and defaulting can be written as:

$$\sigma_t^c(s_{i,t}) = \Pr(a_{i,t} = c) = \frac{exp\left(v_t^c(s_{i,t})\right)}{\sum_{k\in\{c,r,sd\}} exp\left(v_t^k(s_{i,t})\right)}$$

$$\sigma_t^r(s_{i,t}) = \Pr(a_{i,t} = r) = \frac{exp\left(v_t^r(s_{i,t})\right)}{\sum_{k\in\{c,r,sd\}} exp\left(v_t^k(s_{i,t})\right)}$$

$$\sigma_t^{sd}(s_{i,t}) = \Pr(a_{i,t} = sd) = \frac{exp\left(v_t^{sd}(s_{i,t})\right)}{\sum_{k\in\{c,r,sd\}} exp\left(v_t^k(s_{i,t})\right)}$$

By the Hotz-Miller inversion, the following can be derived:

$$v_t^c(s_{i,t}) - v_t^{sd}(s_{i,t}) = log\left(\frac{\sigma_t^c(s_{i,t})}{\sigma_t^{sd}(s_{i,t})}\right)$$

$$v_t^r(s_{i,t}) - v_t^{sd}(s_{i,t}) = log\left(\frac{\sigma_t^r(s_{i,t})}{\sigma_t^{sd}(s_{i,t})}\right)$$

Given the extreme value assumption for the idiosyncratic error terms, the ex ante value function $V_t\ (s_{i,t})$ can be written as in Equation (11), where $\sigma_t^{sd}(s_{i,t})$ can be obtained in the first-stage estimation and the functional form of $v_t^{sd}(s_{i,t})$ can be specified by researchers. Thus, $V_t\ (s_{i,t})$ can be recovered for the second-stage estimation.

$$\begin{aligned} V_t\ (s_{i,t}) &= log\left(exp\left(v_t^c(s_{i,t})\right) + exp\left(v_t^r(s_{i,t})\right) + exp\left(v_t^{sd}(s_{i,t})\right)\right) \\ &= log\left(\frac{1}{\sigma_t^{sd}(s_{i,t})}\right) + v_t^{sd}(s_{i,t}) \end{aligned} \tag{11}$$

4.1 First-stage Estimation

4.1.1 Estimating the Policy Functions

First, the probability of observing a default is equal to the sum of the probability of illiquidity-triggered default and the probability of strategic default. Second, the probability of observing a prepayment is equal to the sum of the probability of moving and the probability of refinancing. Third, the probability of observing a household continuing to pay is equal to the probability that neither illiquidity nor move occurs times the conditional choice probability of the household choosing to continue to pay given that it has the opportunity to make the decision. The probabilities of these three observable events can be written as in Equations (12).

$$
\begin{aligned}
Pr\{default_{i,t}\} &= p_t^{td}(z_{i,t},;\varphi) + \sigma_t^{sd}(s_{i,t};\gamma)\cdot p_t^3(z_{i,t},w_{i,t};\varphi,\psi)\\
Pr\{prepay_{i,t}\} &= p_t^{m}(z_{i,t},w_{i,t};\varphi,\psi) + \sigma_t^{r}(s_{i,t};\gamma)\cdot p_t^3(z_{i,t},w_{i,t};\varphi,\psi)\\
Pr\{continue_{i,t}\} &= \sigma_t^{c}(s_{i,t};\gamma)\cdot p_t^3(z_{i,t},w_{i,t};\varphi,\psi)
\end{aligned}
\tag{12}
$$

$$
\begin{aligned}
logL = \sum_{i=1}^{N}\sum_{t=1}^{T_i}\Big[& I\{default_{i,t}\}\log\big(Pr\{default_{i,t}\}\big) + I\{prepay_{i,t}\}\log\big(Pr\{prepay_{i,t}\}\big)\\
&+ I\{continue_{i,t}\}\log(Pr\{continue_{i,t}\})\Big]
\end{aligned}
\tag{13}
$$

The log likelihood function of the first-stage estimation is shown in Equation (13). I use the MLE to estimate φ, ψ, and γ. The estimation results are denoted as $\{\hat{\varphi},\hat{\psi},\hat{\gamma}\}$; then, I obtain $p_t^{td}(z_{i,t},;\hat{\varphi})$, $p_t^{m}(z_{i,t},w_{i,t};\hat{\varphi},\hat{\psi})$, $p_t^{3}(z_{i,t},w_{i,t};\hat{\varphi},\hat{\psi})$, $\sigma_t^{sd}(s_{i,t};\hat{\gamma})$, $\sigma_t^{r}(s_{i,t};\hat{\gamma})$, and $\sigma_t^{c}(s_{i,t};\hat{\gamma})$.

If the data on aggregate-level illiquidity-triggered default rates, strategic default rates, moving rates, and refinance rates in each period are available from other sources, they may bring more information into the estimation. The first-stage estimation can use the generalized method of moments (GMM) with additional moment conditions that match $\sum_i p_t^{td}(z_{i,t};\varphi)$, $\sum_i \sigma_t^{sd}(s_{i,t};\gamma)p_t^3(z_{i,t},w_{i,t};\varphi,\psi)$, $\sum_i p_t^{m}(z_{i,t},w_{i,t};\varphi,\psi)$, and $\sum_i \sigma_t^{r}(s_{i,t};\gamma)p_t^3(z_{i,t},w_{i,t};\varphi,\psi)$ to the corresponding aggregate-level rates, respectively.[21] As shown in Imbens and Lancaster (1994), Berry, Levinsohn and Pakes (2004), and Petrin (2002), using moment conditions generated from both micro-level data and macro level data

[21] I cannot match $\sum_i p_t^{td}(z_{i,t};\varphi)$and $\sum_i \sigma_t^{sd}(s_{i,t};\gamma)p_t^3(z_{i,t},w_{i,t};\varphi,\psi)$, respectively, to the aggregate illiquidity-triggered default rates and aggregate strategic default ratesin Gerardi, Herkenhoff, Ohanian, and Willen (2013), Tirupattur, Chang and Egan (2010), or Guiso, Sapienza and Zingales (2013). First, these illiquidity-triggered default rates and strategic default rates are sensitive to how these two types of default are defined based on observables or how questions in surveys are phrased. Second, the sample in this paper is not comparable to the samples in those papers. I also cannot match $\sum_i \sigma_t^{r}(s_{i,t};\gamma)p_t^3(z_{i,t},w_{i,t};\varphi,\psi)$ to a proper aggregate-level rate. The aggregate-level migration rates from one country to other countries can be created from the Current Population Survey (CPS), the American Community Survey (ACS), and the Census. However, the moving event considered in this paper includes not only migration across counties but also changing houses within a county.

can improve the efficiency of GMM estimators. However, GMM can only be as efficient as MLE when there are optimal instrumental variables. Alternatively, the first-stage estimation can use MLE with the constraints that $\sum_i p_t^{td}(z_{i,t};\varphi)$, $\sum_i \sigma_t^{sd}(s_{i,t};\gamma)p_t^3(z_{i,t},w_{i,t};\varphi,\psi)$, $\sum_i p_t^m(z_{i,t},w_{i,t};\varphi,\psi)$, and $\sum_i \sigma_t^r(s_{i,t};\gamma)p_t^3(z_{i,t},w_{i,t};\varphi,\psi)$ equal the corresponding aggregate-level rates, respectively. However, first, this is equivalent to the GMM with extremely high weights on those additional moment conditions; second, when there are many parameters to search for and the constraints are irregular, numerical optimizers will have a much less stable performance for optimization with constraints than for optimization without constraints.

4.1.2 Estimating the State Transitions

The state transitions are given by $s_{i,t+1} = f_{i,t}(s_{i,t},a_{i,t},\xi_{i,t+1})$. There are four classes of state variables, for which I model the transitions in different ways.

First, I model the log difference of the house price hp and the innovation of the mortgage market average interest rate r following AR(1) processes as in Equations (14). The results of Dickey-Fuller tests show that the time series of the house price and interest rate have unit roots. Therefore, I model the changes in these two variables rather than their levels as AR(1). Second, some state variables, such as race, the DPA grant indicator, and the monthly payment amount if the household does not refinance, do not change over time. I also assume that income, household size, and credit score do not change over time because I observe them only when the mortgage was originated. Third, the transitions of some state variables, such as age and the unpaid balance, are deterministic. Fourth, some state variables are functions of other state variables. For example, $defaultOpt_{i,t}$ is a function of the unpaid balance, purchase price, house price index, and time to maturity, and $refiOpt_{i,t}$ is a function of the unpaid balance, monthly payment, current market mortgage interest rate, and time to maturity. Once I have the state transitions of the house price index and interest rate, the state transitions of $defaultOpt_{i,t}$ and $refiOpt_{i,t}$ will be implied.

$$\begin{aligned} \Delta \log hp_{t+1} &= \lambda^0 + \lambda^1 \Delta \log hp_t + \xi^1_{t+1} \\ \Delta r_{t+1} &= \lambda_2 \Delta r_t + \xi^2_{t+1} \end{aligned} \tag{14}$$

4.2 Second-stage Estimation

Given $s_{i,t}$ and given that household i chooses action $a_{i,t} = k \in \{c, r\}$ at period t, I simulate $s^{k,(j)}_{i,t+1}$ for J times. The simulations are based on the result of the estimation of the state transition models in Equations (14). Given certain values of the structural parameters α, for each simulation, using $\hat{\gamma}$ in $\sigma^{sd}_t(s_{i,t}; \gamma)$ estimated in the first stage and the specified parametric form of $v^{sd}_{t+1}(s_{i,t+1}; \alpha)$ and employing Equation (15), I recover $V_{t+1}\left(s^{k,(j)}_{i,t+1}\right)$, i.e., the ex ante value functions of period $t+1$ if the household chooses action $a_{i,t} = k \in \{c, r\}$ in period t. Then, I average all of the simulations to obtain the expected ex ante value of period $t+1$ multiplied by the probability that neither illiquidity nor move occurs in period $t+1$ conditional on the state and the action in period t, as expressed in Equation (16).

$$V_{t+1}\left(s^{k,(j)}_{i,t+1}; \alpha\right) = log\left(\frac{1}{\sigma^{sd}_{t+1}\left(s^{k,(j)}_{i,t+1}; \widetilde{\gamma}\right)}\right) + v^{sd}_{t+1}\left(x^{k,(j)}_{i,t+1}; \alpha\right), \ k \in \{c, r\} \tag{15}$$

$$\begin{aligned} &\hat{E}_t\left[p^3_{t+1}(z_{i,t+1}, w_{i,t+1}) V_{t+1}(s_{i,t+1}) | s_{i,t}, a_{i,t} = k\right] \\ &\qquad = \frac{1}{J}\sum_{j=1}^{J} p^3_{t+1}\left(z^{k,(j)}_{i,t+1}, w^{k,(j)}_{i,t+1}; \hat{\varphi}, \hat{\psi}\right) V_{t+1}\left(s^{k,(j)}_{i,t+1}; \alpha\right), \quad k \in \{c, r\} \end{aligned} \tag{16}$$

Similarly, using $\hat{\varphi}, \hat{\psi}$ in $p^{td}_t(z_{i,t}; \varphi)$ and $p^m_{t+1}(z_{i,t+1}, w_{i,t+1}; \varphi, \psi)$ estimated in the first stage and the specified parametric form of $V^{td}_{t+1}(s_{i,t+1}; \alpha)$ and $V^m_{t+1}(s_{i,t+1}; \alpha)$, I calculate $\hat{E}_t\left[p^{td}_{t+1}(z_{i,t+1}) V^{td}_{t+1}(s_{i,t+1}) | s_{i,t}, a_{i,t} = k\right]$ and $\hat{E}_t\left[p^m_{t+1}(z_{i,t+1}, w_{i,t+1}) V^m_{t+1}(s_{i,t+1}) | s_{i,t}, a_{i,t} = k\right]$ according to Equation (17) and Equation (18), respectively.

$$\hat{E}_t\left[p_{t+1}^{td}(z_{i,t+1})V_{t+1}^{td}(s_{i,t+1})|s_{i,t},a_{i,t}=k\right]$$
$$=\frac{1}{J}\sum_{j=1}^{J}p_{t+1}^{td}\left(z_{i,t+1}^{k,(j)};\hat{\varphi}\right)V_{t+1}^{td}\left(s_{i,t+1}^{k,(j)};\alpha\right),\qquad k\in\{c,r\} \tag{17}$$

$$\hat{E}_t\left[p_{t+1}^{m}(z_{i,t+1},w_{i,t+1})V_{t+1}^{m}(s_{i,t+1})|s_{i,t},a_{i,t}=k\right]$$
$$=\frac{1}{J}\sum_{j=1}^{J}p_{t+1}^{m}\left(z_{i,t+1}^{k,(j)},w_{i,t+1}^{k,(j)};\hat{\varphi},\hat{\psi}\right)V_{t+1}^{m}\left(s_{i,t+1}^{k,(j)};\alpha\right),\quad k\in\{c,r\} \tag{18}$$

Then, I obtain $\hat{E}_t\left[W_{t+1}(s_{i,t+1})|s_{i,t},a_{i,t}=k\right]$ as:

$$\begin{aligned}\hat{E}_t&\left[W_{t+1}(s_{i,t+1})|s_{i,t},a_{i,t}=k\right]\\ &=\hat{E}_t\left[p_{t+1}^{td}(z_{i,t+1})V_{t+1}^{td}(s_{i,t+1})|s_{i,t},a_{i,t}=k\right]\\ &+\hat{E}_t\left[p_{t+1}^{m}(z_{i,t+1},w_{i,t+1})V_{t+1}^{m}(s_{i,t+1})|s_{i,t},a_{i,t}=k\right]\\ &+\hat{E}_t\left[p_{t+1}^{3}(z_{i,t+1},w_{i,t+1})V_{t+1}(s_{i,t+1})|s_{i,t},a_{i,t}=k\right],\qquad k\in\{c,r\}\end{aligned}$$

I plug $\hat{E}_t\left[W_{t+1}(s_{i,t+1})|s_{i,t},a_{i,t}=k\right]$ into Equations (10) and obtain:

$$\begin{aligned}\hat{v}_t^c(s_{i,t};\alpha)&=\tilde{u}_t^c(x_{i,t};\alpha)+\beta\hat{E}_t\left[W_{t+1}(s_{i,t+1})|s_{i,t},a_{i,t}=c\right]\\ \hat{v}_t^r(s_{i,t};\alpha)&=\tilde{u}_t^r(x_{i,t};\alpha)+\beta\hat{E}_t\left[W_{t+1}(s_{i,t+1})|s_{i,t},a_{i,t}=r\right]\\ \hat{v}_t^{sd}(s_{i,t;}\alpha)&=\tilde{v}_t^{sd}(x_{i,t};\alpha)\end{aligned} \tag{19}$$

Next, the probabilities of the observable outcomes (default, prepayment, and continuing to pay) for the second-stage estimation are formulated as in Equations (20), where $\hat{\phi}$, $\hat{\psi}$ and $\hat{\gamma}$ are the parameters estimated in the first stage and α are the structural parameters to be estimated in the second stage. The second-stage estimation is also performed by the MLE, as expressed in Equation (21).

$$\widehat{Pr}\{default_{i,t}\} = p_t^{td}(z_{i,t},;\hat{\varphi}) + p_t^3(z_{i,t}, w_{i,t};\hat{\varphi},\hat{\psi})\frac{exp\{\hat{v}_t^{sd}(s_{i,t};\alpha)\}}{\sum_{k\in\{c,r,sd\}}\{exp\{\hat{v}_t^k(s_{i,t};\alpha)\}\}}$$

$$\widehat{Pr}\{prepay_{i,t}\} = p_t^m(z_{i,t}, w_{i,t};\hat{\varphi},\hat{\psi}) + p_t^3(z_{i,t}, w_{i,t};\hat{\varphi},\hat{\psi})\frac{exp\{\hat{v}_t^r(s_{i,t};\alpha)\}}{\sum_{k\in\{c,r,sd\}}\{exp\{\hat{v}_t^k(s_{i,t};\alpha)\}\}} \tag{20}$$

$$\widehat{Pr}\{continue_{i,t}\} = p_t^3(z_{i,t}, w_{i,t};\hat{\varphi},\hat{\psi})\frac{exp\{\hat{v}_t^{sd}(s_{i,t};\alpha)\}}{\sum_{k\in\{c,r,sd\}}\{exp\{\hat{v}_t^k(s_{i,t};\alpha)\}\}}$$

$$\max_{\alpha|\{\hat{\varphi},\hat{\psi},\hat{\gamma}\}} \quad logL = \sum_{i=1}^{N}\sum_{t=1}^{T_i}\left[I\{default_{i,t}\}\log\left(\widehat{Pr}\{default_{i,t}\}\right)\right. \\ \left. + I\{prepay_{i,t}\}\log\left(\widehat{Pr}\{prepay_{i,t}\}\right) + I\{continue_{i,t}\}\log\left(\widehat{Pr}\{continue_{i,t}\}\right)\right] \tag{21}$$

In fact, $p_{t+1}^{td}\left(z_{i,t+1}^{k,(j)};\hat{\varphi}\right)$ and $p_{t+1}^{m}\left(z_{i,t+1}^{k,(j)}, w_{i,t+1}^{k,(j)};\hat{\varphi},\hat{\psi}\right)$ are very close to zero as default, prepayment, and moving are all rare events. Thus, there is not sufficient variation in $p_{t+1}^{td}(z_{i,t+1})V_{t+1}^{td}(s_{i,t+1})$and $p_{t+1}^{m}(z_{i,t+1}, w_{i,t+1})V_{t+1}^{m}(s_{i,t+1})$ across individual households and periods. This causes that the estimation results of the parameters in$V_{i,t+1}^{td}(s_{i,t+1};\alpha)$and $V_{i,t+1}^{m}(s_{i,t+1};\alpha)$ are unstable, although those parameters are identifiable under certain conditions. Therefore, I set $\hat{E}_t\left[p_{t+1}^{td}(z_{i,t+1})V_{t+1}^{td}(s_{i,t+1})|s_{i,t}, a_{i,t} = k\right] = 0$ and $\hat{E}_t[p_{t+1}^{m}(z_{i,t+1}, w_{i,t+1})V_{t+1}^{m}(s_{i,t+1})|s_{i,t}, a_{i,t} = k] = 0$, where $k \in \{c, r\}$, and accordingly, there is no need to estimate the parameters in $V_{t+1}^{td}(s_{i,t+1};\alpha)$ and $V_{t+1}^{m}(s_{i,t+1};\alpha)$. The assumption underlying this approximation is that when making decisions in the current period on whether to default, refinance, or continue to pay, most people in normal status actually may not think about what if either illiquidity or move occurs in the future if they continue to pay or refinance in the current period because the probabilities of those events are very small. There could be a small number of exceptions. For example, a worker might anticipate unemployment because his or her factory is closing down and then stop making mortgage payments immediately before the unemployment spell occurs. However, such borrowers cannot be identified with the observables in the data.

I also conduct Monte Carlo simulations where the data for estimation are artificially generated by high $p_{t+1}^{td}(z_{i,t+1};\varphi)$ and $p_{t+1}^{m}(z_{i,t+1}, w_{i,t+1};\varphi,\psi)$. In this case, the estimation results for the parameters in $V_{t+1}^{td}(s_{i,t+1};\alpha)$ and $V_{t+1}^{m}(s_{i,t+1};\alpha)$ are stable. This indicates that the full methodology developed in this paper should work well for empirical problems in which the events in the partially observable outcomes are not rare events.

4.3 Identification

Using the data in which only default and prepayment are observable, I estimate a structural model in which I treat illiquidity-triggered default and strategic default as separate outcomes and treat moving and refinancing as separate outcomes. Accordingly, exclusion restrictions and identification at infinity arguments provide the identification conditions for the parameters in the probabilities of illiquidity-triggered default, strategic default, moving, and refinancing in the first-stage estimation. Identification at infinity arguments have been widely used in econometric studies, such as Heckman (1990) and Andrews and Schafgans (1998) for sample selections and Tamer (2003), Ciliberto and Tamer (2009), and Bajari, Hong and Ryan (2010) for estimation of discrete games.

Suppose that all the state variables $\{s_{i,t}\} = \{z_{i,t}, w_{i,t}, x_{i,t}\}$. One exclusion restriction requires that there exists at least one state variable with sufficient variation, say ${s_{i,t}}^{(1)}$, appearing in $\sigma_t^{sd}(s_{i,t};\gamma)$ and ${s_{i,t}}^{(1)} \notin z_{i,t}$. ${s_{i,t}}^{(1)}$ could be a variable related to the financial incentive to default but not related to illiquidity (e.g., $defaultOpt_{i,t}$ and $Homeprice_{i,t} \cdot \Delta\log(\mathrm{hp}_t)$). Let ${s_{i,t}}^{(1)} \to +\infty$ or $-\infty$ to make $\sigma_t^{sd}(s_{i,t};\gamma) \to 0$; then, in the first equation in Equations (12), $Pr\{default_{i,t}\} \to p_t^{td}(z_{i,t};\varphi)$. Thus, φ are identified.

The other exclusion restriction requires that there exists at least one variable with sufficient variation, say $s_{i,t}^{(2)}$, appearing in $\sigma_t^{r}(s_{i,t};\gamma)$ and $s_{i,t}^{(2)} \notin z_{i,t}, w_{i,t}$. $s_{i,t}^{(2)}$ could be a variable related to the financial incentive to refinance but not related to move (e.g., $refiOpt_{i,t}$). Let $s_{i,t}^{(2)} \to +\infty$ or $-\infty$ to make $\sigma_t^{r}(s_{i,t};\gamma) \to 0$; then, in the second equation in Equations (12), $Pr\{prepay_{i,t}\} \to p_t^{m}(z_{i,t}, w_{i,t};\varphi,\psi)$. Thus, ψ are identified. Because φ and ψ are already identified, $s_{i,t}\gamma^{sd}$ and $s_{i,t}\gamma^{r}$ can be uniquely solved from Equations (22). If $s_{i,t}$ have the full rank, $\gamma = \{\gamma^{sd}, \gamma^{r}\}$ are identified.

$$\sigma_t^{sd}(s_{i,t};\gamma) = \frac{\exp\{s_{i,t}\gamma^{sd}\}}{1+\exp\{s_{i,t}\gamma^{sd}\}+\exp\{s_{i,t}\gamma^{r}\}}$$
$$\sigma_t^{r}(s_{i,t};\gamma) = \frac{\exp\{s_{i,t}\gamma^{r}\}}{1+\exp\{s_{i,t}\gamma^{sd}\}+\exp\{s_{i,t}\gamma^{r}\}} \tag{22}$$
$$\sigma_t^{c}(s_{i,t};\gamma) = \frac{1}{1+\exp\{s_{i,t}\gamma^{sd}\}+\exp\{s_{i,t}\gamma^{r}\}}$$

Intuitively, when $defaultOpt_{i,t}$ is very low (i.e., the current home value is much higher than the unpaid balance), if the household defaults, it should be illiquidity-triggered default rather than strategic default. Thus, households with highly negative $defaultOpt_{i,t}$ can help identify the parameters in the probability of illiquidity-triggered default. Moreover, when both the current individual house value and the current house price index growth rate are very high, $Homeprice_{i,t} \cdot \Delta\log(\text{hp}_t)$ will be high. Consequently, the household's expectation regarding $Homeprice_{i,t} \cdot \Delta\log(\text{hp}_{t+1})$ will be very high, as by assumption households' expectations regarding $\Delta\log(\text{hp}_{t+1})$ follow AR(1) and the serial correlation coefficients are above 0.79 for all the MSAs. If a household with a very high expectation regarding its home value appreciation defaults, it should be illiquidity-triggered default rather than strategic default. Thus, households with a very high expectation regarding their home value appreciation can also help identify the parameters in the probability of illiquidity-triggered default. Thereafter, the parameters in the conditional choice probability of strategic default can be identified.

When $refiOpt_{i,t}$ is very low (the unpaid balance is very high and the current market interest rate is much higher than the coupon rate of the existing mortgage), if a household prepays its mortgage, it should be caused by moving rather than refinancing. Thus, households with a highly negative $refiOpt_{i,t}$ can help identify the parameters in the probability of moving. Thereafter, the parameters in the conditional choice probability of refinancing can be identified.[22] [23]

[22] $refiOpt_{i,t}$ may not be perfectly excluded from the probability of moving. Quigley (1987) found that if the current interest rate is higher than the coupon rate of the existing mortgage, households will have less mobility owing to the lock-in effect (i.e., if they move and prepay their current mortgages, they will need to borrow new mortgages with high interest rates to buy new houses). However, even with negative home equity, many other factors can still induce households to move. The lock-in effect of $refiOpt_{i,t}$ on moving is weaker than other

4.4 Expectations Regarding Future House Prices

As noted by Manski (2004), with data only on observed choices, researchers cannot infer people's both preferences and expectations simultaneously. If researchers need to infer people's preferences, they have to either make assumptions about people's expectations or obtain additional data on them. Because data on households' expectations are usually unavailable, the first approach is generally used in the literature estimating dynamic choice models. There are a few studies, such as Erdem, Keane, Öncü and Strebel (2005), that use the second approach.

The econometric methodology previously discussed in Section 4 follows the first approach. I assume that households forecast the future house prices and interest rates following the AR(1) processes estimated using the realized house price and interest rate data (i.e., rational expectations). However, households' expectations for the future may not follow the assumption of rational expectations. First, the process of households' price expectation formations may not follow the process of actual price movements.[24] Second, households' perception of current prices may deviate from the actual current prices. Davis and Quintin

---

factors' effects on moving. In the data, there are 252 households that prepaid their mortgages with negative $refiOpt_{i,t}$, 112 of which had a highly negative $refiOpt_{i,t}$ (<-\$5000). Moreover, the effect of $refiOpt_{i,t}$ on moving is also much weaker than its effect on refinancing.

[23] The identification at infinity arguments are used only to provide identification conditions. In reality, no variable can have a value of infinity. When estimating the model, I pool all the households together and estimate all the parameters in $p_t^{td}(z_{i,t};\varphi)$, $p_t^m(z_{i,t},w_{i,t};\varphi,\psi)$, $\sigma_t^{sd}(s_{i,t};\gamma)$, $\sigma_t^r(s_{i,t};\gamma)$, and $\sigma_t^c(s_{i,t};\gamma)$ at once. I do not initially use borrowers with extremely negative $defaultOpt_{i,t}$ to estimate the parameters in $p_t^{td}(z_{i,t};\varphi)$ and then use the remaining borrowers to estimate the parameters in $\sigma_{i,t}^{sd}(s_{i,t};\gamma)$ given $p_t^{td}(z_{i,t};\hat{\varphi})$. The first reason for taking this approach is that estimating all the parameters at once is more efficient. The second reason is that if I initially use a subsample of borrowers with extremely negative $defaultOpt_{i,t}$ to estimate $p_t^{td}(z_{i,t};\varphi)$, I need to choose a threshold for $defaultOpt_{i,t}$ to form the subsample. If the threshold is too negative, then the subsample size will be too small, which will give the estimates of $p_t^{td}(z_{i,t};\varphi)$ large standard errors; if the threshold is not very negative, then it will be far away from infinity and $\sigma_{i,t}^{sd}(s_{i,t};\gamma)$ of borrowers in the subsample will not be close to zero, which will make the estimates of $p_t^{td}(z_{i,t};\varphi)$ substantially biased. The third reason is that there is another partial observability issue for the prepayment outcomes. Thus, I actually need to select borrowers with extremely negative values for both $defaultOpt_{i,t}$ and $refiOpt_{i,t}$ in the first step, which will make the subsample size even smaller. Other studies involving identification at infinity arguments, such as Ciliberto and Tamer (2009) and Bajari, Hong and Ryan (2010), also estimate all the parameters at once.

[24] Both Case, Shiller and Thompson (2012) (based on survey data covering four counties in the U.S.) and Ma (2015b) (based on the Michigan Survey of Consumers) found that households do not have rational expectations regarding house prices. Using the Wall Street Journal's economic forecasting survey, Zhang (2015) found that even professional forecasters' predictions regarding house prices are biased and inefficient.

(2013) found that the changes in self-assessed house prices significantly differ from those of house price indices calculated based on actual house transactions and that the former plays a more important role in mortgage defaults than the latter.

Alternatively, I also use the second approach in which I use the subjective house price expectation data of the Ohio respondents from the Michigan Survey of Consumers to represent the house price expectations of the households in my mortgage loan performance data, who also reside in Ohio.[25] When estimating the state transition equations, Erdem, Keane, Öncü and Strebel (2005) used consumer expectation data on PC prices from a survey for the dependent variable and used current and past realized PC price data for the explanatory variables. However, in contrast to PC analyzed in Erdem, Keane, Öncü and Strebel (2005), detergent analyzed in Hendel and Nevo (2006), and soda analyzed in Wang (2015), the future prices of houses are highly correlated with not only past house prices but also other macroeconomic variables, such as GDP growth, unemployment, and income per capita. One way to solve this problem would be to include other macroeconomic variables on the right-hand side of the state transition equation. However, this method still requires an assumption regarding the structure of the expectation formation process, and it would not work well if the structure is miss-specified.[26] Instead of estimating Equation (14) and then performing one-period forward simulations based on the estimates to obtain the distribution of people's house price expectations for the next period, I directly use the empirical distribution of the house price forecasts in each month by the Ohio respondents of the Michigan Survey of Consumers as the distribution of people's house price forecasts for the next period. This approach does not require any assumption regarding the structure of the state transition process.[27] Accordingly, for the house price dimension,

[25] In this approach, I drop the observations before 2007 in the loan performance data because the subjective house price expectation data in the Michigan Survey of Consumers is available only after 2007.

[26] In the literature estimating dynamic choice models, researchers have generally assumed AR(1) for the structure of state transitions for simplicity. However, in the housing economics literature, many researchers, such as Abraham and Hendershott (1996), Gallin (2006), Gallin (2008), Capozza, Hendershott and Mack (2004), and Malpezzi (1999), have used error correction models to model house price dynamics. It is difficult to implement error correction models as state transitions in estimations of dynamic choice models, and I am unaware of any paper that has done so.

[27] In fact, this distribution captures the disagreement in individuals' point forecasts. Ideally, individuals' density forecasts should be used here to capture individuals' uncertainty about the future. However, no survey asks respondents about their density forecasts of local house prices, although some surveys ask respondents about their density forecasts of other economic variables. Fortunately, Bomberger (1996), Rich and Tracy (2010), and

$\hat{E}_t\left[W_{t+1}\left(s_{i,t+1}\right)|s_{i,t}, a_{i,t} = k\right]$ in Equations (19) is averaged over the distribution of the forecasts by the Ohio respondents of the Michigan Survey of Consumers in each month rather than the distribution generated by the forward simulations based on the estimation results of the first equation in Equations (14).

5 Empirical Results

5.1 Preliminary Regressions

Table III presents the results of multinomial logit regressions with three possible outcomes for the dependent variable: default, prepayment, and continuing to pay. The probability of default increases as the monthly payment, household size (which is positively correlated with household expenditure), local unemployment rate, and default option value increase and as the household income and credit score decrease. Households choosing to take a DPA grant are more likely to default because they may have lower savings levels. The probability of prepayment increases as the refinance option value increases. In addition to refinance, move can lead to prepayment. The coefficients of household size and age for prepayment are significantly negative, as larger and older households have less mobility. The coefficient of the difference between the local and the national employment growth rate is also significantly negative, as households would like to move from areas with few opportunities to areas with more opportunities.

5.2 First-stage Estimation

Table IV displays the results for the policy functions in the first-stage estimation. Panel A of Table IV reports the estimates of the parameters in the probability of illiquidity-triggered default and those in the probability of moving. A larger monthly payment amount, household size, and unemployment rate in the county will make households significantly more likely to have illiquidity-triggered default; a higher income and credit score (a measure of accessibility to other loans) will make households significantly less likely to have illiquidity-triggered default. Further, the probability of moving is significantly lower for larger and older

---

Bachmann, Elstner and Sims (2013) have provided empirical evidence that forecasters' uncertainty about the future is positively correlated with the disagreement in their point forecasts. Many papers have used disagreement as a proxy of uncertainty.

households. These results are consistent with Quigley (1987) and Ferreira, Gyourko and Tracy (2010). The past-12-month moving average of the difference between the local and the national employment growth rate is also significantly negatively correlated with the probability of moving.[28] Minorities have a higher probability of facing illiquidity and a lower mobility, but the coefficients are not statistically significant. The reason for this result could be that the sample is special (first-time homebuyers with low-to moderate income) rather than randomly drawn from the whole population.[29]

For the conditional choice probabilities for strategic default and refinance, I employ a flexible multinomial logit form in the estimation, involving interaction terms and splines.[30] As in Panel B of Table IV, I report only the coefficients for some interpretable variables. For strategic default, the coefficient of $defaultOpt_{i,t}$ is significantly positive; the effect of $defaultOpt_{i,t}$ is larger in the negative domain than in the positive domain. For refinance, the coefficient of $refiOpt_{i,t}$ is significantly positive; the effect of $refiOpt_{i,t}$ is larger in the negative domain than in the positive domain.

[28] The result is robust to changes in the moving average windows.

[29] Using the AHS data, Ferreira, Gyourko and Tracy (2010) found that negative home equity has a lock-in effect on households' mobility. In other specifications, I include $max\{0, defaultOpt_{i,t}\}$ or $defaultOpt_{i,t}$ in the probability of moving, but the coefficient is insignificantly positive rather than significantly negative. One possible reason for this result could also be that the sample is special (first-time homebuyers with low-to-moderate income) rather than randomly drawn from the whole population.

[30] Some papers applying the CCP method have used nonparametric methods to estimate the conditional choice probabilities in the first stage; based on whether they provided a good fit to the sample, the researchers chose the specifications of the nonparametric methods to construct the value functions for the future used in the second-stage estimation. Other papers used flexible parametric forms to estimate the conditional choice probabilities in the first stage; researchers then chose the specifications of the flexible function forms based on whether they provided a good fit to the sample. In this paper, I use flexible parametric forms. The first reason I adopt this approach is that the number of state variables is large; thus, nonparametric methods cannot have a good performance because of the curse of dimensionality. The second reason is that the results of nonparametric models in the first-stage estimation are not interpretable. Whether default is illiquidity-triggered default or strategic default is not observable, and whether prepayment is caused by move or refinance is also not observable; therefore, if I use the first method, I cannot check whether the outcomes predicted by the model have a good fit to the outcomes in the sample at the level of illiquidity-triggered default, strategic default, move, refinance, or continuing to pay; rather, I can check the fit only at the level of default, prepayment, or continuing to pay. Instead, I use flexible parametric forms to estimate the conditional choice probabilities in the first stage and interpret the parameters; then, I decide whether a specification works well based on whether the results for the parameters are consistent with the relevant theories and common sense. The third reason is related to identification. As discussed in Section 4.3, exclusion restrictions and identification at infinity arguments provide the identification conditions. With a parametric function form for the probability of strategic default, after I obtain the estimates in the first stage, I can check whether the probability of strategic default goes to zero when a certain variable goes to infinity. However, it is not straight forward to check this based on the results obtained in the first-stage estimation if nonparametric methods are used.

Other state variables in the conditional choice probabilities include the credit score, interest rate innovation, MSA-level house price index change, purchase price, current home value calculated based on the MSA-level house price index, derivative of $refiOpt_{i,t}$ with respect to the interest rate innovation, and some interaction terms of these variables.

Predicted by the first-stage estimates, up to March 2011, the illiquidity-triggered default rate is 8.37%, while the strategic default rate is 2.88%. This result is reasonable because all of the households in this sample are low-to-moderate income first-time home buyers, who are more likely to encounter illiquidity than the general population. The sum of the two rates is 11.25%, which is equal to the actual default rate. Also predicted by the first-stage estimates, up to March 2011, the moving rate is 1.05%, while the refinance rate is 9.86%. This result is also reasonable because low-to-moderate income households have less mobility. The sum of the two rates is 10.91%, which is equal to the actual prepayment rate.

The results of state transitions are reported in Table V. Compared to the change in the interest rate, the change in the house price index is more closely related to its one-period lag. The coefficients for the house price transition process do not vary much across different MSAs.

5.3 Second-stage Estimation

In the second-stage estimation, I set the monthly discount factor β=0.995.[31] The first column of Table VI reports the structural parameters obtained from the second-stage estimation using the forward simulation based on the AR(1) state transitions estimated in Table V. The coefficients are sensible. The standard errors in the second-stage estimation are adjusted using the correction methods developed by Murphy and Topel (1985).

The second column of Table VI reports the structural parameters obtained from the second-stage estimation using the empirical distribution of households' subjective house price expectations in the Michigan Survey of Consumers. There is no substantial difference between the estimates in the two columns of Table VI. According to Manski (2004), with

[31] In estimations of dynamic choice models, the discount factor is generally not identifiable. The standard approach in the literature is to set a value for this factor. Bajari, Chu, Nekipelov and Park (2013) developed an econometric technique to identify the discount factor. Identifying and estimating the discount factor is beyond the scope of my paper.

only data on observed choices, estimates of preferences depend on assumptions regarding beliefs. Households' subjective expectations in the Michigan Survey are a bit more optimistic than the expectations generated from the AR(1) process estimated using the realized house price data, and more optimal expectations tend to make households less likely to default. Therefore, to explain the same number of observed defaults, the coefficients related to the costs of default (α_4 and α_5) should be a bit smaller in the second column of Table VI.

If I simply treat default outcomes as strategic default and treat prepayment outcomes as refinance and use the traditional CCP method to estimate the structural parameters in households' intertemporal utility-maximizing problem of whether to default, refinance, or continue to pay, then the following two consequences will be generated. First, the conditional choice probabilities for strategic default obtained in the first-stage estimation will be contaminated by the probabilities of illiquidity-triggered default. When the contaminated conditional choice probabilities for strategic default are used to recover the structural parameters in the second stage, α_4, α_5, and α_2 related to the cost of default will be underestimated, and α_1 related to the benefit of default will be overestimated. The second consequence is that the conditional choice probabilities for refinance obtained in the first-stage estimation will be contaminated by the probabilities of move. When the contaminated conditional choice probabilities for refinance are used to recover the structural parameters in the second stage, α_3 related to the cost of refinance will be underestimated, and α_1 related to the benefit of refinance will be overestimated.[32]

6 Counterfactual Analyses

I conduct two sets of counterfactual analyses. The first set involves mortgage loan modification policies aimed at foreclosure mitigation. The second set considers the Federal Reserve CCAR scenarios.

6.1 Loan Modification Policies

Since the 2007-2009 financial crisis, several loan modification programs aimed at foreclosure mitigation, such as the Home Affordable Modification Program (HAMP), have been

[32] These results are available upon request.

designed. Borrowers struggling to make mortgage payments due to financial hardships can apply for those programs if they meet certain eligibility criteria, and their monthly payment amount can be reduced through an interest rate reduction, term extension, principal forbearance, or/and principal forgiveness. In addition to reducing the expected loss in foreclosure, lenders can also benefit from obtaining compensation from the government for modifying mortgages. The lenders participating in the HAMP use a net present value (NPV) model as a tool to decide whether to modify a troubled mortgage. First, the model calculates the probability of default for the troubled mortgage without modification and the probability of default with modification. Second, under certain assumptions, it calculates the expected cash flow that the lender could obtain from this mortgage in the future in the following four scenarios: 1) modified and default; 2) modified and cure; 3) unmodified and default; and 4) unmodified and cure. Then, it calculates the present value if the mortgage is modified and the present value if the mortgage is unmodified to determine whether approving modification for the mortgage gives the lender a positive NPV.[33] While the model has a complex procedure for calculating the cash flow, its methodology for calculating the probability of default is merely a simple and intuitive logit model.

Previous empirical studies on loan modification programs, such as Quercia and Ding (2009) and Haughwout, Okah, and Tracy (2009), have found that in terms of the effect in reducing the overall default rate, writing down the principal is more effective than interest reduction, which is more effective than term extension. They also provided a conceptual discussion that writing down the principal can both increase households' ability to pay through payment reduction and raise their incentive to pay by increasing their home equity, whereas interest reduction and term extension do not have the latter channel.

One contribution of this paper is that the model incorporating partial observability can provide numerical predictions for the changes in the illiquidity-triggered default rate and the strategic default rate caused by each type of loan modification. The first row in Table VII reports the estimated illiquidity-triggered default rate and strategic default rate in the actual history. The other rows report how much each loan modification policy can reduce these

[33] Fannie Mae has developed the NPV model software for HAMP-participating lenders to use. Details on the HAMP NPV model can be found in the HAMP online documentation: https://www.hmpadmin.com/portal/programs/hamp.jsp

rates. As shown in Table VII, writing down the principal can effectively reduce both the illiquidity-triggered default rate and the strategic default rate. Interest reduction can effectively reduce the illiquidity-triggered default rate but cannot effectively reduce the strategic default rate. Term extension can reduce the total default rate. However, after splitting the total default rate into the illiquidity-triggered default rate and the strategic default rate, the result shows that term extension actually increases the strategic default rate. As the mortgage term is extended from 30 to 40 years, the speed at which the homeowners build home equity decreases; therefore, they will have a longer period with low or even negative home equity, which can drive up the strategic default rate. [34]

6.2 Federal Reserve CCAR Scenarios

Starting in 2011, the Federal Reserve provided projections each year for 28 macroeconomic variables over the next 13 quarters under each of the following three hypothetical scenarios: baseline, adverse, and severely adverse. Banks are required to build models to forecast the probabilities of default (PD) for the loans they hold and the loss given default (LGD) in these scenarios to determine whether they have adequate capital to withstand a highly stressful economic environment. They are required to submit their models and planning to the Federal Reserve in order to pass the annual stress test on their strength and resilience.[35]

I use the model developed in this paper to forecast the default rate and refinance rate under the adverse and severely adverse scenarios provided by the Federal Reserve for the loans originated in 2005. I assume that the percent changes in the MSA house price index, county unemployment rate, and mortgage interest rate during 2006 Q1 – 2009 Q1 followed the percent changes in the national house price index, national unemployment rate, and mortgage interest rate projected for 2013 Q4 to 2016 Q4 by the Federal Reserve for the 2014 annual stress tests. All of the other variables remain the same as the actual history.[36] [37]

---

[34] In these counterfactual experiments, loan modification is applied to every household in the sample from the beginning. In practice, loan modification is applied only to the households that currently have financial hardship, apply for the loan modification program, and get approved by the bank.

[35] Details about the CCAR can be found in the Federal Reserve CCAR online documentation: http://www.federalreserve.gov/bankinforeg/ccar.htm

[36] The 2014 projection of the macroeconomic variables by the Federal Reserve is for 2013 Q4 through 2016 Q4 rather than for 2006 Q1 through 2009 Q1. However, as stated in their document, "the adverse and severely adverse scenarios are not forecasts, but rather are hypothetical scenarios designed to assess the strength of

The left column of Figure 2 shows the predicted cumulative total default rate, illiquidity-triggered default rate, and strategic default rate across different scenarios. The right column of Figure 2 shows the movements of the national house price index, mortgage interest rates, and national unemployment rates across different scenarios projected by the Federal Reserve. For the total default rate, the value is higher for the severely adverse scenario than for the adverse scenario, which has a higher value than that for the actual history. Before 2007 Q2, the cumulative total default rate was increasing at a lower speed in the actual history than in the adverse and severely adverse scenarios. After 2007 Q2, its increasing speed in the actual history is close to that in the adverse and severely adverse scenarios. For the illiquidity-triggered default rate and strategic default rate, the level is higher for the severely adverse scenario than for the adverse scenario.[38]

The precision of the counterfactual analyses for loan modifications and the CCAR scenarios requires that the state variables in the counterfactual scenarios remain in the empirical support of the policy functions. As the loan performance data used in this paper include 20,487 mortgages with sufficient variation in characteristics, even after loan modification, the characteristics of most of the mortgages should remain in the empirical support of the policy functions. As the loan performance data span from 2005 to 2011, which covers the boom, crisis, and recovering of the U.S. economy and housing market, the macroeconomic variables and the individual-level variables translated from the macroeconomic variables in the CCAR adverse and severely adverse scenarios should remain in the empirical support of the policy functions.

---

banking organizations and their resilience to adverse economic environments". The counterfactual analyses in this part of the paper examine how the default rate and refinance rate for the 2005 cohort would move if the Federal Reserve's hypothetical adverse and severely adverse scenarios happened during 2006 Q1 – 2009 Q1.

[37] The coefficients of the state transitions of interest rate innovations and house price appreciations underlying these counterfactual scenarios might differ from those for the actual history. However, it is impossible to estimate the state transitions in counterfactual scenarios. I checked the Michigan Survey of Consumers and found that the pattern of households' house price expectation formation is not very different across the pre-crisis, crisis, and recovery periods.

[38] Note that in the actual historical data, the illiquidity-triggered default rate and strategic default rate are not observable owing to the partial observability issue.

7 Conclusion

In this paper, I estimate a dynamic discrete choice model with partially observable outcomes for household mortgage default and prepayment behaviors. When households decide whether to default on or refinance their mortgages, there are tradeoffs both between defaulting right now and defaulting later and between refinancing now and refinancing later, and their expectations regarding house price and interest rate movements in the future affect their decisions in the current period. This motivates estimating a dynamic discrete choice model. In loan performance data, researchers can observe only whether default occurs but not whether it is strategic or triggered by illiquidity. Moreover, researchers can typically observe only whether prepayment occurs, but not whether it is due to refinancing or moving. These facts motivate estimating a dynamic discrete choice model with partially observable outcomes.

The contribution to econometric methodology is that I extend the CCP method to estimate a dynamic discrete choice model with partially observable outcomes (or dependent variables). Arcidiacono and Miller (2011) extended the CCP method to incorporate unobserved heterogeneity of independent variables. Both the CCP method with partial observability of dependent variables and the CCP method with unobserved heterogeneity of independent variables involve estimating a mixture of distributions in the first stage. However, there are significant differences between the two extensions.

First, in the CCP method with unobserved heterogeneity of independent variables, agents in different unobserved exogenous types are solving the same problems. By contrast, in the CCP method with partial observability of dependent variables, agents are solving different problems, but the decisions within these problems are mixed in the observable actions or outcomes. Different distributions in the mixture estimated in the first step are associated with different problems, and they will be used to recover the structural parameters in the corresponding problems in the second step. In addition, in the CCP method with partial observability of dependent variables, the agents could be solving the same problem, but the outcomes can be generated through different mechanisms (censored dependent variables). Moreover, the action in solving one problem can affect the probability of another outcome in the mixture in the future. For example, the probability of illiquidity-triggered default can be lower than before if borrowers chose to refinance to a new loan with lower monthly payment amount.

Second, for single agent dynamic problems (not dynamic games), in the CCP method with unobserved heterogeneity of independent variables, when agents are solving a problem, each agent needs to consider only the probability distribution associated with his or her own exogenous type in the mixture. By contrast, in the CCP method with partial observability of dependent variables, when agents solve a problem, they may also need to consider the probability distribution in the mixture associated with other problems or other mechanisms that can generate the same outcomes. In the second step of the estimation, the structural parameters are recovered based on this fact.

The paper also makes contributions to empirical studies and policy implications. The econometric model in this paper yields separate predictions of the probabilities of illiquidity-triggered default, strategic default, refinancing, and moving. The counterfactual analyses for foreclosure-mitigating mortgage modification policies show that writing down the principal can effectively reduce both the illiquidity-triggered default and the strategic default because it can reduce both the probability of illiquidity and the financial incentive to default. Interest rate reduction can effectively reduce illiquidity-triggered default but cannot effectively reduce strategic default. Term extension can effectively reduce illiquidity-triggered default but will increase strategic default, as households will have a longer period with low or even negative home equity after the extension.

Another innovation is that as an alternative to the usual methods that assume rational expectations for consumers and estimate a structure for the state transitions using the realized data, I follow the spirit of Manski (2004) and directly use the subjective expectation data on future house prices from the Michigan Survey of Consumers in each month to construct the distribution of households' future house price expectations for the state transitions in estimating the dynamic choice model. This method does not require an assumption of rational expectations and an assumption of the structure for house price expectation formation processes.

References

Abbraham, J., and P. Hendershott (1996): Bubbles in Metropolitan Housing Market, Journal of Housing Research, 7, 191–207.

Aguirregabiria, V., and C. Ho (2012): A Dynamic Oligopoly Game of the US Airline Industry: Estimation and Policy Experiments, Journal of Econometrics, 168, 156–173.

Aguirregabiria, V., and P. Mira (2002): Swapping the Nested Fixed Point Algorithm: A Class of Estimators for Discrete Markov Decision Models, Econometrica, 70, 1519–1543.

Aguirregabiria, V., and P. Mira (2010): Dynamic Discrete Choice Structural Models: A Survey, Journal of Econometrics, 156, 38–67.

An, X., J. M. Clapp, and Y. Deng (2010): Omitted Mobility Characteristics and Property Market Dynamics: Application to Mortgage Termination, Journal of Real Estate Finance and Economics, 41, 245–271.

Andrews, D. W. K., and M. M. A. Schafgans (1998): Semiparametric Estimation of the Intercept of a Sample Selection Model, Review of Economic Studies, 65, 497–517.

Archer, W. R., D. C. Ling, and G. A. McGill (1996): The Effect of Income and Collateral Constraints on Residential Mortgage Terminations, Regional Science and Urban Economics, 26, 235–261.

Arcidiacono, P., P. Bayer, J. R. Blevins, and P. B. Ellickson (2014): Estimation of Dynamic Discrete Choice Models in Continuous Time with an Application to Retail Competition, NBER Working Paper, (18449).

Arcidiacono, P., and R. A. Miller (2011): Conditional Choice Probability Estimation of Dynamic Discrete Choice Models with Unobserved Heterogeneity, Econometrica, 79, 1823–1867.

Bachmann, R., S. Elstner, and E. R. Sims (2013): Uncertainty and Economic Activity: Evidence from Business Survey Data, American Economic Journal: Macroeconomics, 5, 217-49.

Bajari, P., C. L. Benkard, and J. Levin (2007): Estimating Dynamic Models of Imperfect Competition, Econometrica, 75, 1331–1370.

Bajari, P., C. S. Chu, D. Nekipelov, and M. Park (2013): A Dynamic Model of Subprime Mortgage Default: Estimation and Policy Implications, NBER Working Paper, (18850).

Bajari, P., C. S. Chu, and M. Park (2010): An Empirical Model of Subprime Mortgage Default from 2000 to 2007, NBER Working Paper, (14625).

Bajari, P., H. Hong, and S. P. Ryan (2010): Identification and Estimation of a Discrete Game of Complete Information, Econometrica, 78, 1529–1568.

Bajari, P., P. Chan, D. Krueger, and D. Miller (2013): A Dynamic Model of Housing Demand: Estimation and Policy Implications, International Economic Review, 54, 409–442.

Benkard, C. L., A. Bodoh-Creed, and J. Lazarev (2010): Simulating the Dynamic Effects of Horizontal Mergers: US Airlines, Working Paper, Yale University.

Benoit, J. (1984): Financially Constrained Entry in a Game with Incomplete Information, RAND Journal of Economics, 15, 490–499.

Berry, S., J. Levinsohn, and A. PAKES (2004): Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market, Journal of Political Economy, 112, 68–105.

Blau, D. M. (2008): Retirement and Consumption in a Life Cycle Model, Journal of Labor Economics, 26, 35–71.

Bomberger, W. A. (1996): Disagreement as a Measure of Uncertainty, Journal of Money Credit and Banking, 28, 381–392.

Calhoun, C. A., and Y. Deng (2002): A Dynamic Analysis of Fixed- and Adjustable-Rate Mortgage Terminations, Journal of Real Estate Finance and Economics, 24, 9–33.

Campbell, J. Y., and J. F. Cocco (2015): A Model of Mortgage Default, The Journal of Finance, 70, 1495–1554.

Capozza, D. R., P. H. Hendershott, and C. Mack (2004): An Anatomy of Price Dynamics in Illiquid Markets: Analysis and Evidence from Local Housing Markets, Real Estate Economics, 32, 1–32.

Carranza, J. E. and S. Navarro (2010): Estimating Dynamic Models with Aggregate Shocks and an Application to Mortgage Default in Colombia, Universidad Icesi Deapartment of Economics Research Paper, (24).

Case, K. E., R. J. Shiller, and A. Thompson (2012): What Have They Been Thinking? Home Buyer Behavior in Hot and Cold Markets, Brookings Papers on Economic Activity, 265–298.

Ciliberto, F., and E. Tamer (2009): Market Structure and Multiple Equilibria in Airline Markets, Econometrica, 77, 1791–1828.

Clapp, J. M., G. M. Goldberg, J. P. Harding, and M. LaCour-Little (2001): Movers and Shuckers: Interdependent Prepayment Decisions, Real Estate Economics, 29, 411–450.

Clapp, J. M., Y. Deng, and X. An (2006): Unobserved Heterogeneity in Models of Competing Mortgage Termination Risks, Real Estate Economics, 34, 243–273.

Conlon, C. T. (2012): A Dynamic Model of Prices and Margins in the Lcd Tv Industry, Working Paper, University of Columbia.

Corbae, D., and E. Quintin (2015): Leverage and the Foreclosure Crisis, Journal of Political Economy, 123, 1-65.

Davis, M., and E. Quintin (2013): Self-Assessed House Prices and Mortgage Defaults, Working Paper, University of Wisconsin-Madison.

Deng, Y. (1997): Mortgage Termination: An Empirical Hazard Model with a Stochastic Term Structure, Journal of Real Estate Finance and Economics, 14, 309–331.

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

Donna, J. (2012): Estimating Switching Costs in Urban Travel Demand in Chicago, Working Paper, Ohio State University.

Erdem, T., M. P. Keane, T. S. Öncü, and J. Strebel (2005): Learning About Computers: An Analysis of Information Search and Technology Choice, Quantitative Marketing and Economics, 3, 207–247.

Ergungor, O. E., and S. Moulton (2014): Beyond the Transaction: Depository Institutions and Reduced Mortgage Default for Low-Income Homebuyers, forthcoming at Journal of Money, Credit and Banking,

Ferreira, F., J. Gyourko, and J. Tracy (2010): Housing Busts and Household Mobility, Journal of Urban Economics, 68, 34–45.

Gallin, J. (2006): The Long-Run Relationship Between House Prices and Income: Evidence from Local Housing Markets, Real Estate Economics, 34, 417–438.

Gallin, J. (2008): The Long-Run Relationship Between House Prices and Rents, Real Estate Economics, 36, 635–658.

Gerardi, K., K. F. Herkenhoff, L. E. Ohanian, and P. Willen (2013): Unemployment, Negative Equity, and Strategic Default, Working Paper.

Glover, B. (2015): The Expected Cost of Default, forthcoming at Journal of Financial Economics.

Gowrisankaran, G., and M. Rysman (2012): Dynamics of Consumer Demand for New Durable Goods, Journal of Political Economy, 120, 1173–1219.

Guiso, L., P. Sapienza, and L. Zingales (2013): The Determinants of Attitudes Toward Strategic Default on Mortgages, The Journal of Finance, 68, 1473–1515.

Haughwout, A. F., E. Okah, and J. S. Tracy (2009): Second Chances: Subprime Mortgage Modification and Re-Default, FRB of New York Staff Report, (417).

Heckman, J. (1990): Varieties of Selection Bias, American Economic Review, 80, 313–318.

Hendel, I., and A. Nevo (2006): Measuring the Implications of Sales and Consumer Inventory Behavior, Econometrica, 74, 1637–1673.

Ho, C. Y. (2010): Switching Cost and the Deposit Demand in China, Working Paper, Georgia Institute of Technology.

Hotz, V. J., and R. A. Miller (1993): Conditional Choice Probabilities and the Estimation of Dynamic Models, The Review of Economic Studies, 60, 497–529.

Hotz, V. J., R. A. Miller, S. Sanders, and J. Smith (1994): A Simulation Estimator for Dynamic Models of Discrete Choice, The Review of Economic Studies, 61, 265–289.

Imbens, G. W., and T. Lancaster (1994): Combining Micro and Macro Data in Microeconometric Models, The Review of Economic Studies, 61, 655–680.

Kau, J. B., D. C. Keenan, W. J. Muller, III, and J. F. Epperson (1995): The Valuation at Origination of Fixed-Rate Mortgages with Default and Prepayment, Journal of Real Estate Finance and Economics, 11, 5–36.

Kau, J. B., D. C. Keenan, and X. Li (2011): An Analysis of Mortgage Termination Risks: A Shared Frailty Approach with Msa-Level Random Effects, Journal of Real Estate Finance and Economics, 42, 51–67.

Kennan, J., and J. R. Walker (2011): The Effect of Expected Income on Individual Migration Decisions, Econometrica, 79, 211–251.

Laufer, S. (2011): Equity Extraction and Mortgage Default, Working Paper.

Lee, R. S. (2013): Vertical Integration and Exclusivity in Platform and Two-Sided Markets, American Economic Review, 103, 2960–3000.

Ma, C. (2015a): House Price Expectations and Mortgage Default Decisions, Working Paper.

Ma, C. (2015b): Momentum and Reversion to Fundamentals: Are They Captured by Households' Subjective Expectations of House Prices?, Working Paper.

Malpezzi, S. (1999): A Simple Error Correction Model of House Prices, Journal of Housing Economics, 8, 27–62.

Manski, C. F. (2004): Measuring Expectations, Econometrica, 72, 1329–1376.

Meng, C., and P. Schmidt (1985): On the Cost of Partial Observability in the Bivariate Probit Model, International Economic Review, 26, 71–85.

Murphy, A. (2013): A Dynamic Model of Housing Supply, Working Paper.

Murphy, K. M., and R. H. Topel (1985): Estimation and Inference in Two-Step Econometric Models, Journal of Business & Economic Statistics, 3, 370–379.

Pakes, A. (1986): Patents as Options: Some Estimates of the Value of Holding European Patent Stocks, Econometrica, 54, 755–784.

Pantano, J., and Y. Zheng (2013): Using Subjective Expectations Data to Allow for Unobserved Heterogeneity in Hotz-Miller Estimation Strategies, Working Paper.

Pavlov, A. D. (2001): Competing Risks of Mortgage Termination: Who Refinances, Who Moves, and Who Defaults?, Journal of Real Estate Finance and Economics, 23, 185–211.

Petrin, A. (2002): Quantifying the Benefits of New Products: The Case of the Minivan, *Journal of Political Economy*, 110, 705-729.

Poirier, D. J. (1980): Partial Observability in Bivariate Probit Models, Journal of Econometrics, 12, 209–217.

Quercia, R. G., and L. Ding (2009): Loan Modifications and Redefault Risk: An Examination of Short-Term Impacts, Cityscape, 11, 171–193.

Quigley, J. M. (1987): Interest Rate Variations, Mortgage Prepayments and Household Mobility, The Review of Economics and Statistics, 69, 636–643.

Rich, R., and J. Tracy (2010): The Relationships Among Expected Inflation, Disagreement, and Uncertainty: Evidence from Matched Point and Density Forecasts, Review of Economics and Statistics, 92, 200–207.

Rose, M. J. (2012): Origination Channel, Prepayment Penalties and Default, Real Estate Economics, 40, 663–708.

Rust, J. (1987): Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher, Econometrica, 55, 999–1033.

Schiraldi, P. (2011): Automobile Replacement: A Dynamic Structural Approach, RAND Journal of Economics, 42, 266–291.

Shcherbakov, O. (2009): The Effect of Consumer Switching Costs on Market Power of Cable Television Providers, Working Paper.

Song, I., and P. K. Chintagunta (2003): A Micromodel of New Product Adoption with Heterogeneous and Forward-Looking Consumers: Application to the Digital Camera Category, Quantitative Marketing and Economics, 1, 371–407.

Stanton, R., and N. Wallace (1998): Mortgage Choice: What's the Point?, Real Estate Economics, 26, 173–205.

Tamer, E. (2003): Incomplete Simultaneous Discrete Response Model with Multiple Equilibria, Review of Economic Studies, 70, 147–165.

Tiruppatur, V., O. Chang, and J. Egan (2010): Abs Market Insights: Understanding Strategic Defaults, Working Paper, Morgan Stanley.

Wang, E. Y. (2015): The Impact of Soda Taxes on Consumer Welfare: Implications of Storability and Taste Heterogeneity, RAND Journal of Economics, 46, 409-411.

Zhang, J. (2015): House Price Expectations: Unbiasedness and Efficiency of Forecasters, forthcoming at Real Estate Economics.

Zhang, Y. (2010): High Default Risk on Down Payment Assistance Program: Adverse Selection V.S. Program Characteristics?, Working Paper.

FIGURE 1.—The model

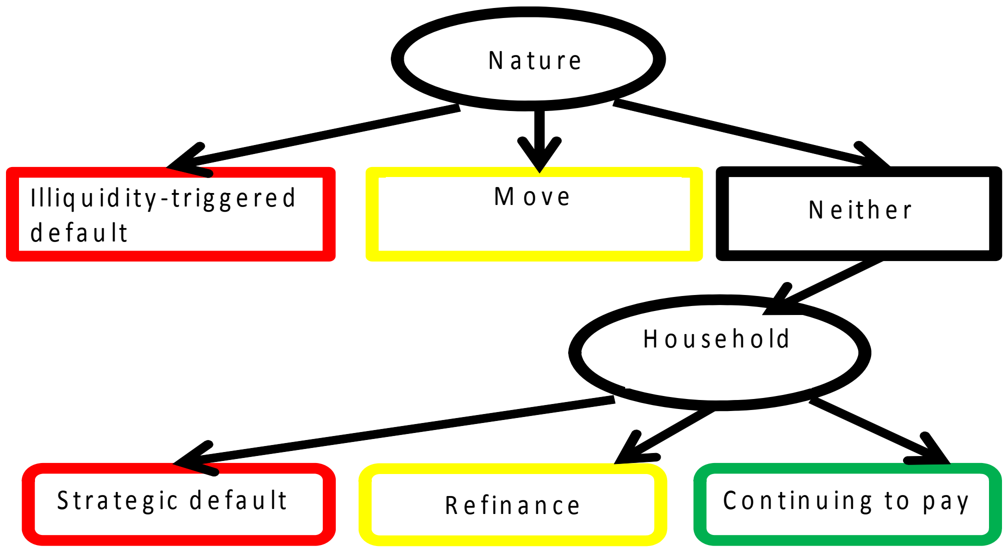
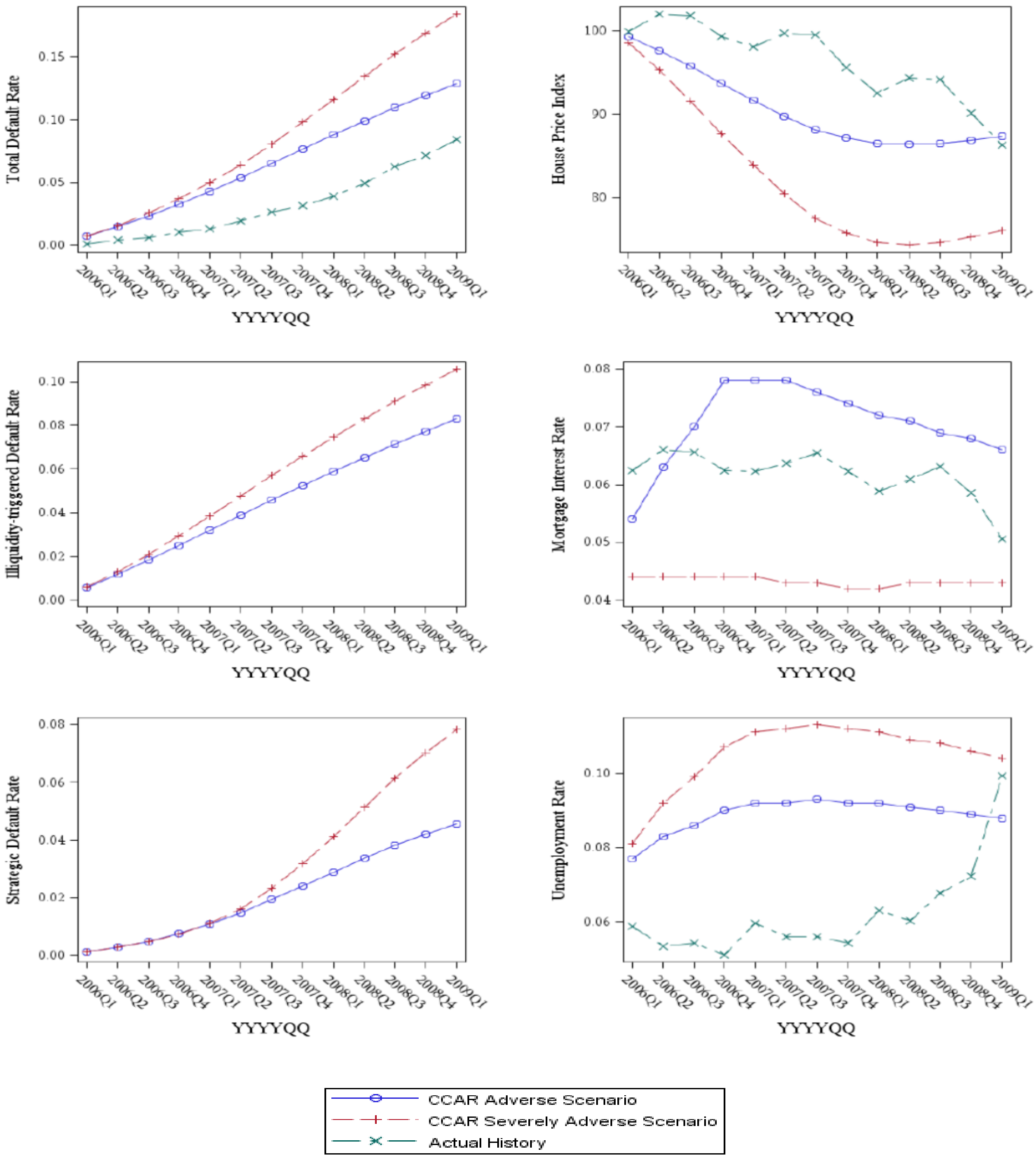


FIGURE 2.—Counterfactual Analyses for CCAR scenarios



Note: House price indices are normalized to 100 in the fourth quarter of 2005

TABLE I

Prepayment, Default, and Delinquency Rate

	Default	Prepayment	Total
Percent	11.25%	10.91%	--
Frequency	2305	2235	20487

TABLE II

Descriptive Statistics for the Loan Performance Data

	N	Mean	Std
Purchase Price of the House	20,487	109584.280	32451.820
Loan Amount	20,487	103347.680	32140.510
LTV	20,487	0.944	0.097
Mortgage Interest Rate	20,487	5.865	0.462
Monthly Payment	20,487	612.544	186.485
Credit Score	20,487	687.337	65.013
Monthly Total Income	20,487	3232.900	942.721
DPA Grant Indicator	20,487	0.254	0.435
African-American	20,487	0.091	0.288
Hispanic	20,487	0.019	0.138
Female Household Head Indicator	20,487	0.397	0.489
Age at Origination	20,487	31.498	10.077
Household Size	20,487	1.940	1.185
$defaultOpt_{i,t}$	896,043	-439.955	10160.316
$refiOpt_{i,t}$	896,043	2294.258	10288.742

Note: $defaultOpt_{i,t}$ and $refiOpt_{i,t}$ are panel data; thus, there are 896,043 observations. Other variables are cross-sectional data.

TABLE III

Logit Regression

	Default	Prepay
Intercept	0.1304	-9.214***
	(0.2767)	(0.3089)
Payment	0.000953***	-0.00016
	(0.000146)	(0.000144)
Income	-0.00031***	0.000069***
	(0.000027)	(0.000026)
Household size	0.0967***	-0.0931***
	(0.0157)	(0.0216)
Credit Score	-0.0104***	0.00379***
	(0.000346)	(0.000368)
Minority	0.042	-0.437***
	(0.0599)	(0.097)
County unemployment rate	12.0249***	6.52***
	(1.0123)	(1.1063)
Employment growth rate diff	-0.2295	-7.646***
(local - national)	(2.3018)	(2.1259)
Female	0.00753	-0.2025***
	(0.0439)	(0.0454)
DPA grant	0.0554	-0.0751
	(0.046)	(0.0503)
Age	0.000448	-0.0145***
	(0.002150	(0.00242)
$defaultOpt_{i,t}$ ($1000)	0.0182***	-0.0221***
	(0.00275)	(0.00141)
$refiOpt_{i,t}$ ($1000)	-0.00347	0.0927***
	(0.00267)	(0.00273)

Note: * denotes significance at a 10% level. ** denotes significance at a 5% level. *** denotes significance at a 1% level. The standard errors are in parentheses.

TABLE IV

Policy Functions in the 1st Stage Estimation

Panel A	Illiquidity-triggered default	Move
Intercept	-0.4786	-6.0533***
	(0.5620)	(0.4333)
Payment ($1000)	1.5963***	
	(0.2798)	
Income ($1000)	-0.4203***	
	(0.0503)	
Household size	0.1219***	-0.1482**
	(0.0223)	(0.0815)
Credit score/1000	-10.3991***	
	(0.7893)	
Minority	0.0562	-12.0282
	(0.0784)	(12.4938)
County unemployment rate	14.4277***	
	(1.5537)	
Employment growth rate diff (local - national)		-21.5773***
		(7.4877)
DPA grant	0.0530	-0.1186
	(0.0618)	(0.2195)
Age		-3.8222***
		(1.3392)
Female		-0.4824***
		(0.2035)

Panel B	Strategic Default	Refinance
Intercept	0.4232	-10.3161***
	(1.3449)	(0.3163)
$defaultOpt_{i,t}$ ($1000) $\times I(\geq 0)$	0.0298***	-0.0693***
	(0.0024)	(0.0071)
$refiOpt_{i,t}$ ($1000) $\times I(\geq 0)$	-0.0293	0.1051***
	(0.0176)	(0.0041)
$defaultOpt_{i,t}$ ($1000) $\times I(< 0)$	0.1998**	0.0374*
	(0.1133)	(0.0259)
$refiOpt_{i,t}$ ($1000) $\times I(< 0)$	-0.0200***	0.2656***
	(0.0026)	(0.0488)
Other state variables & their interactions included	Yes	Yes

Note: * denotes significance at a 10% level. ** denotes significance at a 5% level. *** denotes significance at a 1% level. The standard errors are in parentheses.

TABLE V

State Transitions

MSA code	MSA name	Intercept	1 month lag
Dependent variable: difference of market mortgage interest rate			
	NATION	-	0.4080***
		-	(0.0418)
Dependent variable: log difference of house price index			
10420	Akron	0.00175***	0.8288***
		(0.000562)	(0.02711)
15940	Canton-Massillon	0.00224***	0.84601***
		(0.000578)	(0.02553)
17140	Cincinnati-Middletown	0.000833	0.81338***
		(0.000566)	(0.02797)
17460	Cleveland-Elyria-Mentor	0.000498	0.80605***
		(0.00075)	(0.02888)
18140	Columbus	0.00067	0.79297***
		(0.000567)	(0.02652)
19380	Dayton	0.00107	0.83213***
		(0.000689)	(0.0269)
30620	Lima	0.00176***	0.893***
		(0.000564)	(0.02023)
31900	Mansfield	0.000968**	0.90509***
		(0.000461)	(0.02074)
41780	Sandusky	0.000422	0.91201***
		(0.000562)	(0.01912)
44220	Springfield	0.000438	0.88511***
		(0.000454)	(0.01888)
45780	Toledo	0.0012**	0.85004***
		(0.00066)	(0.02553)
49660	Youngstown-Warren-Boardman	0.0024***	0.87554***
		(0.000608)	(0.02359)
	Ohio State	0.000892*	0.83659***
		(0.00048)	(0.02676)

Note: * denotes significance at a 10% level. ** denotes significance at a 5% level. *** denotes significance at a 1% level. The standard errors are in parentheses.

TABLE VI

The 2nd Stage Estimation

$$u_{i,t}(x_{i,t}, a_{i,t} = c) = -\alpha_1 \cdot payment_i + \alpha_2 \cdot HouseStock_i + \varepsilon_{i,t}^{c}$$
$$u_{i,t}(x_{i,t}, a_{i,t} = r) = -\alpha_3 - \alpha_1 \cdot payment\_refi_{i,t} + \alpha_2 \cdot HouseStock_i + \varepsilon_{i,t}^{r}$$
$$v_{i,t}(x_{i,t}, a_{i,t} = sd) = -\alpha_4 - \alpha_5 \cdot HomeValue_{i,t} + \varepsilon_{i,t}^{sd}$$

	1. AR(1) State Transitions	2. Subjective Expectations
α_1	20.6338***	19.9091***
	(0.4684)	(0.4745)
α_2	1.8788***	2.0197***
	(0.2735)	(0.2643)
α_3	6.8035***	6.7271***
	(0.0421)	(0.0426)
α_4	0.3038**	0.2161**
	(0.1572)	(0.1247)
α_5	3.3895**	2.2466*
	(1.5095)	(1.4440)

Notes. * denotes significance at a 10% level. ** denotes significance at a 5% level. *** denotes significance at a 1% level. The standard errors are in parentheses

TABLE VII

Counterfactual Analyses: Loan Modification

Counterfactual Scenarios	Illiquidity-triggered default rate	Strategic default rate
Baseline	8.37%	2.88%
Principal Writing down 10%	-0.79%	-1.50%
Principal Writing down 20%	-1.51%	-2.18%
Interest reduction 1%	-0.87%	-0.23%
Term extension to 40 year	-0.68%	+0.33%