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

Chao Ma[2]

November 18, 2014

Abstract

When households decide whether to default or refinance on their mortgages, their beliefs about future house prices and interest rates matter. This motivates estimating a dynamic discrete choice model. There are two types of default generated by two different mechanisms: illiquidity-triggered default (being forced to default by illiquidity) and strategic default (strategically choosing to default due to financial incentives). However, researchers can only observe default, but cannot identify whether it is illiquidity-triggered or strategic. Moreover, usually researchers can only observe prepayment, but cannot identify whether it is due to refinancing or moving. In this paper, I extend the conditional choice probability (CCP) method to estimate a dynamic discrete choice model with partially observable outcomes. Exclusion restrictions provide identifying conditions: some variables are only related to the incentive to pay but not the ability to pay; and some variables are only related to the incentive to refinance but not the decision to move. The model yields separate predictions of the probabilities of illiquidity-triggered default, strategic default, refinancing, and moving. Counterfactual analysis for foreclosure-mitigating loan modification policies shows that writing down the principal can reduce both illiquidity-triggered default and strategic default. Interest rate reduction can reduce illiquidity-triggered default, but cannot effectively reduce strategic default. Term extension can reduce illiquidity-triggered default, but will increase strategic default, as households will have a longer period with low or even negative home equities.

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

[1]I would like to thank Jason Blevins, Donald Haurin, Stephen Cosslett, Victor Aguirregabiria, Robin Lee, Chris Conlon, Igal Hendel, Michael Dickstein, N. Edward Coulson, Morris Davis, Marc Francke, David Blau, Lucia Dunn, Lung-Fei Lee, Robert De Jong, Matt Lewis, Maryam Saeedi, Javier Donna, Daeho Kim, Huanxing Yang, Jing Zhang, as well as participants at Econometrics Seminar and Applied Microeconomics Seminar at Ohio 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, 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]Department of Economics, Ohio State University. Email: ma.271@buckeyemail.osu.edu

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

1 Introduction

In the 2007-2009 financial crisis, we observe two significant phenomena in the residential mortgage market: 1) high default rates (as shown in Figure 1); 2) low interest rates and high prepayment rates (as shown in Figure 2). Precisely predicting the probabilities of default and prepayment on mortgages could improve the pricing of the risks of default and prepayment during the following four stages: 1) when mortgages are originated by banks; 2) when the securitization trustees, such as Fannie Mae and Freddie Mac, purchase mortgages from the banks; 3) when the securitization trustees sell mortgage-backed securities to the investors; and 4) when the mortgage-backed securities are traded in the secondary market. Thoroughly understanding what factors could affect households' default and prepayment decisions and how they affect those decisions will help financial institutions control the risks and help the policy maker design interventions mitigating foreclosure.

There are two options embedded in a mortgage. The default choice is a put option, and the refinance choice is a call option. Kau, Keenan, Muller and Epperson (1995) developed a theoretical option-pricing model to evaluate a mortgage. In empirical studies, such as Deng (1997) and Deng, Quigley and Order (2000), a measure of the financial incentive to default can be 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. If the house value is below the unpaid mortgage balance, then the house is underwater and the household has a positive financial incentive to default. A measure of the financial incentive to refinance can be 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 the existing mortgage for household i, and $payment_i$ is the monthly payment flow in the future. The first term 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. If the current market mortgage interest rate is lower than the coupon rate of the household's existing mortgage, then the household has a positive financial incentive to refinance.

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

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

Most of the previous research on mortgage termination was based on static models.[3] However, in reality, households should make decisions in a dynamic setting where their' expectation about the future can affect their choices in the current period. First, suppose in the current period the house prices are low and the households' houses are underwater. If the forward-looking households forecast that the house prices will rebound in the future, they may not default because it will incur a large amount of a fixed cost to them, including losing their homes, moving, and ruining their credit history and reputation. If they forecast that the house prices will not rebound in the future, they will be more likely to choose to default because, given that they will have to default anyway, to default earlier to avoid making payments in more periods would be better than to default later. Second, suppose the current interest rate is attractively low and 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 be left then. If the households forecast that the

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

interest rate will drop further, they may wait, rather than 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; but each time they refinance, they would need to pay a fixed transaction cost, including an application fee for the new loan, appraisal fee for their houses, searching cost for the new loan, and possible prepayment penalty.

There have been a lot of papers estimating dynamic discrete choice models for consumer purchasing behavior: Gowrisankaran and Rysman (2012), Shiraldi (2011), Song and Chintagunta (2003), Conlon (2012), and Erdem, Keane, Oncu and Strebel (2005) on durable goods, Hendel and Nevo (2006) and Wang (2013) on storable goods, Lee (2013) on two-sided markets, and so on. However, only a few papers estimate dynamic discrete choice models for household mortgage default and prepayment behavior, including Bajari, Chu, Nekipelov and Park (2013) and Carranza and Navarro (2010).

Besides the dynamic issue, partial observability of the dependent variables is another important issue. There are two types of default due to two different reasons. The first is strategic default (or ruthless default), which is caused by the financial incentive to default. In this case, default provides the households with higher utility than refinance or continuing to pay. The second type is illiquidity-triggered default, which is caused by illiquidity; i.e., the households do not have enough money to make the monthly mortgage payment. In this case, default may not necessarily provide the households with higher utility than refinance or continuing to pay, because the latter two choices are out of the feasible choice set due to the liquidity constraint. Unfortunately, in mortgage loan performance data sets researchers can only observe people's default action, but cannot observe which reason causes the default. This is the partial observability problem for default. Most previous research estimated reduced form models which include in one equation both the measure for financial incentive to default and the variables related to illiquidity that could trigger default, without treating the illiquidity-triggered default outcome and the strategic default outcome separately. For example, Deng, Quigley and Order (2000) included in a single equation a measure for the financial incentive to default and the local unemployment rates related to illiquidity. Because the two types of default outcome are generated by two different mechanisms (one is that the households strategically choose to default over refinance or continuing to pay, the other is that the households are forced to default due to liquidity constraint binding), if one estimates a structural dynamic model without treating them separately, the structural parameters will be biased; moreover, the model will not be able to separately predict the probability of illiquidity-triggered default and the probability of strategic default. Poirier (1980) firstly developed the econometric methodology to estimate bivariate probit models with partial observability. Following his methodology, Bajari, Chu and Park (2010) place the measure of financial incentive to default and the measure of illiquidity into two different equations and modeled default as the outcome of a two-equation system taking into consideration 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. Bajari, Chu, Nekipelov and Park (2013) estimated a dynamic model and treated the strategic default and the illiquidity-triggered default as the unobserved heterogeneity among borrowers. The difference between partial observability examined in my paper and unobserved heterogeneity used in many other papers is that partial observability here is about dependent variables, while unobserved heterogeneity is about unobserved explanatory variables.

There are also two types of prepayment due to two different reasons. The first is refinance, where the households prepay their current mortgage and originate a new mortgage for the same houses 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, where the households sell their houses and prepay their current mortgages without refinancing. And the moving is due to reasons other than the financial incentive to refinance. Unfortunately in most mortgage loan performance data sets researchers can only observe households' prepayment decisions and do not observe whether they are caused by refinancing or moving. Most previous research either does not distinguish these two types of prepayment, or just assumes that all of the prepayments are for refinancing. An, Clapp and Deng (2010) merged a loan performance data including 1985 mortgages with the residential real estate transaction data in California by the street address to separately identify moving and refinance. Performing a) the estimation pooling moving and refinance together as one choice (prepayment) and b) the estimation distinguishing moving and refinance as two different choices, they found that the results are different. However, for most mortgage loan performance data, especially large data sets, it is either costly or infeasible to identify moving and refinance by merging them with the residential real estate transaction data. I am unaware of any financial institution practising this in their business. Bajari, Chu, Nekipelov and Park (2013) assumed that all of the prepayments were caused by refinance when they estimated their dynamic model.[4]

Incorporating partial observability of dependent variables into a dynamic structural model will not only provide a more precise prediction of default rates and prepayment rates, but also yields better counterfactual analyses for policy interventions. 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 the illiquidity-triggered default and the strategic default. Interest rate reduction can effectively reduce the illiquidity-triggered default as it can reduce the monthly payment amount, but it cannot effectively reduce the strategic default as it cannot dramatically reduce the financial incentive to default. Term extension can effectively reduce the illiquidity-triggered default, but will increase the strategic default as the households will have a longer period with low or even negative home equities.

In this paper I first estimate a multinomial logit model for default and prepayment. In addition to the loan and borrower characteristic variables usually used by the previous research, I include a measure of the households' expectation on future house price appreci-

[4] Besides refinance and sale, there is a third type of prepayment, where the household just prepay without refinance or selling the house. Due to data limitation, this type of prepayment is not considered in the literature of mortgage termination. It is even less likely for the borrowers in the loan performance data used in my paper because all of them are low-to-moderate income first-time homebuyers.

ation rate constructed using the data from the Michigan Survey of Consumers.[5] The result shows that the probability of default will be significantly lower if the households expect a higher house price appreciation rate in the future, which verifies the motivation to estimate a dynamic choice model for the forward-looking households.

There is a rich econometric literature on estimating dynamic models.[6] The most important innovation of my paper on econometric methodology is that I incorporate partial observabilities into the CCP method, and use the exclusion restriction of the variables only related to the exogenous events and those only related to the decision making of agents to ensure the identification of the parameters. Another innovation of my paper is that, besides the usual method dealing with the state transitions in the dynamic model estimation literature (assuming the rational expectation of households and a structure of the state transitions, and using the realized data to estimate the structure in the first step), following the spirit of Manski (2004), I use the data from the Michigan Survey of Consumers on people's subjective expectation about the future house prices for the state transitions, which does not require assuming the rational expectation of households and a structure of the state transitions to infer households' preferences.

The remaining portion of this paper is organized as follows. In Section 2, I construct

[5] Bajari, Chu and Park (2011) used three measures for people's expectation on future house price appreciation rate in their static regressions. The first is the average of realized appreciation rates over the past 12 months. The second measure is the average of realized appreciation rates over the next 12 months, based on the perfect foresight assumption – an extreme case of rational expectation. The third one is implied by the equality that the user cost of homeownership is equal to the rent. Different from that paper, I directly use the subjective survey data to construct the measure for people's expectation on future house price appreciation rate.

[6] 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 structural parameters. Hotz and Miller (1993) developed the conditional choice probability (CCP) method which dramatically reduces the computation burden of computing the value function. Hotz and Miller (1994) developed a simulation-based CCP method which is suitable for the problems with large state space. Aguirregabiria and Mira (2002) developed a recursive CCP method. Arcidiacono and Miller (2011) incorporated unobserved heterogeneity into the CCP method. Arcidiacono, Bayer, Blevins and Ellickson (2014) extended the CCP method to estimate dynamic choice models in continuous time. Bajari, Benkard, and Levin (2006) developed the BBL method which works for both the single agent dynamic problem and dynamic game, the finite horizon problem and infinite horizon problem, and the discrete choice problem and continuous choice problem. More details of this literature can be found in the survey paper by Aguirregabiria and Mira (2010).

the dynamic discrete choice model for the households. 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

The processes modeled in this paper generating the mortgage performance outcomes are displayed in Figure 3. In each period t, there is an exogenous process that determines whether the household i faces illiquidity(denoted as td). Illiquidity-triggered default will happen if and only if $I\left\{z_{i,t}\varphi + e^1_{i,t} < 0\right\} = 1$, where $z_{i,t}$ includes the monthly mortgage payment amount, monthly income, household size, credit score, county level unemployment rate, minority indicator, whether the household took downpayment assistance (DPA) grant. And φ are the parameters to be estimated. The probability of illiquidity-triggered default is given by Equation (3). $z_{i,t}\varphi + e^1_{i,t} < 0$ can be viewed as the household's liquidity constraint. The liquidity constraint is more likely to bind if the monthly expenditure is high. Mortgage payments are part of the monthly expenditure. Household size is positively correlated with the monthly expenditure. The liquidity constraint is less likely to bind if the household has a higher income, higher savings level, and higher accessibility to other loans. The household that chose to take DPA grant may have a lower savings level. The credit score is a measure of the household's accessibility to other loans. The household is more likely to encounter a shortfall of income if the local unemployment rate is high.

$$\Pr\{Illiquidity\_triggered\_default_{i,t}\} = \Pr\left\{z_{i,t}\varphi + e^1_{i,t} < 0\right\} = p^{td}_t(z_{i,t};\varphi) \quad (3)$$
$$= \frac{\exp\{-z_{i,t}\varphi\}}{1+\exp\{-z_{i,t}\varphi\}} = \frac{1}{1+\exp\{z_{i,t}\varphi\}}$$

If the illiquidity-triggered default does not happen, then the household will determine 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 $I\left\{w_{i,t}\psi + e^2_{i,t} < 0\right\} = 1$, where $w_{i,t}$ include the difference between the county employment growth rate and the national one, minority indicator, whether the household took the DPA grant, household size, and age. And ψ are the parameters to be estimated. I assume if both $I\left\{z_{i,t}\varphi + e^1_{i,t} < 0\right\} = 1$ and $I\left\{w_{i,t}\psi + e^2_{i,t} < 0\right\} = 1$, then the illiquidity-triggered default happens rather than move-sell-prepay. For simplicity, I also assume e^1 and e^2 are independent from each other. The probability of move-sell-prepay is given by Equation (4).[7]

Archer, Ling, and McGill (1996) categorized the mobility-driven factors related to mortgage termination into two classes: the location-decision factors and the response-to-housing-disequilibrium factors. The difference between the county employment growth rate and the national one belongs to the first class. People may want to move for new opportunities if there are more opportunities in other places than the place in which they currently reside. Household size also belongs to both of the two classes. First, larger households have less mobility. Second, households of different sizes have different probabilities of restructuring. The empirical mobility literature, such as Quigle (1987) using the Panel Study of Income Dynamics [PSID] data and Ferreira, Gyourko and Tracy (2010) using the American Housing Survey (AHS) data, have provided evidence that larger households, elder households, and

[7] Similar to strategic default and refinance on mortgages, moving should also be a dynamic choice by the forward looking households. However, because of data limitation, it is impossible to estimate a dynamic discrete choice model for moving together with the dynamic choice of default, refinance or continuing to pay on mortgages. Therefore, I model moving as a static choice and assume it is independent from the choice of strategic default, refinance, or continuing to pay on mortgages

minorities have significantly lower probabilities of moving.

I also include the indicator of whether the household took the DPA grant as a variable related to the moving probability. In this mortgage sample, a borrower could take a grant up to 3% of the home's purchase price to pay for the downpayment, origination fee, appraisal fee, and other closing costs. If the borrower took the DPA grant, his interest rate would be increased by 0.5%. Taking the grant is equivalent to choosing low points and high interest rates, while not taking the grant is equivalent to choosing high points and a low interest rates.[8] A borrower expecting to move in the near future would choose a mortgage with low points and a high interest rate, because if he chose high points and low interest rate he would not be able to continue to take advantage of low interest rate after moving. But a borrower expecting to stay in the house over the mortgage horizon would choose a mortgage with high points and low interest rate. Stanton and Wallace (1998) derived a separating equilibrium model, where the borrowers self-select the points and interest rates.[9]

$$
\begin{aligned}
\Pr\{Move_{i,t}\} &= \Pr\left\{z_{i,t}\varphi + e^1_{i,t} \geq 0 \cap w_{i,t}\psi + e^2_{i,t} < 0\right\} \\
&= \Pr\left\{z_{i,t}\varphi + e^1_{i,t} \geq 0\right\} \Pr\{w_{i,t}\psi + e^2_{i,t} < 0\} \\
&= p^m_t(z_{i,t}, w_{i,t}; \varphi, \psi) = \frac{\exp\{z_{i,t}\varphi\}}{1+\exp\{z_{i,t}\varphi\}} \frac{1}{1+\exp\{w_{i,t}\psi\}}
\end{aligned} \tag{4}
$$

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

[9] In most mortgage loan performance data, points are not observable, and only the interest rates are observable. Some empirical papers used a two-step regression for the prepayment decision with the effect of the unobserved point choice. In the first step, they run 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 LTV, payment-to-income ratio, and credit score. And then they use the residual as an measure of points for the second step (the regression for the prepayment choice). A higher residual in the first step means a higher interest rate relative to the market, and implies that the borrower may choose low points. However, the problem of this method in the existing literature is that the first step estimation has endogeneity bias, which will make the second step estimation inconsistent. It is true that borrowers with high payment-to-income ratio and high LTV have higher risk of default, thus the lender will offer high interest rates to them. However, it could also be true that borrowers offered low interest rates are more willing to borrow more, thus they have higher payment-to-income ratio and higher LTV. One advantage of the loan performance data set used in my paper is that it has the information about whether the borrowers take the DPA grant. Therefore, I do not need to use the estimated points, which have endogeneity bias.

If neither illiquidity nor moving does happen to the household in the previous two processes, the household will make a choice among continuing to pay on the mortgage (denoted as c), refinance (denoted as r), and 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).

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

The household's utility maximizing problem is formulated in Equation (6).[10] $x_{i,t}$ are the variables in the period utility functions. $\{s_{i,t}\} = \{z_{i,t}, w_{i,t}, x_{i,t}\}$, which include all of the state variables. The transition of the state variables is given by Equation (7), where the state variables in the next period $s_{i,t+1}$ is functions of the state variables and actions in the current period, and random shocks in the next period. The choice-specific period utility functions are given by Equation (8) in detail. $payment_i$ is the monthly payment amount of the 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 current market average interest rate, the new loan amount they borrow will be equal to the unpaid balance of the old mortgage, and the term of the new mortgage will be equal to the horizon remaining for the old mortgage. If refinance or continue to pay, the household will obtain disutility from making monthly payments and obtain utility from living in its home. The utility living in its home is positively related to

[10] This is a partial equilibrium model only considering the households' default decisions given the house price level. It does not take into consideration that the households' default decisions can also affect the house price level. General equilibrium models are difficult to estimate. Macroeconomists usually use calibration, rather than estimation, to analyze general equilibrium models. Corbae and Quintin (2014) calibrated a dynamic general equilibrium model with heterogeneous households making decisions on mortgage selection and default. Even though I estimate a partial equilibrium model, I include the local economic variables in the estimation to mitigate the endogeniety.

the home quality. I use the purchase price of the house deflated by the CPI at purchase as a measure of the quality of the house. If defaults, the household will ruin its credit history and lose its home equity. A higher credit score can provide the household with more accessibility to other loans. Thus, the higher the credit score and home equity the household has, the more utility the household will obtain from keeping the mortgage alive.[11] [12] I also include intercepts in the period utility functions of refinance and that of default. The reason is that, for refinance, there are unobserved transaction costs;[13] for default, besides ruining its credit history and losing its home, the household will incur the costs of moving to a new home and being socially stigmatized, which are also unobservable. $V_{i,t}(s_{i,t})$ is the expectation of the ex ante 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 of continuing to pay the mortgage, refinance, or strategic default, it takes into consideration that in the

[11] In the period utility functions, firstly, I assume that the coefficient of mortgage payment amount for continuing to pay is the same as that for refinance, because the household should obtain the same amount of disutility from paying a dollar no matter whether it chooses to refinance or continue to pay. Secondly, I assume that the coefficient of home quality for continuing to pay is the same as that for refinance, because no matter the household refinances or continues to pay, it lives in the same house, thus should obtain the same servicing utility from the house. Thirdly, I assume that the coefficient of credit score for continuing to pay is different from that for refinance, because a high credit score is more valuable for the household in the period when they need to borrow from banks.

[12] In the literature of labor economics, using panel data with individuals' consumption levels, such as PSID and the Health and Retirement Study (HRS), some papers, such as Blau (2008), estimated dynamic life cycle models in which the period utility is a function of consumption. In the literature of housing economics, Bajari, Chan, Krueger and Miller (2013) used PSID data and estimated a dynamic model of housing demand where the period utility is a function of non-durable consumption and housing stock. However, for mortgage default and prepayment, it is difficult to estimate a dynamic life cycle model with the 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 the peroid utility function. Laufer (2011) and Zhang (2010) used some assumptions to imply the individual's consumption level and included it in the period utility function. Laufer (2011) matched the consumption levels of the households in PSID to the households in his loan performance data with similar characteristics. Zhang (2010) assumed that the households have zero savings. In the literature of industrial organization, the period utility functions of consumers in dynamic choice models usually do not have consumption.

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

next period there will still be positive probabilities of illiquidity-triggered default and move, and the household assigns values to them. Theoretically, this model is estimatable. And I provide the estimating routine in Section 3. However, because default, refinance and move are all rare events, $p_{t+1}^{td}(z_{i,t+1})$ and $p_{t+1}^{m}(z_{i,t+1}, w_{i,t+1})$ are both very small, which makes the estimation result unstable. Moreover, from the behavioral aspect, when making a decision in the current period on whether default or continue to pay, most people in normal status may not take into consideration the minute probability of illiquidity in the future nor assign a value to it. Instead, they may simply ignore it.[14] Thus, when estimating the model, I assume $E_t[p_{t+1}^{td}(z_{i,t+1})V_{t+1}^{td}(s_{i,t+1}) \mid z_t, a_{i,t}] = 0$ and $E_t[p_{t+1}^{m}(z_{i,t+1}, w_{i,t+1})V_{t+1}^{m}(s_{i,t+1}) \mid z_t, w_t, a_{i,t}] = 0$.

$$V_{i,t}(s_{i,t}) = E_\varepsilon \max_{a_{i,t}\in\{sd,r,c\}} \left\{ \begin{array}{c} u_{i,t}(x_{i,t}, a_{i,t} = c) + \beta E_t[p_{t+1}^{td}(z_{i,t+1})V_{t+1}^{td}(s_{i,t+1}) \mid z_{i,t}, a_{i,t} = c] \\ +\beta E_t[p_{t+1}^{m}(z_{i,t+1}, w_{i,t+1})V_{t+1}^{m}(s_{i,t+1}) \mid z_{i,t}, w_{i,t}, a_{i,t} = c] \\ +\beta E_t\left[p_{t+1}^{3}(z_{i,t+1}, w_{i,t+1})V_{i,t+1}(s_{i,t+1}) \mid s_{i,t}, a_{i,t} = c\right], \\ u_{i,t}(x_{i,t}, a_{i,t} = r) + \beta E_t[p_{t+1}^{td}(z_{i,t+1})V_{t+1}^{td}(s_{i,t+1}) \mid z_{i,t}, a_{i,t} = r] \\ +\beta E_t[p_{t+1}^{m}(z_{i,t+1}, w_{i,t+1})V_{t+1}^{m}(s_{i,t+1}) \mid z_{i,t}, w_{i,t}, a_{i,t} = r] \\ +\beta E_t\left[p_{t+1}^{3}(z_{i,t+1}, w_{i,t+1})V_{i,t+1}(s_{i,t+1}) \mid s_{i,t}, a_{i,t} = r\right]], \\ u_{i,t}(x_{i,t}, a_{i,t} = sd) \end{array} \right\} \tag{6}$$

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

[14] There could be a small number of exceptions. For example, a worker might anticipate unemployment because the factory is closing down, and then stop making mortgage payment immediately before that event occurs.

$$u_{i,t}(x_{i,t}, a_{i,t} = c; \alpha) = \alpha_0 \cdot payment_i + \alpha_1 \cdot HomeQuality_i + \alpha_3 \cdot Credit_{i,t} + \varepsilon^c_{i,t} \tag{8}$$

$$u_{i,t}(x_{i,t}, a_{i,t} = r; \alpha) = \alpha_2 + \alpha_0 \cdot payment\_refi_{i,t} + \alpha_1 \cdot HomeQuality_i + \alpha_4 \cdot Credit_{i,t} + \varepsilon^r_{i,t}$$

$$u_{i,t}(x_{i,t}, a_{i,t} = sd; \alpha) = \alpha_5 + \varepsilon^{sd}_{i,t}$$

3 Data and Descriptive Statistics

3.1 The Loan Performance Data

The loan performance data comes from Mortgage Revenue Bonds data maintained by the Ohio Housing Finance Agency (OHFA).[15] 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 contains the monthly payment history through March 2011.[16]

As shown in Table I, up to March 2011, 11.25% of these borrowers defaulted, and 10.91% prepaid. The data includes the information of the borrower's credit score, monthly income, and the price of the house at purchase, but does not track the change of them over time. The data also contains the mortgage characteristics such as the coupon interest rate, loan amount, monthly payment amount and initial LTV; and the demographic characteristics such as household size, location, race, and ethnicity.[17] The descriptive statistics of important

[15] 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".

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

[17] 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 can not be extended to the whole population. But the model and methodology developed in this paper are extendable to other loan performance data sets, such as the data maintained by Corelogic, containing 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 result of default

variables are displayed in Table II.

3.2 The Macroeconomic Data

The monthly MSA level house price index data comes from the Freddie Mac Housing Price Index (FMHPI). The monthly mortgage market interest rate data comes from the Freddie Mac Primary Mortgage Market Survey (PMMS). Together with individual mortgage data, using Equation (1) and Equation (2), I can compute $defaultOpt_{i,t}$ and $refiOpt_{i,t}$, whose descriptive statistics are displayed in Table II.[18] 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 surveys around 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 on the average house price growth rates over the next year and over the next five years. On average, each month there are 22 people from Ohio reporting their responses to the house price expectation questions. Figure 4 shows the means and standard deviations of the Ohio respondents' expectations on the annual house price growth rate over the next year and the next 5 years within each survey month. 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 The Econometric Methodology

and prepayment risks obtained from this sample is still helpful in pricing the Mortgage Revenue Bonds Securities. The advantage of this data is that it has information on some characteristics of the households, such as household size, which other loan performance data does not have.

[18] The loan performance data does 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 to $Homeprice_{i,t}$.

I incorporate partial observability into 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 are used to guarantee the identification of the parameters. In the first stage, the parameters in the probabilities of illiquidity and moving (the exogenous events) and the parameters in the conditional choice probabilities of strategic default, refinance, and continuing to pay (the policy functions) are estimated. The parameters in the structure of state transitions are estimated, as well. In the second stage, the structural parameters in the dynamic model are estimated.

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

$$
\begin{aligned}
v^c_{i,t}(s_{i,t}) &= \widetilde{u}^c_{i,t}(x_{i,t}) + \beta E_t[p^{td}_{t+1}(z_{i,t+1})V^{td}_{t+1}(s_{i,t+1}) \mid s_{i,t}, a_{i,t} = c] \\
&+\beta E_t[p^{m}_{t+1}(z_{i,t+1}, w_{i,t+1})V^{m}_{t+1}(s_{i,t+1}) \mid s_{i,t}, a_{i,t} = c] \\
&+\beta E_t\left[p^{3}_{t+1}(z_{i,t+1}, w_{i,t+1})V_{i,t+1}(s_{i,t+1}) \mid s_{i,t}, a_{i,t} = c\right], \\
v^r_{i,t}(s_{i,t}) &= \widetilde{u}^r_{i,t}(x_{i,t}) + \beta E_t[p^{td}_{t+1}(z_{i,t+1})V^{td}_{t+1}(s_{i,t+1}) \mid s_{i,t}, a_{i,t} = r] \\
&+\beta E_t[p^{m}_{t+1}(z_{i,t+1}, w_{i,t+1})V^{m}_{t+1}(s_{i,t+1}) \mid s_{i,t}, a_{i,t} = r] \\
&+\beta E_t\left[p^{3}_{t+1}(z_{i,t+1}, w_{i,t+1})V_{i,t+1}(s_{i,t+1}) \mid s_{i,t}, a_{i,t} = r\right], \\
v^{sd}_{i,t}(s_{i,t}) &= \widetilde{u}^{sd}_{i,t}(x_{i,t}).
\end{aligned}
$$

Given the state variables $s_{i,t}$, and conditional on the household not facing illiquidity or moving, because of the assumption that the idiosyncratic shocks follow Type I extreme value distribution, the equilibrium probabilities of continuing to pay, refinance, and default can be written as in Equation (9)

$$
\begin{aligned}
\sigma^c_{i,t}(s_{i,t}) &= \Pr(a_{i,t}=c) = \frac{\exp\left(v^c_{i,t}(s_{i,t})\right)}{\sum\limits_{k\in\{c,r,sd\}} \exp\left(v^k_{i,t}(s_{i,t})\right)} \\
\sigma^r_{i,t}(s_{i,t}) &= \Pr(a_{i,t}=r) = \frac{\exp\left(v^r_{i,t}(s_{i,t})\right)}{\sum\limits_{k\in\{c,r,sd\}} \exp\left(v^k_{i,t}(s_{i,t})\right)} \\
\sigma^d_{i,t}(s_{i,t}) &= \Pr(a_{i,t}=sd) = \frac{\exp\left(v^{sd}_{i,t}(s_{i,t})\right)}{\sum\limits_{k\in\{c,r,sd\}} \exp\left(v^k_{i,t}(s_{i,t})\right)}
\end{aligned} \tag{9}
$$

By the Hotz-Miller inversion, Equation (10) can be derived, where $v^{sd}_{i,t}(s_{i,t}) = \widetilde{u}^{sd}_{i,t}(x_{i,t})$.

$$
\begin{aligned}
v^c_{i,t}(s_{i,t}) - v^{sd}_{i,t}(s_{i,t}) &= \log\left(\frac{\sigma^c_{i,t}(s_{i,t})}{\sigma^{sd}_{i,t}(s_{i,t})}\right) \\
v^r_{i,t}(s_{i,t}) - v^{sd}_{i,t}(s_{i,t}) &= \log\left(\frac{\sigma^r_{i,t}(s_{i,t})}{\sigma^{sd}_{i,t}(s_{i,t})}\right)
\end{aligned} \tag{10}
$$

Given the extreme value assumption for the idiosyncratic error terms, the ex ante value function can be written as in Equation (11).[19]

$$
V_{i,t}(s_{i,t}) = \log\left(\exp\{v^c_{i,t}(s_{i,t})\} + \exp\{v^r_{i,t}(s_{i,t})\} + \exp\{v^{sd}_{i,t}(s_{i,t})\}\right) = \log\left(\frac{1}{\sigma^{sd}_{i,t}(s_{i,t})}\right) + \widetilde{u}^{sd}_{i,t}(x_{i,t}) \tag{11}
$$

[19] Because this is a finite horizon problem for the household, for each t, $V_{i,t}(S_{i,t})$ should be different from each other. Instead of parameterizing it as $V_{i,t}(S_{i,t}) = V_i(t, S_{i,t})$, I parameterize it as $V_{i,t}(S_{i,t}) = V_i(S_{i,t})$, where some elements of the state variables S are functions of t. For example, $defaultOpt_{i,t}$ is defined as the unpaid loan balance minus the house price. As the mortgage is going toward maturity, the loan balance will decrease, and then $defaultOpt_{i,t}$ will decrease. Therefore, the value function of continuing to pay versus default $(v^c_{i,t}(s_{i,t}) - v^{sd}_{i,t}(s_{i,t}))$ and the vaule function of refinance versus default $(v^r_{i,t}(s_{i,t}) - v^{sd}_{i,t}(s_{i,t}))$ will increase. $refiOpt_{i,t}$ is defined as the present value of the future payment flow discounted by the current market interest rate minus that discounted by the existing mortgage interest rate. As the mortgage is going toward maturity, the magnitude of $refiOpt_{i,t}$ will decrease, i.e., the households can obtain less benefit or incur less loss from refinancing because fewer payments are left. Therefore, the vaule function of refinance versus default $(v^r_{i,t}(s_{i,t}) - v^{sd}_{i,t}(s_{i,t}))$ will decrease.

4.1 The 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 prepay is equal to the sum of the probability of moving and that of refinance. Third, the probability of observing a continuing to pay is equal to the probability of neither illiquidity nor move happening, times the probability of choosing to continue to pay conditional on neither illiquidity nor move happening. The probabilities of these three observable events can be written as in Equation (12).

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

$$\log L = \sum_{i=1}^{N}\sum_{t=1}^{T_i}\left[\begin{array}{c} I\{default_{i,t}\}\log(\Pr\{default_{i,t}\}) + I\{prepay_{i,t}\}\log(\Pr\{prepay_{i,t}\}) \\ +I\{continue_{i,t}\}\log(\Pr\{continue_{i,t}\}) \end{array}\right] \tag{13}$$

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

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). 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 change of these two variables rather than the level of them as AR(1). Second, some state variables, such as race, DPA grant indicator, and monthly payment amount if not refinance, do not change over time. I also assume that income, household size, and credit score do not change over time because I only observe them when the mortgage was originated. Third, the transition of some state variables, such as age and unpaid balance, are deterministic. Fourth, some state variables are functions of other state variables. For example, $defaultOpt_{i,t}$ is a function of unpaid balance, purchase price, house price index, and time to maturity. $refiOpt_{i,t}$ is a function of 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 The Second Stage Estimation

Given $s_{i,t}$ and the 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 Equation (14). For each simulation, with the parameters estimated in the first stage, I recover the ex ante value functions of period $t+1$ after the household chooses to refinance or continue to pay in period t by Equation (15). Then I average all of the simulations to obtain the expected ex ante value of period $t+1$ multiplied by the probability of neither illiquidity nor move happening at period $t+1$ conditional on the state and action in period t, as in Equation (16).

$$V_{i,t+1}(s_{i,t+1}^{k,(j)}) = \log\left(\frac{1}{\sigma_{i,t+1}^{sd}(s_{i,t+1}^{k,(j)};\widetilde{\gamma})}\right) + \widetilde{u}_{i,t+1}^{sd}(x_{i,t+1}^{k,(j)}), k \in \{c, r\} \tag{15}$$

$$\begin{aligned} &\widehat{E_t}\left[p_{t+1}^3(z_{i,t+1}, w_{i,t+1})V_{i,t+1}(s_{i,t+1}) \mid a_{i,t} = k, s_{i,t}\right] \\ &= \frac{1}{J}\sum_{j=1}^{J} p_{t+1}^3(z_{i,t+1}^{k,(j)}, w_{i,t+1}^{k,(j)};\widetilde{\varphi},\widetilde{\psi})V_{i,t+1}(s_{i,t+1}^{k,(j)}), \quad k \in \{c, r\} \end{aligned} \tag{16}$$

Ideally, I could parameterize the value associated with illiquidity-triggered default in period $t+1$ as $V_{t+1}^{td}(s_{i,t+1};\theta)$, and parameterize the value associated with move in period $t+1$ as $V_{t+1}^{m}(s_{i,t+1};\theta)$. Then I average all of the simulations to obtain the expected value of illiquidity-triggered default in period $t+1$ multiplied by the probability of illiquidity-triggered default in period $t+1$ conditional on the state and action in period t, as in Equation (17), and obtain the the expected value of move in period $t+1$ multiplied by the probability of move in period $t+1$ conditional on the state and action in period t, as in Equation (18). Then I obtain the choice-specific value functions in period t, as in Equation (19). Next, the probabilities of observable outcomes (default, prepayment, and continuing to pay) for the second stage estimation are formulated as in Equation (20), where $\widetilde{\varphi}$ and $\widetilde{\psi}$ are the parameters estimated in the first stage, and α and θ are the parameters to be estimated in the second stage. However, the estimation results are unstable. One possible reason is that $p_{t+1}^{td}(z_{i,t+1}^{k,(j)};\widetilde{\varphi})$ and $p_{t+1}^{m}(z_{i,t+1}^{k,(j)}, w_{i,t+1}^{k,(j)};\widetilde{\varphi},\widetilde{\psi})$ are very close to zero because default, prepayment, and move are all rare events. As discussed in Section 2, I set $\widehat{E_t}[p_{t+1}^{td}(z_{i,t+1})V_{t+1}^{td}(s_{i,t+1}) \mid s_{i,t}, a_{i,t} = k] = 0$, $\widehat{E_t}[p_{t+1}^{m}(z_{i,t+1}, w_{i,t+1})V_{t+1}^{m}(s_{i,t+1}) \mid s_{i,t}, a_{i,t} = k]$, and $p_{t+1}^3(z_{i,t+1}^{k,(j)}, w_{i,t+1}^{k,(j)};\widetilde{\varphi},\widetilde{\psi}) = 1$, where $k \in \{c, r\}$. 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 happens in the future if they continue to pay in the current period, because the probabilities of those events are very small.

$$\widehat{E_t}[p^{td}_{t+1}(z_{i,t+1})V^{td}_{t+1}(s_{i,t+1}) \mid s_{i,t}, a_{i,t} = k] = \frac{1}{J}\sum_{j=1}^{J} p^{td}_{t+1}(z^{k,(j)}_{i,t+1}; \widetilde{\varphi})V^{td}_{t+1}(s^{k,(j)}_{i,t+1}; \theta), k \in \{c, r\} \quad (17)$$

$$\begin{aligned} \widehat{E_t}[p^{m}_{t+1}(z_{i,t+1}, w_{i,t+1})V^{m}_{t+1}(s_{i,t+1}) \quad &| \quad s_{i,t}, a_{i,t} = k] \qquad (18) \\ &= \quad \frac{1}{J}\sum_{j=1}^{J} p^{m}_{t+1}(z^{k,(j)}_{i,t+1}, w^{k,(j)}_{i,t+1}; \widetilde{\varphi}, \widetilde{\psi})V^{m}_{t+1}(s^{k,(j)}_{i,t+1}; \theta), \; k \in \{c, r\} \end{aligned}$$

$$\begin{gathered} \widehat{v}^{c}_{i,t}(s_{i,t}; \alpha, \theta) = \widetilde{u}^{c}_{i,t}(x_{i,t}; \alpha) + \beta\widehat{E_t}[p^{td}_{t+1}(z_{i,t+1})V^{td}_{t+1}(s_{i,t+1}) \mid s_{i,t}, a_{i,t} = c] \qquad (19) \\ +\beta\widehat{E_t}[p^{m}_{t+1}(z_{i,t+1}, w_{i,t+1})V^{m}_{t+1}(s_{i,t+1}) \mid s_{i,t}, a_{i,t} = c] \\ +\beta\widehat{E_t}\left[p^{3}_{t+1}(z_{i,t+1}, w_{i,t+1})V_{i,t+1}(s_{i,t+1}) \mid s_{i,t}, a_{i,t} = c\right] \\ \widehat{v}^{r}_{i,t}(s_{i,t}; \alpha, \theta) = \widetilde{u}^{r}_{i,t}(x_{i,t}; \alpha) + \beta\widehat{E_t}[p^{td}_{t+1}(z_{i,t+1})V^{td}_{t+1}(s_{i,t+1}) \mid s_{i,t}, a_{i,t} = r] \\ +\beta\widehat{E_t}[p^{m}_{t+1}(z_{i,t+1}, w_{i,t+1})V^{m}_{t+1}(s_{i,t+1}) \mid s_{i,t}, a_{i,t} = r] \\ +\beta\widehat{E_t}\left[p^{3}_{t+1}(z_{i,t+1}, w_{i,t+1})V_{i,t+1}(s_{i,t+1}) \mid s_{i,t}, a_{i,t} = r\right] \\ \widehat{v}^{d}_{i,t}(s_{i,t}; \alpha, \theta) = \widetilde{u}^{sd}_{i,t}(x_{i,t}; \alpha) \end{gathered}$$

$$\begin{aligned} \widetilde{\Pr}\{default_{i,t}; \alpha, \theta\} &= p^{td}_{t}(z_{i,t}; \widetilde{\varphi}) + \qquad (20) \\ &\quad p^{3}_{t}(z_i, w_i; \widetilde{\varphi}, \widetilde{\psi})\frac{\exp\{\widehat{v}^{sd}_{i,t}(s_{i,t}; \alpha, \theta)\}}{\exp\{\widehat{v}^{sd}_{i,t}(s_{i,t}; \alpha, \theta)\} + \exp\{\widehat{v}^{r}_{i,t}(s_{i,t}; \alpha, \theta)\} + \exp\{\widehat{v}^{c}_{i,t}(s_{i,t}; \alpha, \theta)\}} \\ \widetilde{\Pr}\{prepay_{i,t}; \alpha, \theta\} &= p^{m}_{t}(z_{i,t}; \widetilde{\varphi}) + \\ &\quad p^{3}_{t}(z_i, w_i; \widetilde{\varphi}, \widetilde{\psi})\frac{\exp\{\widehat{v}^{r}_{i,t}(s_{i,t}; \alpha, \theta)\}}{\exp\{\widehat{v}^{sd}_{i,t}(s_{i,t}; \alpha, \theta)\} + \exp\{\widehat{v}^{r}_{i,t}(s_{i,t}; \alpha, \theta)\} + \exp\{\widehat{v}^{c}_{i,t}(s_{i,t}; \alpha, \theta)\}} \\ \widetilde{\Pr}\{continue_{i,t}; \alpha, \theta\} &= p^{3}_{t}(z_i, w_i; \widetilde{\varphi}, \widetilde{\psi})\frac{\exp\{\widehat{v}^{c}_{i,t}(s_{i,t}; \alpha, \theta)\}}{\exp\{\widehat{v}^{sd}_{i,t}(s_{i,t}; \alpha, \theta)\} + \exp\{\widehat{v}^{r}_{i,t}(s_{i,t}; \alpha, \theta)\} + \exp\{\widehat{v}^{c}_{i,t}(s_{i,t}; \alpha, \theta)\}} \end{aligned}$$

The second stage estimation is also performed by MLE. As in Equation (21), $\widetilde{\varphi}$ and $\widetilde{\psi}$ are obtained from the first stage estimation. The structural parameters α are searched for to maximize the log likelihood.

$$\max_{\{\alpha|\widetilde{\varphi},\widetilde{\psi}\}} \log L = \sum_{i=1}^{N}\sum_{t=1}^{T_i}\left[\begin{array}{c} I\{default_{i,t}\}\log(\widetilde{\Pr}\{default_{i,t}\}) + I\{prepay_{i,t}\}\log(\widetilde{\Pr}\{prepay_{i,t}\}) \\ +I\{continue_{i,t}\}\log(\widetilde{\Pr}\{continue_{i,t}\}) \end{array}\right] \tag{21}$$

4.3 Identification

Using the data in which only default and prepayment are observable, I estimate a structural model treating illiquidity-triggered default and strategic default as separate outcomes, as well as treating move and refinance as separate outcomes. Therefore, I use exclusion restrictions to ensure the identification of the parameters in the probabilities of illiquidity-triggered default, strategic default, move, and refinance in the first-stage estimation. One exclusion restriction requires that there exists at least one variable with sufficient variation, $x_{i,t}^{(1)}$, such that $x_{i,t}^{(1)} \in x_{i,t}$ and $x_{i,t}^{(1)} \notin z_{i,t}$. $x_{i,t}^{(1)}$ can be a variable related to the financial incentive to default but not related to illiquidity (e.g., $defaultOpt_{i,t}$ and $Homeprice_{i,t} \cdot \Delta \log hp_t$). Let $x_{i,t}^{(1)} \to +\infty$ or $-\infty$ to make $\sigma_{i,t}^{sd}(s_{i,t};\gamma) \to 0$, then in the first equation in (12), $\Pr\{default_{i,t}\} = p_t^{td}(z_{i,t};\varphi) + \sigma_{i,t}^{sd}(s_{i,t};\gamma) \cdot p_t^3(z_{i,t}, w_{i,t};\varphi,\psi) \to p_t^{td}(z_{i,t};\varphi)$. Thus, φ are identified. The other exclusion restriction requires that there exists at least one variable with sufficient variation, $x_{i,t}^{(2)}$, such that $x_{i,t}^{(2)} \in x_{i,t}$, and $x_{i,t}^{(2)} \notin z_{i,t}, w_{i,t}$. $x_{i,t}^{(2)}$ can be a variable related to the financial incentive to refinance but not related to move (e.g., $refiOpt_{i,t}$). Let $x_{i,t}^{(2)} \to +\infty$ or $-\infty$ to make $\sigma_{i,t}^{r}(s_{i,t};\gamma) \to 0$, then in the second equation in (12), $\Pr\{prepay_{i,t}\} = p_t^{m}(z_{i,t}, w_{i,t};\varphi,\psi) + \sigma_{i,t}^{r}(s_{i,t};\gamma)p_t^3(z_{i,t}, w_{i,t};\varphi,\psi) \to p_t^{m}(z_{i,t}, w_{i,t};\varphi,\psi)$. 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 Equation (23). Because $s_{i,t}$ have full rank, $\gamma = \{\gamma^{sd}, \gamma^{r}\}$ are identified.

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, the 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, $Homeprice_{i,t} \cdot \Delta \log hp_t$ will be high. Consequently, the household's expectation on $Homeprice_{i,t} \cdot \Delta \log hp_{t+1}$ will be very high, because people's expectations on $\Delta \log hp_{t+1}$ follows AR(1) by the assumption and the serial correlation coefficients are above 0.79 for all the MSAs. If the household with a very high expectation of home value appreciation defaults, it should be illiquidity-triggered default, rather than strategic default. Thus, the households with a very high expectation of home value appreciation can also help identify the parameters in the probability of illiquidity-triggered default. Thereafter, the parameters in the 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 the household prepays their mortgage, it should be caused moving rather than refinance. Thus, the households with highly negative $refiOpt_{i,t}$ could help identify the parameters in the probability of moving. Thereafter, the parameters in the probability of refinance could be identified.[20]

$$
\begin{aligned}
\sigma^{sd}_{i,t}(s_{i,t};\gamma) &= \frac{\exp\{s_{i,t}\gamma^{sd}\}}{1+\exp\{s_{i,t}\gamma^{sd}\}+\exp\{s_{i,t}\gamma^{r}\}} \\
\sigma^{r}_{i,t}(s_{i,t};\gamma) &= \frac{\exp\{s_{i,t}\gamma^{r}\}}{1+\exp\{s_{i,t}\gamma^{sd}\}+\exp\{s_{i,t}\gamma^{r}\}} \\
\sigma^{c}_{i,t}(s_{i,t};\gamma) &= \frac{1}{1+\exp\{s_{i,t}\gamma^{sd}\}+\exp\{s_{i,t}\gamma^{r}\}}
\end{aligned} \tag{22}
$$

[20] $refiOpt_{i,t}$ 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, the households will have less mobility due 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 equities, there are many other factors captured in my paper that still can make households move. In the data, there are 252 households that prepaid their mortgages with negative $refiOpt_{i,t}$, 112 of them with highly negative $refiOpt_{i,t}$ (<-\$5000).

$$
\begin{aligned}
\Pr\{prepay_{i,t}\} &= \frac{\exp\{s_{i,t}\gamma^{r}\}}{1+\exp\{s_{i,t}\gamma^{sd}\}+\exp\{s_{i,t}\gamma^{r}\}}(1-p_t^{td}(z_{i,t};\varphi)) \qquad (23)\\
\Pr\{continue_{i,t}\} &= \frac{1}{1+\exp\{s_{i,t}\gamma^{sd}\}+\exp\{s_{i,t}\gamma^{r}\}}\cdot(1-p_t^{td}(z_{i,t};\varphi))
\end{aligned}
$$

4.4 Expectations on Future House Prices

As pointed out by Manski (2004), with data on observed choices, researchers cannot infer both people's preferences and expectations at the same time. If researchers need to infer people's preferences, they have to either make assumptions on people's expectations or obtain additional data on them. Because data on people's expectations are usually unavailable, the first routine is generally used in the literature estimating dynamic choice models. In contrast, Erdem, Keane, Oncu and Strebel (2005) estimated a dynamic choice model for personal computer (PC) purchasing behavior using a survey data on consumers' expectations on future PC prices.

The econometric methodology discussed previously in Section 4 follows the first routine. I assume that people forecast the future house prices and interest rates following the stochastic processes estimated using the realized house prices and interest rates (i.e., rational expectation). However, people's expectations for the future may not follow the rational expectation assumption. First, the process of their expectations on prices may not follow the process of actual price movement. Second, their perception of current prices may deviate from the actual current prices. Morris and Quintin (2013) found that the change of self-assessed house prices is significantly different from that of house price indices calculated based on the actual house transactions, and the former plays a more important role in mortgage defaults than the latter.

Starting in January 2007, the Michigan Survey of Consumers asked subjects each month to provide their expectations on the average house price growth rates over the next year and over the next five years. On average, there are 22 respondents in each month from Ohio. I use this data to represent the house price expectations of the households in my mortgage loan performance data, who also reside in Ohio. When estimating the state transition equations, Erdem, Keane, Oncu and Strebel (2005) used the consumer expectation data on PC prices from survey for the dependent variable on the lefthand side, and used the current and past realized PC price data for the variables on the righthand side. However, different from PC analyzed in Erdem, Keane, Oncu and Strebel (2005), detergent analyzed in Hendel and Nevo (2006), and soda analyzed in Wang (2013), future house prices are not only highly correlated with past house prices, but also with other macroeconomic variables, such as GDP growth, unemployment, income per capita, and so on. One way to solve this problem would be to include other macroeconomic variables on the righthand side of the regression equation, (14). But this method still requires assuming a structure on the state transition process and would not work well if the structure is misspecified.[21] 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 in the Michigan Survey of Consumers as the distribution of people's house price forecasts for the next period, which does not require assuming any structure for the state transition process. Accordingly, for house prices, Equation (16) is the value function averaged over the distribution of the forecasts by the Ohio respondents in the Michigan Survey of Consumers in each month, rather than the distribution of forward simulations based on the estimation result of Equation (14).

[21] In the literature estimating dynamic choice models, the researchers usually assume AR(1) for the structure of state transitions for simplicity. However, in the housing economics literature, many researchers, such as Abraham & Hendershott (1996), Gallin (2006), Gallin (2008), and Malpezzi (1999), use error correction models to model the house price dynamics. It is difficult to implement error correction models as state transitions in the estimation of dynamic choice models, and I am unaware of any paper performing that.

5 The Empirical Results

5.1 The Preliminary Regressions

Table III shows the results of multinomial logit regressions with three possible outcomes for the dependent variable: default, prepayment, and continuing to pay. Model (I) represents the regression without the measure of households' future house price expectations. Model (II) includes the interaction term between $Homeprice_{i,t}$ and the average of the Ohio respondents' expectations on the house price growth rate for the next year in the Michigan Survey of Consumers. Model (III) includes the interaction term between $Homeprice_{i,t}$ and the average of the Ohio respondents' expectations on the average house price growth rate over the next five years. Model (IV) includes both measures.[22] As shown in Model (III), the coefficient of the 5-year house price growth rate expectation for default is significantly negative, and is larger than that for prepayment in magnitude, which indicates that the unconditional probability of default will be lower if people expect larger house price increases in the long run. As shown in Model (II), the coefficient of the 1-year house price growth rate expectation for default is significantly negative, even though smaller than that for prepayment in magnitude, which indicates that at least the probability of default conditional on not prepaying will be lower if people expect large house price increases in the short run. When I include both measures of house price expectations in Model (IV), only the 5-year expectation is significant, which indicates that the long-run house price expectation plays a more important role in mortgage defaults than the short-run house price expectation.

These results provide evidence that households are forward looking when they make decisions about defaulting on their mortgages, which verifies the motivation to estimate a dynamic choice model. The results of the other independent variables also make sense. The probability of default increases as monthly payment, household size (which is positively

[22] Because the mortgage loan performance data used in this paper start in 2005, while the subjective house price expectation data in Michigan Survey of Consumers start in 2007, in the regression of Models (II), (III), and (IV), I delete the observations before 2007 in the mortgage loan performance data.

correlated with household expenditure), local unemployment rate, and default option (put option) value increase, and as income and credit score decrease. The households choosing to take DPA grant are more likely to default because they may have lower savings levels. The probability of prepayment increases as refinance option (call option) value increases. Moreover, the coefficient of default option (put option) value for prepayment is significantly negative. This is consistent with the fact that households will be more likely to be eligible for refinance as they have more home equities. The probability of prepayment increases as income and credit score increase, because households with higher income and credit score are more likely to get offers with lower interest rates for refinance. Besides refinance, move can lead to prepayment. The coefficients of household size and age for prepayment are significantly negative, because larger households and older ones have less mobility. People living in counties with local employment growth rates higher than the national average are less likely to move. The coefficient of minority indicator for prepayment is also negative. It could be either because minorities have less mobility, or because they are less aware of refinance opportunities.

5.2 The First Stage Estimation

Table IV displays the results for the first stage estimation. Larger monthly payment amount, household size, and unemployment rate in the county will make the households significantly more likely to default; and higher income and credit score (a measure of accessibility to other loans) will make the households significantly less likely to defualt. Larger household size, older age, and larger difference between the local employment growth rate and the national average will make the households significantly less likely to move. Minorities have higher probability of illiquidity and lower mobility, but the coefficients are not statistically significant. The reason could be that the sample is special (first-time homebuyers with low-to-moderate income), rather than randomly drawn from the whole population. According to the theory in Stanton and Wallace (1998), the households who select the DPA grant should have higher probability of move. But the coefficient is not significant.[23]

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.[24] In Table IV, I report only the coefficients for some interpretable variables. For strategic default, the coefficient of $defaultOpt_{i,t}$ is significantly positive; and the marginal 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; and the marginal effect of $refiOpt_{i,t}$ is larger in the negative domain than in the positive domain. Besides, the coefficient of $defaultOpt_{i,t}$ for refinance is significantly negative, which indicates that households will be more likely to be eligible for refinance as they have more home equities.[25] The coefficient of credit score for strategic default is significantly negative, as people with higher credit score have more to lose if choose to default. The coefficient of credit score for refinance is significantly positive,

[23] Using AHS data, Ferreira, Gyourko and Tracy (2010) found that negative home equities have a lock-in effect on the households' mobility. In another specification, I include $\max\{0, defaultOpt_{i,t}\}$ in the probability of moving, but the coefficient is positive, rather than negative. One possible reason is that the sample (Ohio low-to-moderate income first-time homebuyers) used in my paper is special.

[24] Some papers applying CCP 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. Theoretically, nonparametric methods could provide less biased estimates for the value functions for the future. In practice, however, if the number of state variables is large, then the curse of of dimensionality will generate very poor estimates. Moreover, the results of nonparametric models in the first stage estimation are not interpretable. In this paper, I use the second method. The first reason is that the number of state variables is large. The second reason is related to the model specification selection. Whether the default is illiquidity-triggered default or strategic default is not observable, and whether the 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, and continuing to pay. Instead, I use flexible parametric forms to estimate the conditional choice probabilities in the first stage and interpret the parameters; and 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 second reason is related to identification. As discussed in Section 3.3, I use exclusion restriction and infinity argument to ensure identification. 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 cannot be checked based on the results obtained in the first stage estimation if nonparametric methods are used, .

[25] In the CCP method, the structure of the state transitions is embedded in the first-step policy function estimates. Because I assume that the state transitions of house price appreciation rate are different across MSAs, the first-step policy functions should also be different across MSAs. However, when I allow that the coefficients of $defaultOpt_{i,t}$ and $refiOpt_{i,t}$ are different across MSAs in the estimation, the result turns out that those coefficients are similar across MSAs.

as people with higher credit score are more likely to get a lower interest offer for refinance. Other state variables in the conditional choice probabilities include interest rate innovation, the MSA house price index change, purchase price, current home value calculated based on the MSA house price index, the derivative of $refiOpt_{i,t}$ with respect to interest rate innovation, and some interaction terms of them.

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 the 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 of interest rate, the change of 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 The Second Stage Estimation

In the second stage estimation, I set the monthly discount factor $\beta = 0.995$.[26] Table VI reports the structural parameters in the second stage estimation using the forward simulation based on the state transitions estimated in Table V. The coefficients are sensible. The coefficient of monthly payment amount is significantly negative, and the coefficient of home quality is significantly positive. If refinance or continue to pay, the household will obtain disutility from making monthly payments and obtain utility from living in its home. The

[26] When estimating dynamic choice models, the discount factor is generally not identifiable. The standard approach in the literature is setting a value for it. 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.

higher quality the house has, the more servicing utility flow the household can obtain from living in the house. If default, the household will ruin its credit history. If not, it should obtain utility from maintaining a high credit score, especially in the period when there is a need to borrow money from banks because borrowers with high credit scores can be offered interest rates lower than the market averages. The coefficient of credit score for continuing to pay is not significant; but that for refinance is significantly positive.[27] The standard errors in the second stage estimation are adjusted, following the correction methods developed by Murphy and Topel (1985).

Table VII reports the structural parameters in the second stage estimation using the state transitions based on people's subjective expectations on house prices extracted from the Michigan Survey of Consumers. The results are similar to those shown in Table VI.

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 Comprehensive Capital Analysis and Review (CCAR) scenarios. I use the first-stage policy function estimates to project the illiquidity-triggered default rate, strategic default rate, refinance rate, and moving rate in those counterfactual scenarios.[28]

[27] In the literature estimating dynamic models with an intercept capturing the unobserved switching cost or transaction cost, the researchers usually convert the intercept into a dollar value by dividing it by the coefficient of price. If I divide the intercept for refinance by the coefficient of payment amount, the dollar value of refinance transection cost is (8.1326/20.4213)*1000=$398, which is very low. In practice, the upfront cost for refinance is usually above $1000. Even though some lenders offer zero-upfront-cost refinance, the interest rates will be higher. One possible reason for this low transaction cost estimates is that when calculating the monthly payment after refinance, I assume the term of the new loan is equal to the remaining term of the old loan. In practice, most people refinance into a mortgage with a standard term such as 30 or 15 years. It is possible that in this low-to-moderate income first-home buyer sample, more people would choose to refinance into a 30-year mortgage. Thus, the actual monthly payment after refinance is lower than my calculation, which could draw down the estimate of the intercept for refinance. The term of the mortgage people refinance into is not observable. And providing a precise estimate for refinance transaction costs is not the focus of this paper.

[28] This method only incurs a reduced computational burden. Bajari, Chu, Nekipelov and Park (2013) provided the justification for this method.

6.1 Loan Modification Policies

After the 2007-2009 financial crisis occured, several loan modification programs aimed at foreclosure mitigation have been designed, such as the Home Affordable Modification Program (HAMP). Borrowers struggling to make mortgage payments due to financial hardships can apply for those programs if they meet certain eligibility criteria. Their monthly payment amount could be reduced through interest rate reduction, term extension, principal forbearance, or/and principal forgiveness. In addition to reducing the expected loss in foreclosure, banks can also benefit from obtaining compensation from the government for modifying a mortgage. The banks participating the HAMP use a net present value (NPV) model as a tool for deciding whether to modify a troubled mortgage. Firstly, the model calculates the probability of default for the troubled mortgage without modification and the probability of default with modification. Secondly, under certain assumptions, it calculates the expected cash flow the bank 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 modified and the present value if unmodified to see whether approving modification for the mortgage gives the bank a positive net present value.[29] While the model has a complex procedure to calculate the cash flow, its methodology for calculating the probability of default is just a simple and intuitive logit model.

The methodology developed in this paper can provide a better estimate of the mortgage default probability for mortgage modification programs. Moreover, it can help the government improve the design of mortgage modification policies on several aspects, such as how to modify a mortgage so as to effectively reduce the default probability, and how much compensation should be given to banks for mortgage modification.

Previous empirical studies on loan modification programs, such as Quercia and Ding

[29] Software of NPV models are developed by Fannie Mae for HAMP-participating banks to use. The detail of the HAMP NPV model can be found in HAMP on-line documentation. https://www.hmpadmin.com/portal/programs/hamp.jsp

(2009) and Haughwout, Okah, and Tracy (2009), found that in terms of the effect on reducing the overall default rate, writing down the principal is more significant than interest reduction, which is more significant than term extension. They also provided a conceptual discussion that writing down the principal can increase households' ability to pay through payment reduction and raise their incentive to pay through the increase of home equity, while interest reduction and term extension do not have the latter channel.

One contribution of my paper is that the model incorporating partial observability can provide numerical predictions for the change of the illiquidity-triggered default rate and that of the strategic default rate caused by each type of loan modifications. As shown in Table VIII, writing down the principal , interest reduction, and term extension can all effectively reduce the illiquidity-triggered default rate (from 8.37% to 7.58%, 7.54%, and 7.69% respectively). Writing down principal can also effectively reduce the strategic default rate from 2.88% to 1.38%.[30] Interest reduction has almost no effect on the strategic default rate. Term extension even increases the strategic default rate from 2.88% to 3.21%. As the mortgage term is extended from 30 to 40 years, the speed at which the homeowners are building up their home equities is slowed down; therefore, they will have a longer period with low or even negative home equities, which can drive up the strategic default rate.

6.2 Federal Reserve CCAR Scenarios

Starting in 2011, each year the Federal Reserve provided projections of 28 macroeconomic variables over the next 13 quarters under each of the following three 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

[30] In these counterfactual experiments, the loan modification is applied to every household in the sample from the beginning. In practice, the loan modification is only applied to the households who currently have financial hardship, apply for the loan modification program, and get approved by the bank. Therefore, the loan modification program could be more effective in reducing the default rate within this particular mortgage pool which actually get modified.

environment. They are required to submit their models and plannings to the Federal Reserve to pass the annual stress test on their strength and resilience.[31]

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 in 2006 and 2007 the percent changes of the MSA house price index, county unemployment rate, and mortgage interest rate followed the percent changes of the national house price index, national unemployment rate, and mortgage interest rate projected for the fourth quarter in 2013 to the fourth quarter in 2015 by the Federal Reserve for the 2014 annual stress tests. All of the other variables remain the same as the actual history.[32] [33] [34] Figure 5 shows the predicted cumulative illiquidity-triggered default rate, strategic default rate, total default rate, refinance rate, moving rate, and total prepayment rate. Figure 6 shows the movements of house prices, mortgage interest rates, and unemployment rates in different scenarios. For the total default rate, severely adverse scenario is higher than adverse scenario, which is higher than actual history. For

[31] Details about CCAR can be found in Federal Reserve CCAR on-line documentation. http://www.federalreserve.gov/bankinforeg/ccar.htm

[32] The 2014 projection of the macroeconomic variables by Federal Reserve is for the last quarter in 2013 through the last quarter in 2017 rather than for the first quarter in 2006 through the first quarter in 2009. 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 banking organizations and their resilience to adverse economic environments". The counterfactual analyses in this part examine how the default rate and refinance rate of the 2005 cohort would move if Federal Reserve's hypothetical adverse and severely adverse scenarios happened in 2006 and 2007.

[33] In the literature, the counterfactual scenario is usually constructed as only one macroeconomic variable, such as the house price, deviates from the actual value. Then how the variable of interest would respond due to the pure effect of this one-variable deviation is analyzed. However, macroeconomic variables are highly correlated to each other. For example, the drop of the house price is usually accompanied by the increase of unemployment. It is not reasonable to assume that one macroeconomic variable changes while others hold the same. Therefore, I use Federal Reserve CCAR scenarios, which provides projection of 28 macroeconomic variables representing the entire macroeconomic environment, to perform the counterfactual analysis. Moreover, this counterfactual analysis is of practical use for the banking industry.

[34] In the CCP method, the structure of the state transitions is embedded in the first-step policy function estimates. If the structure of the state transitions in the counterfactual scenarios is different from that in the actual history, it will be problematic to apply the first-step policy functions estimated using the actual historical data to the counterfactual scenarios. In the adverse and severely adverse CCAR scenarios, the structure of state transitions of the actual data (not people's expectations) may be different from that of the realized data in the history. However, the structure of people's subjective expectation may be stable across the realized history and those scenarios. I checked Michigan Survey of Consumers and found that the pattern of households' expectations for house price appreciation rates after the financial crisis is not very different from that before the financial crisis.

the illiquidity-triggered default rate and strategic default rate, severely adverse scenario is higher than adverse scenario.[35] The reason why the refinance rate is higher in the severely adverse scenario than in the adverse scenario is that the interest rate is lower in the severely adverse scenario than in the adverse scenario.[36]

The precision of the counterfactual analyses based on the first-stage policy function estimates 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 have 20,487 mortgages with sufficient variation of characteristics, even after loan modification, the characteristics of most of the mortgages should still remain in the empirical support of the policy functions. As the loan performance data spans from 2005 to 2011, which covers the boom, crisis, and recovering of the U.S. economy and housing market, the macroeconomic variables in the CCAR adverse and severely adverse scenarios should remain in the empirical support of the policy functions.

7 Conclusion

When households decide whether to default or refinance on their mortgages, their expectations of future house prices and interest rates matter. This motivates estimating a dynamic discrete choice model. In loan performance data, researchers can only observe default, but cannot identify whether it is illiquidity-triggered or strategic. Moreover, usually researchers can only observe prepayment, but cannot identify whether it is due to refinancing or moving. These facts motivate estimating a dynamic discrete choice model with partial observability.

The paper makes several contributions to both empirical studies and econometric method-

[35] Note that in actual historical data the illiquidity-triggered default rate and strategic default rate are not observable due to the partial observability issue, as well as the refinance rate and moving rate.

[36] The predicted moving rates are the same across adverse and severely adverse scenarios because I assume that in those scenarios the difference between the local county employment growth rate and the national average is the same as the actual history. However, using data sets spanning the entire postwar era, Saks and Wozniak (2011) found that internal migration within U.S. is procyclical. After controlling the relative economic conditions in the origin and destination states or MSAs, the migration rate is significantly higher during booms of the national economy, which suggests that the net benefit of moving rises during booms.

ologies. First, I construct a measure of people's expectations on future house price appreciation rates using the subjective data from the Michigan Survey of Consumers and include it in a logit regression. The result shows that the probability of default will be lower if the households expect increases in house prices. This provides evidence that households are forward looking when making decisions of default on their mortgages, and verifies the motivation to estimate a dynamic choice model. Second, I incorporate partial observability into the conditional choice probability (CCP) method to estimate a dynamic discrete choice model with partially observable outcomes. Exclusion restrictions are used to guarantee the identification of the parameters: some variables with sufficient variation are only related to the incentive to pay but not the ability to pay; and some variables with sufficient variation are only related to the incentive to refinance but not the decision to move. Third, as an alternative to the usual methods that assume consumer rational expectation and estimate a structure of the state transitions using the realized data, following the spirit of Manski (2004), I directly use the subjective expectation data in each month on the future house prices from the Michigan Survey of Consumers to construct the distribution of people's future house price expectations for the state transitions in the estimation of dynamic choice model, which does not require assuming rational expectation for the household and any structure for the state transition process of house prices.

The model 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 the illiquidity-triggered default, but cannot effectively reduce the strategic default. Term extension can effectively reduce the illiquidity-triggered default, but will increase the strategic default as households are going to have a longer period with low or even negative home equities.

Reference

Abraham, Jesse and Patric Hendershott. 1996. "Bubbles in Metropolitan Housing Market." *Journal of housing research,* 7(2), 191-207.

Aguirregabiria, Victor and Pedro Mira. 2010. "Dynamic Discrete Choice Structural Models: A Survey." *Journal of Econometrics*, 156(1), 38-67.

____. 2002. "Swapping the Nested Fixed Point Algorithm: A Class of Estimators for Discrete Markov Decision Models." *Econometrica*, 70(4), 1519-43.

An, Xudong; JohnM Clapp and Yongheng Deng. 2010. "Omitted Mobility Characteristics and Property Market Dynamics: Application to Mortgage Termination." *The Journal of Real Estate Finance and Economics*, 41(3), 245-71.

Archer, Wayne R; David C Ling and Gary A McGill. 1996. "The Effect of Income and Collateral Constraints on Residential Mortgage Terminations." *Regional Science and Urban Economics*, 26(3), 235-61.

Arcidiacono, Peter; Patrick Bayer; Jason R Blevins and Paul B Ellickson. 2012. "Estimation of Dynamic Discrete Choice Models in Continuous Time with an Application to Retail Competition," National Bureau of Economic Research,

Arcidiacono, Peter and Robert A. Miller. 2011. "Conditional Choice Probability Estimation of Dynamic Discrete Choice Models with Unobserved Heterogeneity." *Econometrica*, 79(6), 1823-67.

Bajari, Patrick; C. Lanier Benkard and Jonathan Levin. 2007. "Estimating Dynamic Models of Imperfect Competition." *Econometrica*, 75(5), 1331-70.

Bajari, Patrick; Phoebe Chan; Dirk Krueger and Daniel Miller. 2013a. "A Dynamic Model of Housing Demand: Estimation and Policy Implications*." *International Economic Review*, 54(2), 409-42.

Bajari, Patrick; Chenghuan Sean Chu; Denis Nekipelov and Minjung Park. 2013b. "A Dynamic Model of Subprime Mortgage Default: Estimation and Policy Implications," National Bureau of Economic Research,

Bajari, Patrick; Chenghuan Sean Chu and Minjung Park. 2008. "An Empirical Model of Subprime Mortgage Default from 2000 to 2007," National Bureau of Economic Research,

Calhoun, CharlesA and Yongheng Deng. 2002. "A Dynamic Analysis of Fixed- and Adjustable-Rate Mortgage Terminations." *The Journal of Real Estate Finance and Economics*, 24(1-2), 9-33.

Capozza, Dennis R.; Patric H. Hendershott and Charlotte Mack. 2004. "An Anatomy of Price Dynamics in Illiquid Markets: Analysis and Evidence from Local Housing Markets." *Real Estate Economics*, 32(1), 1-32.

Carranza, Juan Esteban and Salvador Navarro. 2010. "Estimating Dynamic Models with Aggregate Shocks and an Application to Mortgage Default in Colombia." *Borradores de Economía y Finanzas; No. 24-Septiembre 2010*.

Clapp, John M.; Yongheng Deng and Xudong An. 2006. "Unobserved Heterogeneity in Models of Competing Mortgage Termination Risks." *Real Estate Economics*, 34(2), 243-73.

Clapp, John M.; Gerson M. Goldberg; John P. Harding and Michael LaCour-Little. 2001. "Movers and Shuckers: Interdependent Prepayment Decisions." *Real Estate Economics*, 29(3), 411-50.

Conlon, Christopher T. 2012. "A Dynamic Model of Prices and Margins in the Lcd Tv Industry." *Manuscript, Columbia University*.

Corbae, Dean and Ervan Quitin. 2011. "Leverage and the Foreclosure Crisis." Forthcoming at *Journal of Political Economy.*

David M. Blau. 2008. "Retirement and Consumption in a Life Cycle Model." *Journal of Labor Economics*, 26(1), 35-71.

Davis, Morris and Erwan Quintin . 2013. "Self-Assessed House Prices and Mortgage Defaults." *Unpublished manuscript*.

Deng, Yongheng. 1997. "Mortgage Termination: An Empirical Hazard Model with a Stochastic Term Structure." *The Journal of Real Estate Finance and Economics*, 14(3), 309-31.

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

Donna, Javier. 2011. "Estimating Switching Costs in Urban Travel Demand in Chicago." *Northwestern University*.

Engle, Robert F. and C. W. J. Granger. 1987. "Co-Integration and Error Correction: Representation, Estimation, and Testing." *Econometrica*, 55(2), 251-76.

Erdem, Tülin; Michael P Keane; T Sabri Öncü and Judi Strebel. 2005. "Learning About Computers: An Analysis of Information Search and Technology Choice." *Quantitative Marketing and Economics*, 3(3), 207-47.

Ergungor, O Emre and Stephanie Moulton. 2011. "Beyond the Transaction: Depository Institutions and Reduced Mortgage Default for Low-Income Homebuyers." Forthcoming at *Journal of Money, Credit and Banking.*

Ferreira, Fernando; Joseph Gyourko and Joseph Tracy. 2010. "Housing Busts and Household Mobility." *Journal of Urban Economics*, 68(1), 34-45.

Gallin, Joshua. 2006. "The Long-Run Relationship between House Prices and Income: Evidence from Local Housing Markets." *Real Estate Economics*, 34(3), 417-38.

____. 2008. "The Long-Run Relationship between House Prices and Rents." *Real Estate Economics*, 36(4), 635-58.

Gowrisankaran, Gautam and Marc Rysman. 2012. "Dynamics of Consumer Demand for New Durable Goods." *Journal of Political Economy*, 120(6), 1173-219.

Haughwout, Andrew; Ebiere Okah and Joseph Tracy. 2009. "Second Chances: Subprime Mortgage Modification and Re-Default," Staff Report, Federal Reserve Bank of New York,

Hendel, Igal and Aviv Nevo. 2006. "Measuring the Implications of Sales and Consumer Inventory Behavior." *Econometrica*, 74(6), 1637-73.

Ho, Chun-Yu. 2008. "Switching Cost and the Deposit Demand in China." *Unpublished Manuscript, Georgia Institute of Technology*.

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

Hotz, V. Joseph; Robert A. Miller; Seth Sanders and Jeffrey Smith. 1994. "A Simulation Estimator for Dynamic Models of Discrete Choice." *The Review of Economic Studies*, 61(2), 265-89.

Kau, JamesB; DonaldC Keenan and Xiaowei Li. 2011. "An Analysis of Mortgage Termination Risks: A Shared Frailty Approach with Msa-Level Random Effects." *The Journal of Real Estate Finance and Economics*, 42(1), 51-67.

Kau, JamesB; DonaldC Keenan; WalterJ Muller, III and JamesF Epperson. 1995. "The Valuation at Origination of Fixed-Rate Mortgages with Default and Prepayment." *The Journal of Real Estate Finance and Economics*, 11(1), 5-36.

Laufer, Steven. 2011. "Equity Extraction and Mortgage Default," Working Paper. New York University.

Lee, Robin S. 2013. "Vertical Integration and Exclusivity in Platform and Two-Sided Markets." *American Economic Review*, 103(7), 2960-3000.

Malpezzi, Stephen. 1999. "A Simple Error Correction Model of House Prices." *Journal of Housing Economics*, 8(1), 27-62.

Manski, Charles F. 2004. "Measuring Expectations." *Econometrica*, 72(5), 1329-76.

Meng, Chun-Lo and Peter Schmidt. 1985. "On the Cost of Partial Observability in the Bivariate Probit Model." *International Economic Review*, 26(1), 71-85.

Murphy, Alvin. 2013. "A Dynamic Model of Housing Supply." *Unpublished manuscript available upon request from author*.

Murphy, Kevin M. and Robert H. Topel. 1985. "Estimation and Inference in Two-Step Econometric Models." *Journal of Business & Economic Statistics*, 3(4), 370-79.

Pavlov, AndreyD. 2001. "Competing Risks of Mortgage Termination: Who Refinances, Who Moves, and Who Defaults?" *The Journal of Real Estate Finance and Economics*, 23(2), 185-211.

Poirier, Dale J. 1980. "Partial Observability in Bivariate Probit Models." *Journal of Econometrics*, 12(2), 209-17.

Quercia, Roberto G. and Lei Ding. 2009. "Loan Modifications and Redefault Risk: An Examination of Short-Term Impacts." *Cityscape*, 11(3), 171-93.

Quigley, John M. 1987. "Interest Rate Variations, Mortgage Prepayments and Household Mobility." *The Review of Economics and Statistics*, 69(4), 636-43.

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

Saks, Raven E. and Abigail Wozniak. 2011. "Labor Reallocation over the Business Cycle: New Evidence from Internal Migration." *Journal of Labor Economics*, 29(4), 697-739.

Schiraldi, Pasquale. 2011. "Automobile Replacement: A Dynamic Structural Approach." *The RAND Journal of Economics*, 42(2), 266-91.

Shcherbakov, Oleksandr. 2009. "The Effect of Consumer Switching Costs on Market Power of Cable Television Providers." *Unpublished manuscript*.

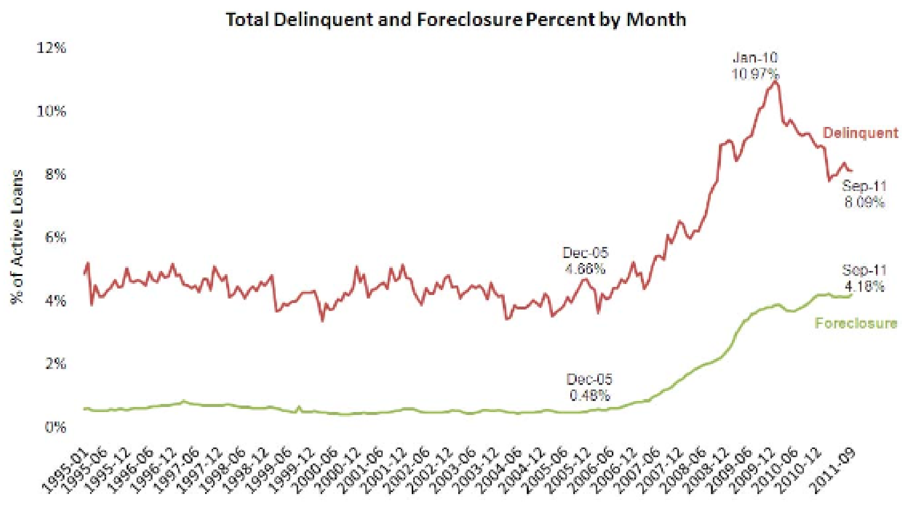
Song, Inseong and PradeepK Chintagunta. 2003. "A Micromodel of New Product Adoption with Heterogeneous and Forward-Looking Consumers: Application to the Digital Camera Category." *Quantitative Marketing and Economics*, 1(4), 371-407.

Stanton, Richard and Nancy Wallace. 1998. "Mortgage Choice: What's the Point?" *Real Estate Economics*, 26(2), 173-205.

Wang, Emily Yucai. 2013. "The Impact of Soda Taxes on Consumer Welfare: Implications of Storability and Taste Heterogeneity," Mimeo,

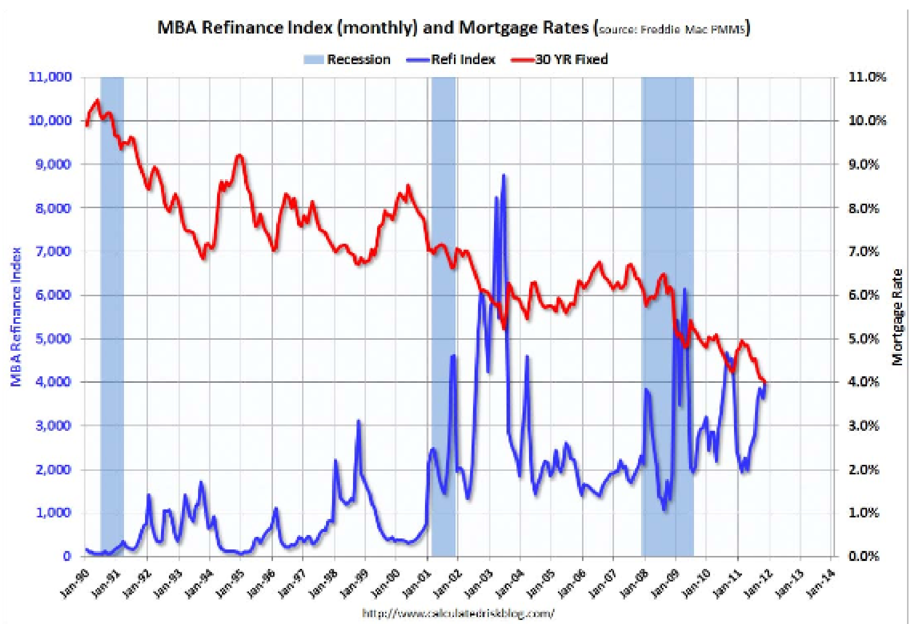
Yuanjie, Zhang. 2010. "High Default Risk on Down Payment Assistance Program: Adverse Selection v.s. Program Characteristics?" *Unpublished manuscript*.

Figure 1: National Foreclosure Rate



(Source: http://www.calculatedriskblog.com)

Figure 2: National Refinance Rate



(Source: http://www.calculatedriskblog.com)

Figure 3: The Model

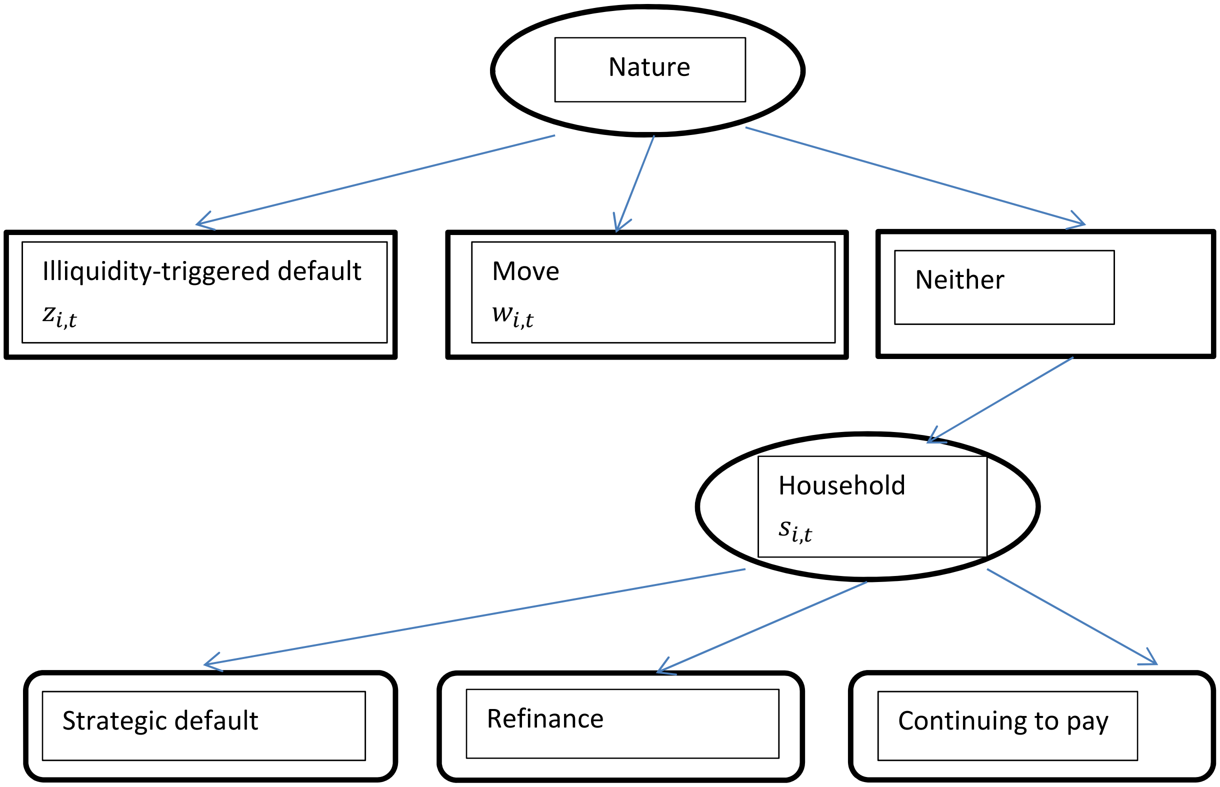
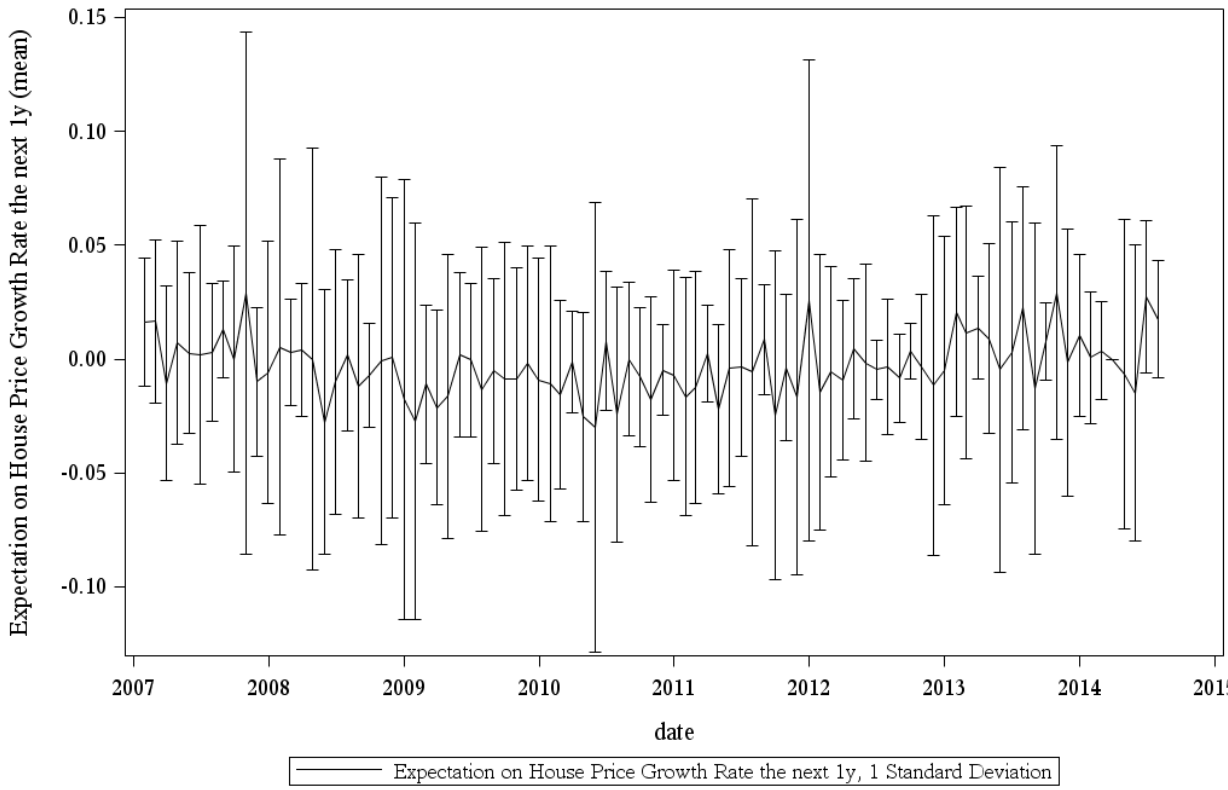


Figure 4: Expectation on Future House Price Growth



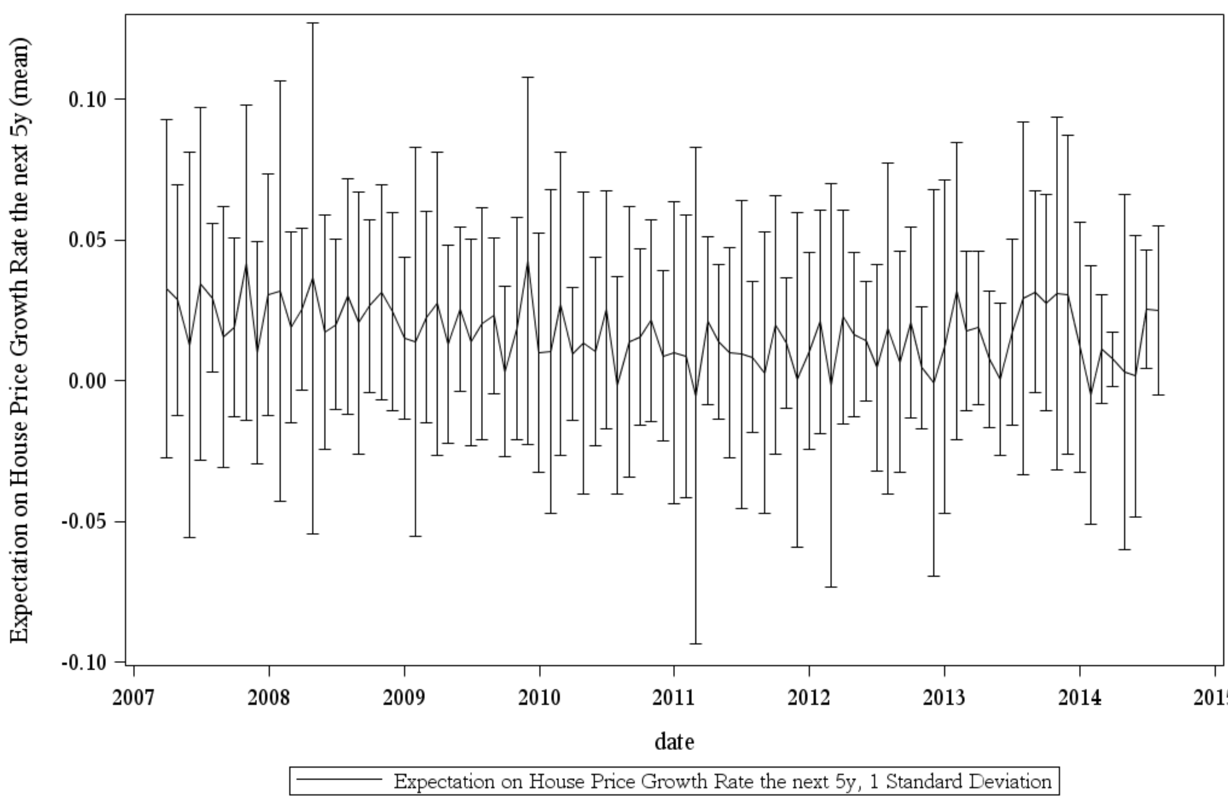


Fig 5: Counterfactural Analyses for CCAR scenarios

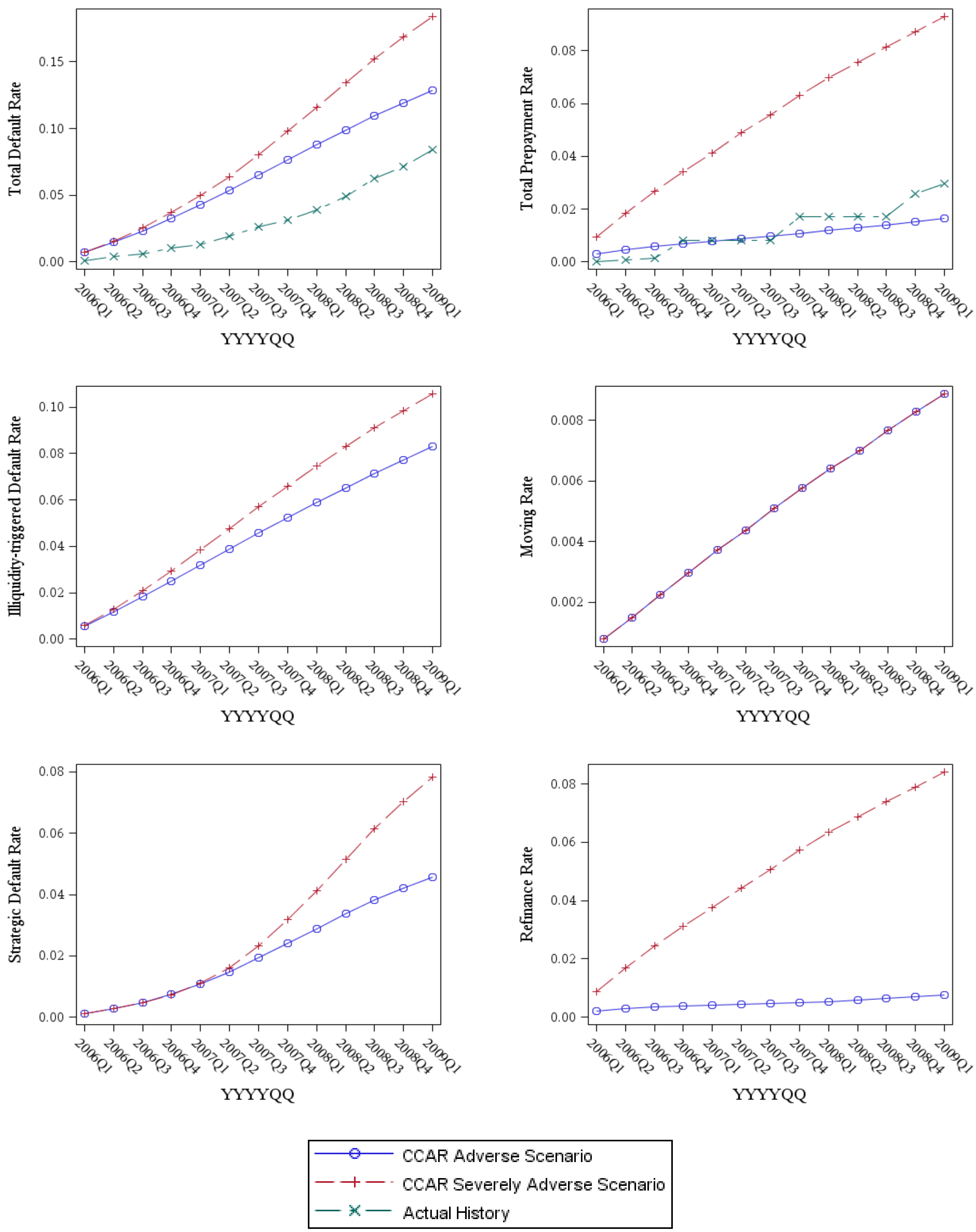
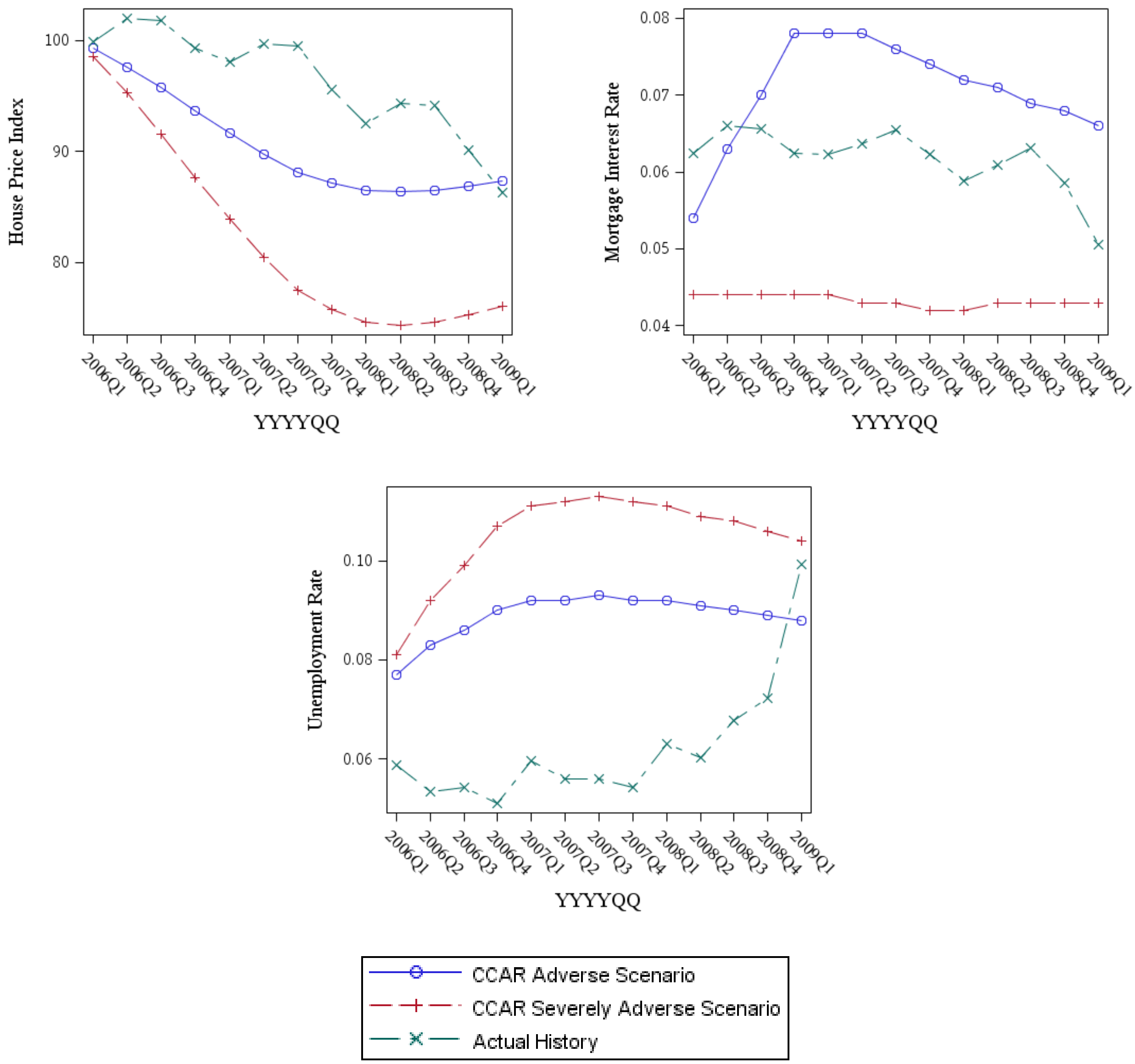


Fig 6: Macro Variables in CCAR scenarios



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

Notes. $defaultOpt_{i,t}$ and $refiOpt_{i,t}$ are panel data, thus has 896,043 observations. Other variables are cross-sectional data.

TABLE III: Logit Regression

	Model (I)		Model (II)	
	Default	Prepay	Default	Prepay
Intercept	0.1304	-9.214***	0.285	-9.0358***
	(0.2767)	(0.3089)	(0.2825)	(0.314)
Payment	0.000953***	-0.00016	0.000888***	-0.00032**
	(0.000146)	(0.000144)	(0.00015)	(0.000151)
Income	-0.00031***	0.000069***	-0.0003***	0.000062**
	(0.000027)	(0.000026)	(0.000028)	(0.000026)
Household size	0.0967***	-0.0931***	0.0996***	-0.0931***
	(0.0157)	(0.0216)	(0.0159)	(0.0218)
Credit Score	-0.0104***	0.00379***	-0.0103***	0.00378***
	(0.000346)	(0.000368)	(0.000351)	(0.000371)
Minority	0.042	-0.437***	0.0544	-0.4316***
	(0.0599)	(0.097)	(0.0606)	(0.0975)
County unemployment rate	12.0249***	6.52***	10.2067***	4.9439***
	(1.0123)	(1.1063)	(1.0585)	(1.1586)
Employment growth rate diff	-0.2295	-7.646***	-1.0441	-7.2935***
(local - national)	(2.3018)	(2.1259)	(2.3217)	(2.1256)
Female	0.00753	-0.2025***	0.0149	-0.2018***
	(0.0439)	(0.0454)	(0.0445)	(0.0457)
DPA grant	0.0554	-0.0751	0.0959**	-0.0699
	(0.046)	(0.0503)	(0.0469)	(0.051)
Age	0.000448	-0.0145***	0.000168	-0.0145***
	(0.002150	(0.00242)	(0.00218)	(0.00245)
$defaultOpt_{i,t}$ (\$1000)	0.0182***	-0.0221***	0.0163***	-0.0212***
	(0.00275)	(0.00141)	(0.0027)	(0.00143)
$refiOpt_{i,t}$ (\$1000)	-0.00347	0.0927***	-0.0107***	0.0925***
	(0.00267)	(0.00273)	(0.00282)	(0.00286)
$Homeprice_{i,t}$ × 1 year			-0.00005***	-0.00009***
expectation growth rate			(0.000016)	(0.000016)
$Homeprice_{i,t}$ × 5 year				
expectation growth rate				

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

Table III continued

	Model (III)		Model (IV)	
	Default	Prepay	Default	Prepay
Intercept	0.2516	-9.1077***	0.2524	-9.0119***
	(0.2836)	(0.3131)	(0.2843)	(0.3141)
Payment	0.00139***	0.000169	0.00139***	-0.00011
	(0.000166)	(0.00016)	(0.000177)	(0.000174)
Income	-0.0003***	0.000062**	-0.0003***	0.000062**
	(0.000028)	(0.000026)	(0.000028)	(0.000026)
Household size	0.099***	-0.0944***	0.0991***	-0.093***
	(0.016)	(0.0218)	(0.016)	(0.0218)
Credit Score	-0.0103***	0.00377***	-0.0103***	0.00378***
	(0.000353)	(0.000371)	(0.000353)	(0.000371)
Minority	0.0617	-0.4282***	0.0617	-0.4318***
	(0.0608)	(0.0975)	(0.0608)	(0.0975)
County unemployment rate	10.2939***	5.6928***	10.2852***	4.8198***
	(1.0448)	(1.1363)	(1.0651)	(1.1606)
Employment growth rate	-0.1908	-6.8308***	-0.1943	-7.0876***
diff (local - national)	(2.3585)	(2.1723)	(2.3601)	(2.1491)
Female	0.0169	-0.2008***	0.0169	-0.2021***
	(0.0448)	(0.0457)	(0.0448)	(0.0457)
DPA grant	0.1048**	-0.0539	0.1049**	-0.0579
	(0.0472)	(0.051)	(0.0472)	(0.0511)
Age	0.000329	-0.0146***	0.000328	-0.0147***
	(0.00219)	(0.00245)	(0.00219)	(0.00245)
$defaultOpt_{i,t}$ (\$1000)	0.0135***	-0.0235***	0.0135***	-0.0222***
	(0.00274)	(0.00145)	(0.00276)	(0.00149)
$refiOpt_{i,t}$ (\$1000)	-0.0159***	0.0867***	-0.0159***	0.0893***
	(0.00295)	(0.00296)	(0.00295)	(0.00304)
$Homeprice_{i,t}$ × 1 year			-7.38E-07	-0.00007***
expectation growth rate			(0.000018)	(0.000017)
$Homeprice_{i,t}$ × 5 year	-0.00012***	-0.00007***	-0.00012***	-0.00004*
expectation growth rate	(0.000022)	(0.00002)	(0.000025)	(0.000021)

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

TABLE IV: Policy Functions in the 1st Stage Estimation

	Illiquidity-triggered default	Move	Strategic Default	Refinance
Intercept	-0.4786	-6.0533***	0.4232	-10.3161***
	(0.5620)	(0.4333)	(1.3449)	(0.3163)
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***		-10.5136***	5.0952***
	(0.7893)		(2.0153)	(0.4351)
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)		
$defaultOpt_{i,t}$ ($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 and their interactions included			Yes	Yes

Notes. *Denotes significance at 10% level. **Denotes significance at 5% level. ***Denotes significance at 1% level. The standard errors are in the 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)

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

TABLE VI: The 2nd Stage Estimation using Estimated State Transitions

	Strategic default	Refinance	Continue to pay
Intercept	- 89.1999	- 8.1326***	
	(90.9935)	(0.1092)	
Payment ($1000)		- 20.4213***	-20.4213***
		(0.4677)	(0.4677)
Credit score/1000		4.8499***	-0.4844
		(1.7629)	(1.7181)
House purchase price		24.8045***	24.8045***
(deflated by CPI)		(0.8972)	(0.8972)

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

TABLE VII: The 2nd Stage Estimation using Subjective Expectation

	Strategic default	Refinance	Continue to pay
Intercept	- 70.5634	- 8.0481***	
	(86.6624)	(0.1093)	
Payment ($1000)		- 19.7154***	-19.7154***
		(0.4740)	(0.4740)
Credit score/1000		4.7422 ***	-0.5687
		(1.6818)	(1.6349)
House purchase price		23.9765***	23.9765***
(deflated by CPI)		(0.8804)	(0.8804)

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

TABLE VIII: Counterfactual Analyses: Loan Modification

Counterfactual Scenarios	Illiquidity-triggered default rate	Strategic default rate	Move rate	Refinance rate
Baseline	8.37%	2.88%	1.05%	9.86%
Principal Writing down 10%	7.58%	1.38%	1.05%	11.54%
Principal Writing down 20%	6.86%	0.70%	1.05%	13.67%
Interest reduction 1%	7.54%	2.92%	1.05%	3.09%
Term extension to 40 year	7.69%	3.21%	1.05%	14.37%