Unobserved Heterogeneity and Its Effects on Mortgage Options*

Min Hwang
George Washington University

Raphael Kuznetsovski
George Washington University

May 30, 2011

** Very preliminary. Please do not quote without authors' explicit consent.*

Abstract

This paper extends the literature on modeling unobserved heterogeneity among mortgage holders and investigates reasons for non-optimal exercising of mortgage termination options: prepayment and default. Using a large sample of diversified population of prime and Alt-A loans, the research shows that unobserved heterogeneity can be substantially reduced after incorporating expanded list of covariates, including time-varying measures of borrower's creditworthiness. The research further shows that a more accurate classification of borrowers into various risk groups can be obtained by increasing number of mass points used to approximate the unknown distribution of unobserved heterogeneity parameters. It also highlights challenges one might face when trying to impose simplifying parametric assumptions on the joint distribution of unobserved heterogeneity. Finally, we present more appropriate capital allocation when defaults are estimated under the optimal number of mass points.

1. Introduction and Literature Review

The literature on the theoretical aspects of residential mortgage valuation is vast and still expanding. The notable contributions were made by works of Dunn and McConnell (1981), Cunningham and Hendershott (1984), Epperson et al. (1985), and Kau et al. (1992, 1993, 1994, 1995), Capozza, Kazarian, and Thomson (1998), and Ambrose and Buttimer (2000). Early theoretical models in this literature focused on application of rational option pricing methods to the problem of mortgage pricing. In option pricing models of consumer behavior, the borrower weighs the value of the future payments discounted with the current mortgage rate against the current mortgage balance, and the value of the collateral against the outstanding principal balance of the mortgage, which guides him or her to make a choice between two early mortgage termination options: default or prepayment.

The theory assumes that mortgage holders are rational agents and that they will try to minimize the market value of their mortgage liabilities through exercising of mortgage termination options. Therefore, modern pricing models of both standalone mortgages and Mortgage-Backed Securities (MBS) must explicitly recognize both termination options embedded in mortgage contracts -- the right to default and the right to prepay. In fact, as argued by Kau et al. (1992), "… since prepayment and default substitute for one another, contracts with only one of the default or prepayment provisions lead the borrower to behave differently then when both are present. This substitution effect means that one cannot accurately value either the individual provisions or their interaction without both options being present." Kau et al. (1992, 1994) then showed how application of option pricing techniques under a contingent claim framework can lead to theoretical value of the mortgage by solving a multi-dimensional partial differential equation (PDE) using backward induction method under appropriate terminal and boundary conditions.

At the same time, there have been numerous empirical studies that tried to exploit advances of option pricing models, aiming to show the link between theoretical mortgage models on default and prepayment and the empirical models driven by the data. Among them were Schwartz and Torous (1989) and Ambrose and Capone (1998), who studied Federal-Housing Authority (FHA)-insured loans, Phillips and VanderHoff (2004) who modeled conventional loans, VanderHoff (1996) who analyzed historical data on 15-year fixed-rate mortgages (FRMs) and 1-year adjustable-rate mortgages (ARMs), and Capozza and Thompson (2005) who studied subprime mortgages. The most notable finding from numerous studies is that, despite options-based theory, borrower's exercise of options often fails – completely or partially – to follow theoretical predictions (Archer and Ling (1993), Quigley and Van Order (1990, 1995)). In one attempt, focusing primarily on prepayments, Archer and Ling (1993) classified models in two categories: those motivated by borrower financial options and those using an empirical approach to study exogenous terminations. The progress in this area of research to try to reconcile theory with observed empirical data was likely impeded by the lack of high quality borrower-level data available for academic research. Most of borrower-level origination and servicing mortgage data is proprietary and is owned by commercial entities.

There were several attempts made by different authors to deal with apparent disagreement between option-based theory on mortgage terminations and observed empirical data. In the literature, this is often referred to as unobserved heterogeneity of the borrowers that cannot be fully explained by available data. One way to explain unobserved heterogeneity is to consider transaction costs related to mortgage termination options. Stanton (1995) proposed an extension of the structural option-based prepayment model by incorporating three exogenous components with which he tried to explain non-rational option exercising by the borrowers. According to Stanton, the borrower's decision to prepay the mortgage is not only affected by heterogeneous transaction costs, which determine the optimal prepayment strategy, but also by time intervals at which these prepayment decisions are evaluated. Stanton suggested that the prepayment decisions are made at random discrete intervals, rather than continuously, which had been implicitly assumed by previous authors. The third component of the Stanton's model is a random process of forced prepayments for reasons other than refinancing. However, Stanton only considered prepayment assuming away borrowers' default decisions. Since then, there have been no published studies that extend Stanton's prepayment model to a more general model that combines both optimal prepayment and optimal default decisions under a unified framework.

In 2000, Deng, Quigley, and Van Order considered the issue of unobserved heterogeneity in the context of hazard modeling. They presented a unified model of competing risks of mortgage terminations, considering these two hazards – prepayment and default – as dependent competing risks that are estimated jointly. A noble contribution of their approach is to explicitly consider unobserved heterogeneity by adding discretely distributed mass point mixed hazard. This approach has been followed by Clapp, Deng, and An (2006), who considered the multinomial logit (MNL) model with mass point mixed hazard and compared with the model by Deng, Quigley and Van Order (2000) in the general framework of mortgage terminations.

2. Motivation and Preview

The methods applied by Stanton (1995), Deng et al. (2000), and Clapp et al. (2006), were developed to fill the gap between the current state of the options-based mortgage pricing theory and the wide range of empirical results that often disagree with theoretical predictions. However, these methods have not yet provided convincing evidence to sway research community to support one side or the other. For example, Stanton's explicitly modeled and estimated heterogeneity in the transactions costs faced by mortgage holders produced the distribution of random transaction costs with the mean value of 41% of the loan balance. This estimate is between five to seven times higher than the average explicit monetary costs normally associated with the refinancing process. This example of the big gap between the model output and widely held beliefs of what the actual transaction costs should be is one of the reasons why this area of research was, and still remains, very active.

Two possible hypotheses have been mentioned with regularity that might help to address some of the disagreements between the data and the simple option pricing models. The first one is related to uncertainty and possible large measurement error in the way actual values of the underlying properties are estimated. Indeed, short of requesting a frequently updated detailed appraisal report for every property of interest, it is almost impossible to be certain about the actual value of a given property at any given point in time. Therefore, as Deng et al. (2000) argued, "we cannot even measure directly the extent to which the default option is in the money without the data on the course of individual house prices."

The second hypothesis is related to borrowers' inability to qualify for new loans due to their deteriorating credit rating and/or temporary financial difficulties. (Stanton (1995), Clapp et al. (2006).) Despite recent progress documented in Clapp et al. (2006), neither of the two hypotheses has been thoroughly investigated using a large representative data sample.

The motivation for our research is to provide first answers to some aspects of the widely hypothesized, but not yet tested, explanations regarding origins of seemingly irrational borrowers' behavior of mortgage options exercising. While recognizing that ordering quarterly appraisals for each property is prohibitively expensive and impractical, we focus our attention on trying to address the second hypothesis regarding potential impact of borrower's changing creditworthiness on non-optimal mortgage options exercising.

As mentioned earlier, the slow advancement of our understanding of borrower unobserved heterogeneity was partially related to unavailability of the relevant low-level data, without which, it is very difficult to differentiate between different classes of borrowers. For this analysis, we had access to a much richer mortgage origination and performance data that includes additional risk characteristics such as loan discount points, prepaid finance charges, updated FICO scores, and bankruptcy indicators – these were usually missing from previous studies. With this research, we try to quantitatively assess the importance of these previously omitted risk drivers.

The point of departure for this analysis is the joint competing risks hazard framework, which was also considered by Deng et al. (2000) and by Clapp et al. (2006). With our research, we attempt to make several novel contributions. First, we extend Deng et al. (2000) and Clapp et al. (2006) results by studying how the relationship between unobserved heterogeneity parameters changes while using an expanded list of borrower- and mortgage-related covariates. For the first time, by incorporating time-varying credit risk indicators, such as periodically updated FICO scores and bankruptcy filings, into the jointly estimated competing risks mortgage termination models, we show that deterioration of borrowers' credit rating definitely plays a significant role in their inability to take advantage of attractive refinance opportunities. In addition, we indirectly measure the improvement the new information brings to the model by analyzing the changes in estimated unobserved heterogeneity parameters. My results show that unobserved heterogeneity is reduced by 33% after including both additional asymmetric information collected during mortgage underwriting process and time-varying borrower-level information that is revealed after mortgage origination.

The second contribution of our research is in advancing our understanding about distributional assumptions on unobservable heterogeneity parameters when modeling a diverse population of mortgage holders. None of the previous research papers that used competing risks hazard framework have addressed this issue in great detail. Previously, the maximum number of mass points that was used to classify borrowers into homogenous groups was three (Deng et al. (2000)). We expand the population partition to five mass points and observe the evolution of the changing distribution shape, for both prepayment and default heterogeneity parameters.

3. Data

The empirical data that was used to analyze the hypotheses came from a major U.S. commercial bank with a national presence. The data consisted of first-lien residential mortgages originated and serviced by the mortgage division of the bank. All of the mortgages in the sample were conventional (non-jumbo) owner-occupied fully amortized fixed-rate mortgages with a 30-year repayment term.

3.1. Data Coverage

The analysis sample covers new mortgage originations starting from January of 1999 through June of 2008. Loan performance was tracked at quarterly intervals, through March of 2009, thus providing an observation window of more than 10 years. This observation window includes a variety of economic conditions and several business cycles, starting from the dot-com boom and bust, through the mild recession of 2001, through the housing bubble, and, finally, catching a big portion of the current housing crash and the world-wide financial crisis. We should note that most of the mortgages in the sample were originated with the purpose of refinancing of a previously originated mortgage (which was not necessarily serviced by the same lender). In many instances, the borrowed amount exceeded the payoff amount of the previously held mortgage, thus resulting in equity cash-out.

3.2. Data Profile

For each loan, the data provides very detailed information about origination terms of the mortgage (including note rate, amount of points and/or prepaid finance charges, total origination fees, etc.); as well as key borrower-related characteristics, such as borrower's age, family and employment status; financial disclosures, including income and savings; and an indicator of borrower's credit risk, summarized as FICO score.

Most borrowers in the sample had relatively high credit scores -- the average FICO score of all originated mortgages was 728. Nevertheless, relaxed underwriting standards on original loan-to-value ratio, or OLTV, (18% of the loans had OLTV higher than 90%) and on debt-to-income ratio, or DTI, (28% of the loans had DTI higher than 45%), as well as reduced loan documentation requirements, make the analysis sample an interesting blend of Prime and Alt-A quality paper. Table 1 and Table 2 below show summary statistics for key data elements by origination year and overall for the population.

The originated loans were geographically diversified. Properties from 49 out of 51 states plus the District of Columbia are represented in the data. None of the state shares exceed 11% of total loan counts. (See Appendix 1 for more details.) In addition, in Appendix 2 at the end, we provide plots of empirically estimated quarterly hazard rates for both prepayment and default terminations, by key origination variables (FICO, OLTV, and DTI).

Figure 1 presents visual distribution of note rates for all loans in the population, by origination date. Each loan is represented by either a red dot or an orange dot on the chart, depending on the origination FICO score. Orange dots represent mortgages with FICO>=720, while red dots are for mortgages with FICO<720. In addition, the market's monthly average commitment rate on 30-year fixed rate mortgages, which is compiled by Freddie Mac, is superimposed on the same chart with a 2-month lag relative to the mortgage origination month. The prevailing market rate is shown as a solid black line on the chart.

From the chart we can see that the note rates of originated loans in the sample are in line with the market rate. Observed positive spread during some of the periods is to be expected because of the special mixture of Prime and Alt-A type mortgages in the population. Most of outliers corresponding to extremely high note rates were originations for borrowers with less than stellar credit history. Therefore, based on note rate patterns exhibited by loans in Figure 1, it is reasonable to assume that most of the mortgages included in the study were fairly priced.

Starting from the time of origination, the servicer of the mortgage portfolio also observed and kept track of important variables that change over time, such as unpaid principal balance and delinquency status, which served as the key identifiers for the termination events. In addition, updated FICO scores were obtained each quarter from one of the major credit bureaus for each borrower with an outstanding principal balance. The servicer of the mortgages also had information about whether or not the borrower had filed for or was in bankruptcy. The updated FICO scores and bankruptcy indicators provide us with a glimpse of the financial strength (or lack of thereof) of the borrower at different points in time.

All of these static and time-varying loan-level data were combined with key economic indicators, such as prevailing market mortgage interest rates, state-level house price index (HPI), and state-level unemployment rates, in one comprehensive data set.

3.3. Mortgage Delinquency and Default Identification

A mortgage is legally declared in default when borrowers fail to abide by any of the provisions of the mortgage contract. The most common form of default arises from the borrower's nonpayment. Borrowers who miss regular mortgage payments move through a classification system of delinquency, which is normally based on the number of missed payments (30, 60, or 90 days delinquent). As the number of missed payments increases, the seriousness of the delinquency increases. In this study, borrowers are classified as "in default" after becoming 90 days delinquent (that is, when the fourth payment is due).

Based on historic delinquency information for each loan, a default indicator variable δ_{jt}^{d} was constructed as follows: $\delta_{jt}^{d}=1$ if the j^{th} mortgage for the first time reached 90 days delinquency status in month t, and $\delta_{jt}^{d}=0$ otherwise. In total, there were 1229 instances of default identified over the observation period. This comprises 3.3% of all loans in the sample.

3.4. Prepayment Identification

All of the mortgages in the data were fully-amortized 30-year fixed rate mortgages. However, majority of the borrowers don't wait until their mortgage fully amortizes. Instead mortgage holders have the right to voluntarily prepay the remaining principal balance of the mortgage at any time ("call option"). In the data, prepayments are identified by a special flag. Using this prepayment flag indicator, a prepayment variable δ_{jt}^{p} was constructed as follows: $\delta_{jt}^{p}=1$ if the prepayment indicator for j^{th} mortgage is set to on and if the mortgage has not been previously identified as default; otherwise δ_{jt}^{p} is set to 0. Of the 37,342 mortgages included in the study and originated between 1999 and 2008, 14,953 (or 40%) have prepaid by the end of observation period in March of 2009.

4. Methodology

4.1. Competing Risks Hazards Model with Unobserved Heterogeneity

Hazard risk modeling began to gain popularity following the seminal work of Cox (1972) in which he introduced a convenient method to efficiently estimate proportional shifts in flexible hazard functions due to effects from observed covariates. Cox's methodology relied on partial maximum likelihood estimation of the unknown parameters and was a semi-parametric in nature, because it did not require a specific functional form for the underlying baseline function. Cox's work spurred a huge interest in survival and hazard modeling, which soon resulted in substantial extensions to the original Cox's modeling methodology. In this analysis, we will be particularly interested in the following two important extensions of the classical survival theory: decomposition of the terminating events into multiple (and distinct) outcomes as the result of competing hazard risks, and removal of proportionality assumptions to allow the hazard function to depend on time-varying covariates.

Han and Hausman (1990) applied proportional hazards modeling with competing risks in the presence of unobserved heterogeneity and tested it using unemployment duration data. Sueyoshi (1992) extended Han and Hausman's (1990) results by relaxing proportionality restrictions and allowing covariates to change over time. McCall (1996) further advanced the competing risks hazards framework and performed an empirical study of the unemployment terminating events. McCall (1996) also provided mathematical expressions for the joint survivor function and the maximum likelihood function, conditional on both observed and unobserved parameters, for grouped duration data.

Deng, Quigley, and Van Order (2000) were the first to apply the methodology of competing risks hazards with unobserved heterogeneity to estimate a reduced-form early termination model for mortgages for both prepayment and default. Deng et al. (2000) showed that the options theory is still relevant when it comes to explaining the early termination patterns of large population of conventional fixed rate mortgages. However, they also showed, through simultaneous estimation of discrete heterogeneity distribution, that while some borrowers have propensity to respond quickly to changing economic conditions by exercising their options, others, for whatever reasons, often miss out on presented to them opportunities to either "enrich" themselves by refinancing out of the mortgage, or to minimize their losses by defaulting on the mortgage. Deng et al. (2000) also provided expressions for the joint survivor function and the maximum likelihood function for combined competing risks of default and prepayment. In this paper we will generally follow the notations and functional forms established by McCall (1996) and Deng et al. (2000).

4.2. Model Specification

As discussed earlier, a mortgage loan contract can prematurely terminate either because of default, when the borrower stops making regular mortgage payments (and surrenders the collateral), or because of voluntary prepayment, when the borrower "calls" the mortgage early and repays the outstanding principal balance at once. Let T_p be a random duration variable describing time elapsed from origination until voluntarily prepayment, and let T_d be a random variable describing time until default. Then, following Deng et al. (2000), the joint survivor function conditional on heterogeneity variables for prepayment, η_p, and default, η_d, can be expressed as follows:

$$S(t_p, t_d | X, \eta_p, \eta_d) = exp\left(-\eta_p \sum_{k=1}^{t_p} exp(\gamma_{pk} + \beta_p' X) \ - \eta_d \sum_{k=1}^{t_d} exp(\gamma_{dk} + \beta_d' X)\right),$$

where γ_p and γ_d are log-transformations of the piece-wise constant baseline hazards for prepayment and default, X is a vector of (possibly time-varying) observed covariates, and β_p and β_d are the vectors of unknown parameters. It is common to assume that unobserved $\eta_p$and η_d are independent of the observed explanatory variables X. In most general case, the unobserved random vector (η_p, η_d) could belong to a broad family of distributions, both continuous and discrete. Let G be the distribution of the unobservables η_p and η_d. Then the unconditional survivor function can be presented as

$$S(t_p, t_d | X) = \int exp\left(-\eta_p \sum_{k=1}^{t_p} exp(\gamma_{pk} + \beta_p' X) \ - \eta_d \sum_{k=1}^{t_d} exp(\gamma_{dk} + \beta_d' X)\right) dG(\eta_p, \eta_d)\,.$$

As was shown in McCall & McCall (2008), the log likelihood function of the hazard model with two competing risks in the presence of unobserved heterogeneity can be expressed as follows:

$$\begin{aligned}\ln(L) = \sum_{\mathrm{j}=1}^{\mathrm{N}} [\delta_j^p \ln\left(\int \{S(t-1, t-1 | X_j, \eta_p, \eta_d) - S(t-1, t | X_j, \eta_p, \eta_d) - A(t | X_j, \eta_p, \eta_d)\} dG(\eta_p, \eta_d)\right) \\ + \delta_j^d \ln\left(\int \{S(t-1, t-1 | X_j, \eta_p, \eta_d) - S(t, t-1 | X_j, \eta_p, \eta_d) - A(t | X_j, \eta_p, \eta_d)\} dG(\eta_p, \eta_d)\right) \\ + \delta_j^c \ln\left(\int \{S(t-1, t-1 | X_j, \eta_p, \eta_d)\} dG(\eta_p, \eta_d)\right)]\,,\end{aligned}$$

where δ_j^p, δ_j^d, and δ_j^c are indicator variables for prepaid, defaulted, and censored observations, respectively, and $A(t | X, \eta_p, \eta_d)$ is given by

$$
\begin{aligned}
A(t|X,\eta_p,\eta_d) \quad &= 0.5\big[S(t-1,t-1|X,\eta_p,\eta_d) + S(t,t|X,\eta_p,\eta_d) \\
&\quad -S(t-1,t|X,\eta_p,\eta_d) - S(t,t-1|X,\eta_p,\eta_d)\big]\,.
\end{aligned}
$$

To the best of our knowledge, up until now there have not been any complete empirical studies done to identify possible parametric families of distributions the vector (η_p, η_d) is likely to belong to, when dealing with mortgage terminations under competing risks hazard framework. Without simple parameterization, the unconditional survivor function $S(t_p, t_d|X)$ and the corresponding log likelihood function $\ln(L)$ become computationally untractable for practical purposes.

Instead, what has been done in practice (McCall (1996), Deng et al. (2000)) is to assume discrete mass-point distribution for (η_p, η_d). In other words, mass-point distribution assumes that the population of all borrowers, Ω, is comprised of a finite number of distinct subpopulations, Ω_m, each occurring with probability p_m, where m varies from 1 to M, and $\Omega = \bigcup_{m=1}^{M} \Omega_m$ and $\sum_{m=1}^{M} p_m = 1$. By increasing number of mass points, M, one might expect that the properly estimated mass-point distribution should converge to the true distribution of the unobserved parameters (η_p, η_d). In our analysis, we also follow the same approach and assume discrete mass-point distribution for the unobserved heterogeneity vector.

Deng et al. (2000) estimated several models of the "interdependent competing risks of mortgage prepayment and default with unobserved heterogeneity" using a sample of 22,294 conventional fixed-rate mortgages originated between 1976 and 1983. All models were fitted using maximum likelihood optimization approach by jointly estimating parameters for baseline hazard function, covariates, and unobserved heterogeneities. Deng at al. (2000) only considered cases where the true distribution of unobserved heterogeneity parameters was approximated using discrete mass-point distributions with less or equal than three mass points. Their main results are summarized in Table IV on p.291 of their paper. The first -- "ruthless" -- model, while unquestionably providing support for the option theory, highlighted a big difference in exercising of prepayment options between borrowers in "high risk" and "low risk" groups. Borrowers in the high risk group were 4.73 times more likely to prepay their mortgages, holding everything else equal. Their second model, after adding several covariates which represented asymmetric information and trigger events, provided slightly better overall fit to the data. Still, the estimates of the heterogeneity parameters were similar to those estimated in the first model: borrowers in the high risk group were about 4.58 times more likely to exercise their prepayment options than borrowers in the low risk group.

When explaining their results, Deng et al. (2000) suggested that the differences between high risk and low risk groups could be attributed to several factors: "… unmeasured house-specific factors (such as unexpected depreciation or appreciation in property values) as well as to borrower tastes and abilities", p.290. They did not, however, empirically investigate these claims.

The goal of our research, presented below, is threefold. First, we replicate Deng et al. (2000) original results using a large sample of recently originated mortgages that were somewhat different from the sample used by Deng et al. (2000). Second, we extend the Deng at al. (2000) analysis by studying the relationship between unobserved heterogeneity parameters and an expanded list of both static and time-varying covariates that help explain some of the unobserved heterogeneity in borrower's exercising of the options. In particular, we incorporate time-varying credit risk indicators, such as periodically updated FICO scores and bankruptcy filings, into the jointly estimated competing risks mortgage termination models. Third, we extend the analysis by increasing number of mass points used to describe heterogeneity parameters while aiming to obtain a better approximation to the true (possibly continuous) distribution of unobserved borrower's heterogeneity.

4.3. Static Explanatory Variables

As was summarized by Clapp et al. (2006), previous studies have shown that many observable variables that are available to lenders at the time of origination can provide strong signals and be useful in predicting future mortgage terminations. Among most commonly mentioned variables are payment-to-income and debt-to-income ratios, loan-to-value ratio, credit history, employment status, number of borrowers, borrower age, marital status, and number of dependents. Low credit score and high loan-to-value ratio usually provide strong opposite effects on two mortgage termination options: negative on prepayment and positive on default. Summarizing previously reported findings, we also expect to find negative effects on prepayment for age, borrowers who are self-employed, and cash-out refinancers.

Discount points have long been hypothesized to provide an indirect signal from borrowers about their intentions to remain in the house for a long period of time. However, this relationship has been rarely tested empirically due to unavailability of the data. Clapp et al. (2006) were able to test this relationship indirectly. They estimated discount points by regressing mortgage note rates on market rates and on other loan and borrower characteristics. Clapp et al. (2006) reported strong negative relationship between discount points and probability of prepayment. My analysis provides an opportunity to test the relationship between discount points and termination options directly, because the data we use contains loan-level information on actual discount points and prepaid finance charges paid at loan origination. we expect to confirm negative relationship between discount points and probability of prepayment reported by other authors.

The observation period that is covered by this study is characterized by a new and rapidly growing (at least through 2007) segment of the mortgage market: alternative documentation loans, or Alt-A low/no-doc loans. Alt-A mortgages were generally issued to borrowers with high credit scores but who were either unable or unwilling to fully disclose their financial statements and sources of income. This product is relatively new to the market, with limited available information on its performance history. Therefore, there have not been many studies that analyzed different types of risks associated with Alt-A loans. The loan-level data used for this study contains a flag that identifies borrowers who applied for the mortgage under alternative documentation guidelines and for whom no detailed income verification process was completed. This provides new opportunity to analyze the effects of Alt-A mortgages on options-based models and to test whether there are any inherent risks associated with these types of loans. Even though this group of borrowers is not very big (under 2% of the total population), we intend to include the limited documentation dummy in the estimated mortgage termination models. we expect to find negative relationship between limited documentation dummy and prepayment rates, because it is reasonable to assume that people who apply for Alt-A loans will find it more difficult to refinance in the future, comparing to full disclosure borrowers.

Table 3 presents definitions and brief descriptions of static explanatory variables that we used to jointly model mortgage terminations. Key option variables and a few related time-varying variables will be introduced in the next section as we start describing the modeling approach along with various model specifications. Additional summary statistics for both static and time-varying variables at different time snapshots are provided in Appendix 3.

5. Results

As Deng et al. (2000), we empirically estimate competing risks hazard models using the maximum likelihood approach on the large sample of fixed rate mortgages described earlier. The maximum likelihood estimates (MLE) of all unknown parameters, including prepayment and default baseline hazards, β-coefficients of covariates, and unobserved heterogeneity parameters $\eta_p$and η_d, are the values that maximize log-likelihood function $\ln(L)$, which is a log transformation of the joint density function of competing risks for all observations in the sample. Mathematically, $\ln(L)$ was defined in the previous section. For exact derivation, see McCall & McCall (2008).

5.1. Effects of observed variables

Table 4 present estimation results with two mass points. Since most of previous studies using similar methods use two mass points, it would serve as a well defined starting point. Call option is computed the same way it was done in Deng et al., and it is consistent with prior expectation and previous results. It is positively related to the probability of prepayment; a borrower is more likely to prepay her loan when the call option is larger. However, its effect is not linear, and negative estimates for squared call option implies its effect is strongly concave. Call option does not appear to be related to borrowers' default decision, implying that a large call option has little impact of borrower's decision to default. To gauge borrower's default option, we use the level of equity in the house rather than put option used by Deng et al. We choose to use the level of equity for the following two reasons. First, to construct put option, we need cross sectional standard deviation of housing returns, which is only available at state level publicly. We feel that state-wide standard errors might mask the true variability of housing returns in some areas. Second, during the data period, the housing markets went through historical level of volatility. But the put option variable by DQV is the probability of negative equity, and large price changes at the tail of the distribution might be underestimated by their put option measurement. The effect of current equity is consistent with prior expectation. A higher current equity implies a smaller probability of going default. As in case of call option, it appears non-linear, but its economic and statistical significance is much smaller. Interestingly, the level of equity also has large effects on prepayment; borrowers are more likely to prepay when the current equities are higher. DTI at origination does not have significant effects on prepayment, but has significant effects on default. Borrowers with higher DTI tend to default more often. Borrowers without documentation, with the loan for second home, or self-employed are more likely to default, but do not exhibit significantly different behavior for default. Downing, Stanton, and Wallace (2005) shows that rate-point pair can serve as separation mechanism between borrowers who are more likely to prepay early and borrowers who are not when prepayment is private information. Even though their model does not consider the effect of borrowers' private information about their default probability, it is likely that borrowers who are privately informed of their likelihood of default might choose lower points. Our results show that this might not be the case; borrowers with higher points are more likely to default. Borrowers with high FICO scores at origination are more likely to prepay and less likely to default as expected. In estimating effects of current FICO scores, we divide the current FICO scores into two variables, "FICO up", which indicates the FICO scores have increased since origination, and "FICO down", which indicates the FICO scores have fallen since origination. It is expected that "FICO up" is more correlated with higher prepayment and "FICO down" with default. Surprisingly, "FICO up" is related with lower prepayment, which we are not sure at this point why it is the case. However, the rest of effects are consistent with prior expectation; "FICO up" is highly correlated to lower default, "FICO down" is highly correlated with higher default and lower prepayment. Lastly, relaxation of underwriting standards is highly correlated with prepayment as expected.

5.2. Models with Varying Number of Mass Points

For the next test, we explore implications of having more mass points for unobserved heterogeneity. For this, we iteratively increase number of mass points for the unobserved heterogeneity vector (η_p, η_d) from two to five. The results are summarized in Table 5-1 through 5-3. The estimates of the corresponding baseline hazards are presented in Appendix 5 at the end.

The covariates coefficients for prepayment changed very little as we kept increasing number of mass points. This shows that prepayment effects identified by the simpler model are very stable. The covariates coefficients for default were also largely consistent across different mass-point parameterizations; all coefficients kept the same sign. However, there were a few observable changes that are worth mentioning.

As we suggested earlier, increasing number of mass points could help with better approximating the true distribution of the unobserved heterogeneity vector (η_p, η_d). In theory, (η_p, η_d) could belong to a wide range of distributions and the large number of mass points could be required to identify a particular type of distribution with any degree of precision. In practice, using a large number of mass points is computationally burdensome. Deng et al. (2000) only considered cases with two or three mass points and Clapp et al. (2006) only considered one case with two mass points. In our research, we refine discrete approximation of the unobserved heterogeneity by increasing number of mass points to five. The plots in Appendix 6 show how this discrete approximation of prepayment and default heterogeneities evolves as we move from the base case of M=2 to the case where M=5.

Analyzing the evolution of distribution of η_p, we can see that distribution of mass points for prepayment is rather stable, and consistent across different specification for mass point. In particular, the distribution of mass points for prepayment quite resembles bell-shaped and symmetric distribution such as normal distributions. This result generally agrees with previous result for two mass points obtained by Deng et al. (2000) who also identified the low prepayment risk group to be relatively large (25%) in the case of M=2. However, our results for three mass points diverge from Deng et al. (2000), who found that the low risk group is reduced to just 5% after going from the case with M=2 to the case with M=3. In our analysis, the portion of slow prepayers never drops below 20%.

Furthermore, we show that by increasing number of mass points to four and five, we can clearly separate the high risk group from the rest of the population. In our case, when M=5, the model identifies a subpopulation of borrowers who are substantially more likely to exercise their prepayment options comparing to an average borrower. This subpopulation of "financially savvy" borrowers is not very big: about 6% of the total population. Coincidentally, the same small group of savvy borrowers is also a lot more likely to exercise their default options, when it is optimal to do so. It would be interesting to try to understand what makes the group of financially savvy borrowers different from the rest of the population, but we clearly need to have more borrower-level data to carry on this additional analysis. Differences in mass points for prepayment are much smaller compared to differences in default.

Finally, it is interesting to see how the shape of the distribution of heterogeneity parameters changes with increased number of mass points. (See Appendix 6.) In the case of prepayment, the distribution of η_p becomes more and more concentrated around the mean, just like a normal (or similar "humped"-shaped) distribution. This is not, however, the case for the default heterogeneity parameter. The default heterogeneity distribution of η_d remains very skewed to the left and it is difficult to see what family of distributions it might belong to. This divergence between distribution shapes for prepayment and default heterogeneity parameters has not been documented prior to this research, and it is noteworthy by itself. It identifies and highlights the challenges one might face when trying to impose parametric assumptions on the unobserved heterogeneity vector (η_p, η_d). This finding also suggests that usual symmetric bell-shape distributions would not fit joint distribution of mass points for prepayment and default.

6. IMPLICATIONS FOR REGULATORY CAPITAL AND BASEL II

6.1 Basel I and Basel II

Mortgages are considered risky assets on which banks are required to hold a minimum reserve, or minimum Regulatory Capital (RC). In 1988, the Basel Committee on Banking Supervision introduced its first Basel Accord, which later became known as Basel I. The result of Basel I was a general framework around standardization of capital requirements for banks and financial institutions around the world. Under Basel I, the main rule for calculating minimum capital requirements was rather simplistic: banks had to demonstrate that their total capital was larger than 8% of their risk weighted assets (RWA). Here, RWA was calculated as weighted average of individual bank.

$$RWA = \sum_{i}^{I} A_i r_i$$

where A_i is the value of the i-th asset and r_i is the risk weight of the exposure. The main weakness of Basel I was that there were only a handful of different categories of assets, and fixed risk weights were assigned to each category. For example, under Basel I all mortgages were assigned a fixed 50% risk weight regardless of borrowers' creditworthiness and regardless of numerous types of mortgage products available in the market. Thus, both subprime and prime mortgage lenders were subject to the same minimum regulatory capital thresholds and were required to hold capital of 4% of their respective mortgage portfolios.

Consequently, the Basel Committee issued the New Basel Accord, which became known as Basel II. The purpose of Basel II was to replace the much criticized "one-size-fits-all" framework with a more risk-sensitive and more sophisticated approach. Under Basel II, the banks are now allowed to choose between Standardized approach and Internal Ratings Based (IRB) approach to calculating risk-weighted assets. Furthermore, banks with sufficiently developed internal modeling capabilities and capital allocation infrastructure could opt in to select Advanced IRB approach, which allows the use of internally-developed credit risk models for the calculation of regulatory capital. Though this approach is inherently more complex, both from development and implementation points of view, it is expected that large financial institutions will still find it more advantageous to pursue Advanced IRB approach. Through application of sophisticated credit risk models, banks have a potential to attain superior precision in risk quantification, comparing to simpler alternatives.

6.2 RC for Mortgages

When it comes to residential mortgage assets, banks choosing the Advanced IRB approach will calculate the risk weights of their retail mortgage portfolio as follows:

$$RWA = 12.5\% \times K(PD) \times EAD \times LGD$$

and

$$K(PD) = \Phi\left(\frac{\Phi^{-1}(PD) + \sqrt{0.15}\Phi^{-1}(0.999)}{\sqrt{1-0.15}}\right) - PD\ ,$$

where PD is probability of exposure defaulting in the next 12 months, EAD is exposure at default, LGD is loss given default, and Φ is the standard normal distribution function. Banks are allowed to estimate their own versions of PD, EAD, and LGD models under a variety of statistical methodologies. To estimate the models, banks are encouraged to use their own internal data that satisfies minimum data requirements as defined by the final Basel II ruling.

For mortgage products, EAD and LGD models could be very simple and range from segment-level averages to loan-level regressions. Estimation of PD rates, on the other hand, is a wide open proposition. Direct estimation of 12-month cumulative default rate is a simple option and it could be accomplished using a variety of statistical techniques such as classification and regression tree analysis, cluster regression analysis, logit or probit regressions, etc. The competing risks hazard methodology described in this paper is a more sophisticated approach which could result in a more precise way of estimating Basel-compliant PD rates. Superior precision of competing risks hazard with unobserved heterogeneity approach comes from the following two notions.

First, when estimating mortgage terminations using competing risks hazard methodology we explicitly account not only for default risk but also for prepayment risk, which could be correlated with defaults. Accounting for prepayment is important because mortgage payoffs reduce number of loans at risk and, thus, affect the denominator for calculating mortgage default rates.

Second, using competing risks hazards with unobserved heterogeneity allows identification of homogeneous groups of borrowers which, as we have shown previously, could have quite different propensities to exercising their default and prepayment options. By increasing number of mass points in the distribution of unobserved heterogeneity parameters, we can estimate the tails of the PD distribution and, therefore, regulatory capital requirements more precisely.

6.3. Empirical tests.

To understand the impact unobserved heterogeneity can cause on calculation of minimum capital requirements under Basel II Advanced IRB approach, we applied the models estimated earlier to the quarterly snapshots of the available performance data. We then derived Basel PD rates from the estimated conditional default rates. Since we did not have access to the final loss and recovery data on the defaulted accounts, we used state-level average LGD rates from industry level data. Basel II guidelines suggest estimating LGD over a mix of economic conditions or during economic downturn conditions, and then use the most conservative of the two estimates. For our tests, we used average LGD rates associated with 2008 mortgage defaults which generated the most conservative LGD estimates. EAD was approximated by loan balance at the time of quarterly snapshots.

Our results demonstrate that RC calculations could be affected by selection of unobserved heterogeneity distribution. The largest impact on regulatory capital was observed when moving from the distribution approximated by two mass points to the distribution approximated by three mass points. We attribute this to the emergence of a small group of high risk borrowers that we were able to detect and identify only after increasing number of mass points to three and higher.

7. Conclusions

This research extends the results of Deng, Quigley, and Van Order (2000) on unobserved heterogeneity of borrowers with respect to optimal exercising of mortgage termination options. First, we show that by including more static borrower-specific and loan-specific explanatory variables that are generally available at the time of mortgage origination, the unobserved heterogeneity can be reduced when comparing to that of a simple ruthless competing risks hazard model. The unobserved heterogeneity is reduced even further, by 33%, after we incorporate time-varying information on borrower creditworthiness and previously missed refinance opportunities into the model. Still, even after including more than a dozen covariates into the model, we estimate that between 13% and 18% of all borrowers will respond very slowly, if at all, to good refinance opportunities presented to them. This non-response cannot be explained by the available data.

Second, we demonstrate that by increasing number of mass points that are used to discretely approximate unknown distributions of unobserved heterogeneity parameters, we obtain improved model fit and we observe that the distribution of prepayment heterogeneity is starting to converge to a hump-shaped distribution.

This paper provides interesting insights on the relationships between extended list of explanatory variables and unobserved heterogeneity between mortgage holders. In particular, it shows that even though current borrower credit rating is definitely very valuable information in identifying subpopulation of borrowers eligible to take advantage of refinancing, this information by itself does not guarantee that borrowers would immediately jump on this opportunity. Therefore, there could be other factors, objective or subjective, that might be holding them off. Do mortgage holders who do not refinance do so because their true LTV ratio, including any additional junior lien loans and home equity lines of credit, no longer meets underwriting guidelines for prime loans? Or, do they miss out because they were unaware that such refinance opportunities existed? Additional research, involving even more borrower-level and property-level explanatory variables, might be necessary in order to answer these important questions.

Lastly, we show the implications of using right specification for unobserved heterogeneity for minimum capital requirement. Our results demonstrate that RC calculations could be affected by selection of unobserved heterogeneity distribution. The largest impact on regulatory capital was observed when moving from the distribution approximated by two mass points to the distribution approximated by three mass points. We attribute this to the emergence of a small group of high risk borrowers that we were able to detect and identify only after increasing number of mass points to three and higher.

8. References

Ambrose, B. W., & Buttimer, R. J.,Jr. (2000). Embedded options in the mortgage contract. *Journal of Real Estate Finance and Economics, 21*(2), 95.

Ambrose, B. W., & Capone, C. A. (1998). Modeling the conditional probability of foreclosure in the context of single-family mortgage default resolutions. *Real Estate Economics, 26*(3), 391.

Archer, W. R., & Ling, D. C. (1993). Pricing mortgage-backed securities: Integrating optimal call and empirical models of prepayment. *Journal of the American Real Estate and Urban Economics Association, 21*(4), 373.

Capozza, D. R., Kazarian, D., & Thomson, T. A. (1998). The conditional probability of mortgage default. *Real Estate Economics, 26*(3), 359.

Capozza, D. R., & Thomson, T. A. (2005). Optimal stopping and losses on subprime mortgages. *Journal of Real Estate Finance and Economics, 30*(2), 115.

Clapp, J. M., Deng, Y., & An, X. (2006). Unobserved heterogeneity in models of competing mortgage termination risks. *Real Estate Economics, 34*(2), 243.

Cox, D. (1972). Regression models with life tables (with discussion). *Journal of the Royal Statistical Society B, 34*, 187.

Cunningham, D. F., & Hendershott, P. H. (1984). Pricing FHA mortgage default insurance. *Housing Finance Review, 3*(4), 373-392.

Deng, Y., Quigley, J. M., & Van Order, R. (2000). Mortgage terminations, heterogeneity and the exercise of mortgage options. *Econometrica, 68*(2), 275.

Downing, C., Stanton, R. & Wallace, N. (2005). An empirical test of a two-factor mortgage valuation model: How much do house prices matter? *The Real Estate Economics (2005), 33*(4), 681.

Dunn, K. B., & McConnell, J. J. (1981). Valuation of GNMA mortgage-backed securities. *Journal of Finance, 36*(3), 599-616.

Epperson, J. F., Kau, J. B., Keenan, D. C., & Muller, W. J.,III. (1985). Pricing default risk in mortgages. *AREUEA Journal, 13*(3), 261.

Han, A., & Hausman, J. A. (1990). Flexible parametric estimation of duration and competing risk models. *Journal of Applied Econometrics, 5*, 1-28.

Kau, J. B., & Keenan, D. C. (1995). An overview of the option-theoretic pricing of mortgages. *Journal of Housing Research, 6*(2), 217.

Kau, J. B., & Keenan, D. C. (1999). Patterns of rational default. *Regional Science and Urban Economics, 29*(6), 765.

Kau, J. B., Keenan, D. C., & Kim, T. (1993). Transaction costs, suboptimal termination and default probabilities. *Journal of the American Real Estate and Urban Economics Association, 21*(3), 247.

Kau, J. B., Keenan, D. C., Muller, W. J.,III, & Epperson, J. F. (1992). A generalized valuation model for fixed-rate residential mortgages. *Journal of Money, Credit, and Banking, 24*(3), 279.

Kau, J. B., & Kim, T. (1994). Waiting to default: The value of delay. *Journal of the American Real Estate and Urban Economics Association, 22*(3), 539.

McCall, B. P. (1996). Unemployment insurance rules, joblessness, and part-time work. *Econometrica, 64*(3), 647.

McCall, B. P., & McCall, J. J. (2008). *The economics of search.* Routledge Advances in Experimental and Computable Economics; London and New York:; Taylor and Francis, Routledge.

Phillips, R. A., & VanderHoff, J. H. (2004). The conditional probability of foreclosure: An empirical analysis of conventional mortgage loan defaults. *Real Estate Economics, 32*(4), 571.

Quigley, J. M., & Van Order, R. (1995). Explicit tests of contingent claims models of mortgage default. *Journal of Real Estate Finance and Economics, 11*(2), 99-117.

Quigley, J. M., & Van Order, R. (1990). Efficiency in the mortgage market: The borrower's perspective. *AREUEA Journal, 18*(3), 237.

Schwartz, E. S., & Torous, W. N. (1989). Prepayment and the valuation of mortgage-backed securities. *The Journal of Finance, 44*(2), 375.

Stanton, R. (1995). Rational prepayment and the valuation of mortgage-backed securities. *The Review of Financial Studies (1986-1998), 8*(3), 677.

Sueyoshi, G. T. (1992). Semiparametric proportional hazards estimation of competing risks models with time-varying covariates. *Journal of Econometrics, 51*(1-2), 25.

VanderHoff, J. (1996). Adjustable and fixed rate mortgage termination, option values and local market conditions: An empirical analysis. *Real Estate Economics, 24*(3), 379.

Table 1. Loan-specific Summary Statistics (averages), by Origination Year

Origination Year	Loan Count	FICO Score	Loan-to-Value (LTV)	Debt-to-Income (DTI)	Second Home (%)	%Equity Cashout (x100%)	Prepaid Finance Charges (incl. Points)
1999	2,305	712	0.89	38.7	1.7%	9.0	2.0
2000	758	702	0.86	38.1	3.7%	9.5	2.1
2001	4,180	719	0.81	37.1	2.7%	6.4	0.6
2002	4,690	738	0.77	37.0	1.7%	5.7	0.2
2003	3,094	741	0.75	36.0	1.3%	4.6	0.3
2004	2,170	730	0.76	37.9	1.9%	5.7	0.3
2005	3,800	719	0.81	38.7	1.6%	6.9	0.3
2006	5,336	714	0.84	39.4	2.6%	7.1	0.4
2007	6,034	725	0.80	38.3	2.9%	6.9	0.3
2008	4,975	752	0.65	35.1	3.2%	4.8	0.4
Total	37,342	728	0.78	37.6	2.3%	6.4	0.5

Table 2. Borrower-specific Summary Statistics (averages), by Origination Year

Origination Year	Loan Count	Borrower's Age (avg, if 2 borrowers)	Years at Current Residency	Self-Employed (%)	No-income Verification (%)	Two Borrowers (%)	Married (%)	Have Dependants (%)
1999	2,305	46.7	10.0	12.0%	0.0%	53.8%	74.0%	27.2%
2000	758	48.8	12.1	11.6%	0.0%	52.5%	67.2%	17.0%
2001	4,180	44.9	8.6	8.3%	1.4%	56.7%	77.1%	22.5%
2002	4,690	44.3	7.3	9.0%	5.4%	56.8%	80.8%	30.3%
2003	3,094	43.3	6.7	7.0%	0.6%	55.8%	78.3%	44.3%
2004	2,170	44.0	7.4	6.8%	1.1%	52.0%	73.9%	26.8%
2005	3,800	42.5	6.8	6.1%	1.4%	49.5%	70.4%	16.8%
2006	5,336	43.3	7.3	7.1%	1.2%	50.1%	71.9%	11.9%
2007	6,034	44.4	8.1	9.0%	1.1%	48.1%	72.1%	11.8%
2008	4,975	47.0	8.2	10.2%	0.3%	53.0%	76.9%	14.1%
Total	37,342	44.6	7.9	8.5%	1.5%	52.5%	74.8%	20.8%

Table 3. Definitions of Static Variables

Variable Name	Variable Name (Short)	Description	Expected Sign in the Model	
			Prepayment	Default
Original Loan-to-Value Ratio	*OrigLTV(0,1)*	Combined value of all known liens at the time of origination divided by the property value at origination (as ratio)	-	+
Debt-to-Income Ratio	*DTI(%)*	Total monthly debt obligation divided by the gross monthly income, at origination (%)	-	+
Origination FICO Score	*OrigFICO*	Fair, Isaac and Company credit score at origination	+	-
Percent Equity Cashout	*COR(%)*	Difference between new and old loan amounts divided by the property value at origination (%)	-	?
Discount Points and Prepaid Finance Charges	*Points*	Total amount of prepaid finance charges, including discount points to buydown the note rate, expressed as percent of loan amount (%)	-	+
Limited Income Documentation Dummy	*No Income Verif*	Equals one if the loan was underwriten using low/no-income documentation guidelines, zero otherwise	-	+
Self-employed Dummy	*Self Employed?*	Equals one if the borrower was self-emplyed at the time of origination, zero otherwise	-	+
Number of Borrowers	*2 borrowers?*	Equals one if two borrowers applied for the loan, zero otherwise		
Borrower's Age	*Borrower's Age*	Age of the borrower in years. (Average age, if two borrowers.)	-	?
Marital Status Dummy	*Married?*	Equals one if the borrower was married at the time of origination, zero otherwise	?	?
Dependants Dummy	*Has Dependants?*	Equals one if the borrower had dependants at the time of origination, zero otherwise	?	?
Age and Dependants Dummy	*Age<50 w/Kids*	Equals one if the borrower is less than 50 years old and had dependants at the time of origination, zero otherwise	?	?
Years at Current Residency	*Years @cur resid*	Number of years the borrower lived in the current house	?	?

Table 4 : Estimation results with 2 Mass Points

Variables	Prepay	t-stat	Default	t-stat
Call Option	7.6219	37.55	1.6516	1.94
Equity(%)	5.0461	33.11	-4.0510	-17.17
Call Option^2	-8.9555	-9.42	0.8318	0.25
Equity^2(%)	0.8067	4.03	-2.0335	-6.44
OrigLTV(0,1)	4.8878	22.44	-5.3281	-9.50
DTI	-0.0007	-0.73	0.0186	4.85
Low-Doc	-0.4273	-5.18	0.1249	0.32
2nd-Home	-0.1920	-3.08	0.2549	1.09
Self-Employed	-0.1338	-3.77	0.2174	1.83
Points	-0.1552	-10.16	0.1154	2.20
COR%	0.0000	0.00	-0.0010	-0.32
2-Borrowers	0.0473	2.08	-0.1540	-1.86
Origination FICO	0.0021	8.10	-0.0161	-18.24
Borrower Age	-0.0361	-6.24	-0.0719	-4.14
Borrowers Age^2	0.0003	5.14	0.0008	4.70
Married	0.0967	3.65	-0.2262	-2.60
Dependents	-0.1235	-2.59	0.0428	0.23
Age<50 with Kids	0.1353	2.66	0.0301	0.14
Years@CurResid	-0.0210	-5.63	-0.0226	-1.78
Years@CurResid^2	0.0003	2.62	0.0005	1.52
FICO up	-0.0039	-5.47	-0.0358	-4.83
FICO down	-0.0033	-11.05	0.0172	32.25
FICO up + COITM	0.0019	2.45	0.0171	2.22
UER(%)	-0.0055	-0.80	0.0116	0.62
Post 2007	-0.2755	-6.67	0.2737	1.53
Bk Ch 7	-0.4853	-1.54	0.1850	0.24
Bk Ch 13	-0.8896	-2.46	0.4823	0.80
Bk 07 filed BAR	0.2347	0.70	0.0785	0.10
Underwriting Idx	0.0028	4.92	0.0002	0.09

LOC1	0.00002		11.04194	0.25
LOC2	0.00011		1.14400	0.75

E(LOC)	0.00009		3.63374	
std(LOC)	0.00004		4.29471	
Correlation		-1		
Log Likelihood		64508.74		
AIC		64631.74		

Table 5-1: Estimation results with 3 Mass Points

	Prepay	t-stat	Default	t-stat
Call Option	7.6694	37.22	2.7521	2.97
Equity(%)	5.0484	32.84	-4.1907	-14.75
Call Option^2	-8.5466	-8.86	3.2049	0.89
Equity^2(%)	0.8488	4.21	-1.9551	-5.40
OrigLTV(0,1)	4.9485	22.38	-5.7312	-8.78
DTI	-0.0005	-0.45	0.0219	4.96
Low-Doc	-0.4391	-5.34	-0.1720	-0.35
2nd-Home	-0.1869	-2.98	0.2382	0.87
Self-Employed	-0.1331	-3.74	0.1788	1.29
Points	-0.1559	-10.11	0.0694	1.18
COR%	0.0001	0.07	-0.0013	-0.34
2-Borrowers	0.0453	1.99	-0.1724	-1.80
Origination FICO	0.0019	7.57	-0.0182	-17.27
Borrower Age	-0.0366	-6.29	-0.0886	-4.49
Borrowers Age^2	0.0003	5.21	0.0010	5.14
Married	0.0964	3.62	-0.2202	-2.19
Dependents	-0.1274	-2.67	-0.0171	-0.08
Age<50 with Kids	0.1404	2.75	0.1007	0.43
Years@CurResid	-0.0214	-5.71	-0.0256	-1.79
Years@CurResid^2	0.0003	2.70	0.0005	1.49
FICO up	-0.0040	-5.59	-0.0340	-4.55
FICO down	-0.0030	-9.84	0.0195	28.60
FICO up + COITM	0.0019	2.38	0.0136	1.74
UER(%)	-0.0059	-0.85	0.0081	0.39
Post 2007	-0.2727	-6.59	0.2757	1.45
Bk Ch 7	-0.3839	-1.20	0.2525	0.33
Bk Ch 13	-0.8435	-2.27	0.5894	1.01
Bk 07 filed BAR	0.2160	0.64	0.0701	0.09
Underwriting Idx	0.0028	5.00	0.0013	0.63

LOC1	0.00015		25.35172	0.15
LOC2	0.00011		0.36393	0.64
LOC3	0.00002		2.53903	0.21

E(LOC)	0.00009		4.52912	
std(LOC)	0.00004		8.72389	
Correlation		0.4515		
Log Likelihood		64466.72		
AIC		64592.72		

Table 5-2: Estimation results with 4 Mass Points

Variables	Prepay	t-stat	Default	t-stat
Call Option	7.6436	37.54	3.1250	3.14
Equity(%)	5.0364	32.92	-4.6465	-14.21
Call Option^2	-8.2957	-8.54	4.1607	1.08
Equity^2(%)	0.8436	4.21	-2.0673	-4.91
OrigLTV(0,1)	4.9412	22.51	-6.5599	-8.80
DTI	-0.0004	-0.37	0.0189	3.91
Low-Doc	-0.4445	-5.47	-0.2937	-0.54
2nd-Home	-0.1808	-2.89	0.2325	0.81
Self-Employed	-0.1296	-3.66	0.2476	1.63
Points	-0.1542	-10.03	0.0458	0.71
COR%	0.0002	0.20	-0.0010	-0.24
2-Borrowers	0.0431	1.90	-0.1610	-1.53
Origination FICO	0.0019	7.47	-0.0206	-16.86
Borrower Age	-0.0367	-6.34	-0.0970	-4.36
Borrowers Age^2	0.0003	5.27	0.0011	5.05
Married	0.0947	3.57	-0.2807	-2.53
Dependents	-0.1261	-2.66	0.0286	0.12
Age<50 with Kids	0.1392	2.75	0.0528	0.21
Years@CurResid	-0.0218	-5.85	-0.0368	-2.34
Years@CurResid^2	0.0003	2.84	0.0006	1.75
FICO up	-0.0040	-5.61	-0.0359	-4.75
FICO down	-0.0029	-9.45	0.0208	26.35
FICO up + COITM	0.0018	2.36	0.0130	1.65
UER(%)	-0.0054	-0.78	0.0027	0.12
Post 2007	-0.2716	-6.58	0.3218	1.63
Bk Ch 7	-0.3259	-1.00	0.2754	0.36
Bk Ch 13	-0.7855	-2.00	0.4875	0.80
Bk 07 filed BAR	0.1994	0.57	0.0815	0.10
Underwriting Idx	0.0028	5.03	0.0024	1.07

LOC1	0.00016		42.20653	0.07
LOC2	0.00011		0.02383	0.44
LOC3	0.00010		2.29894	0.3
LOC4	0.00002		0.56177	0.19

E(LOC)	0.00009		3.65104	
std(LOC)	0.00004		10.41647	
Correlation		0.4733		
Log Likelihood		64448.10		
AIC		64577.10		

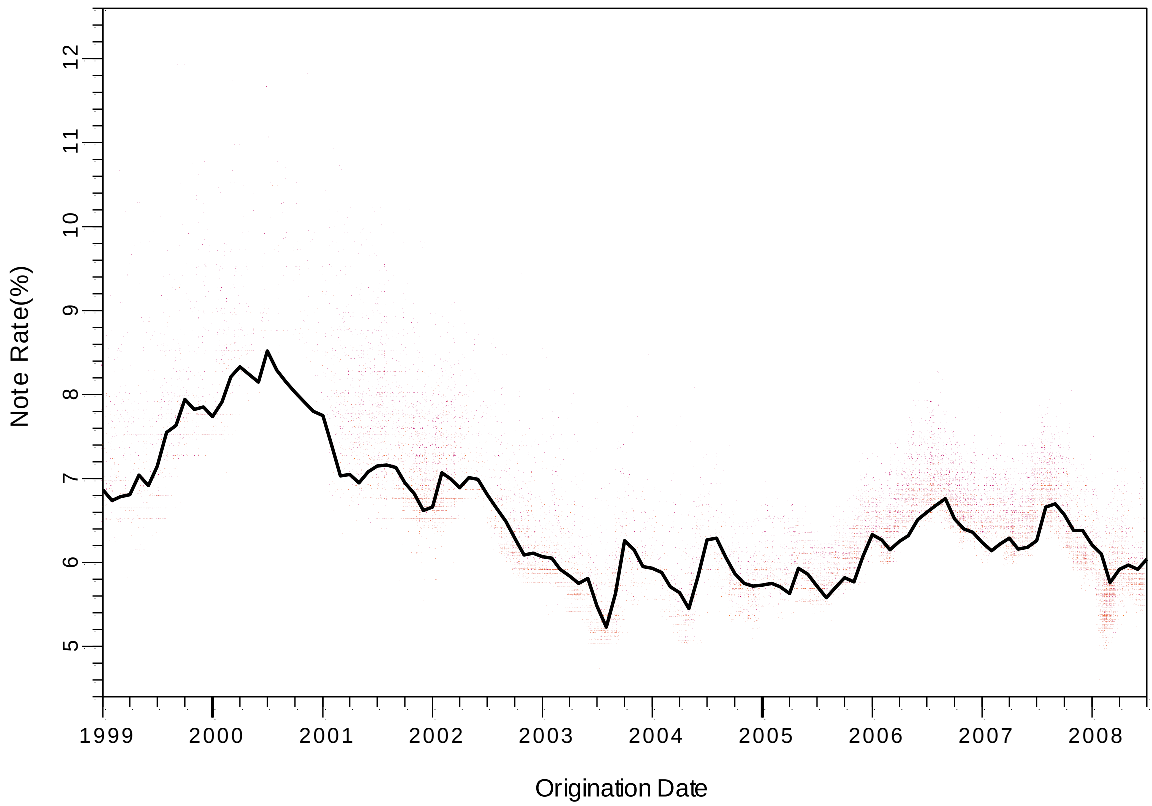
Table 5-3: Estimation results with 5 Mass Points

Variables	Prepay	t-stat	Default	t-stat
Call Option	7.6837	37.12	3.2485	3.27
Equity(%)	5.0629	32.59	-4.6667	-14.05
Call Option^2	-8.3180	-8.54	3.4234	0.89
Equity^2(%)	0.8791	4.33	-2.0522	-4.82
OrigLTV(0,1)	5.0066	22.10	-6.5549	-8.73
DTI	-0.0004	-0.41	0.0188	3.87
Low-Doc	-0.4462	-5.45	-0.2971	-0.55
2nd-Home	-0.1740	-2.72	0.2354	0.82
Self-Employed	-0.1326	-3.72	0.2591	1.70
Points	-0.1572	-10.03	0.0461	0.71
COR%	0.0003	0.31	-0.0008	-0.20
2-Borrowers	0.0432	1.89	-0.1685	-1.60
Origination FICO	0.0019	7.47	-0.0207	-16.82
Borrower Age	-0.0374	-6.36	-0.0984	-4.39
Borrowers Age^2	0.0003	5.30	0.0011	5.09
Married	0.0950	3.55	-0.2773	-2.47
Dependents	-0.1302	-2.72	0.0139	0.06
Age<50 with Kids	0.1441	2.83	0.0680	0.27
Years@CurResid	-0.0217	-5.77	-0.0367	-2.33
Years@CurResid^2	0.0003	2.76	0.0006	1.72
FICO up	-0.0040	-5.61	-0.0358	-4.73
FICO down	-0.0030	-9.40	0.0208	26.18
FICO up + COITM	0.0018	2.33	0.0128	1.62
UER(%)	-0.0059	-0.85	0.0031	0.13
Post 2007	-0.2680	-6.47	0.3180	1.61
Bk Ch 7	-0.2873	-0.81	0.2857	0.37
Bk Ch 13	-0.7966	-2.03	0.5135	0.84
Bk 07 filed BAR	0.1527	0.41	0.0824	0.10
Underwriting Idx	0.0029	5.08	0.0024	1.09

LOC1	0.00015		36.84580	0.06
LOC2	0.00011		0.01436	0.43
LOC3	0.00010		2.08804	0.28
LOC4	0.00001		0.62738	0.04
LOC5	0.00002		0.42939	0.19

E(LOC)	0.00009		3.08662	
std(LOC)	0.00004		8.93065	
Correlation		0.43381		
Log Likelihood		64447.39		
AIC		64579.39		

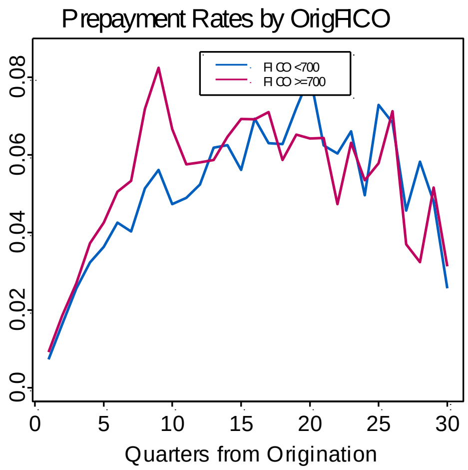
Figure 1. Note Rates (%) of Originated Mortgages

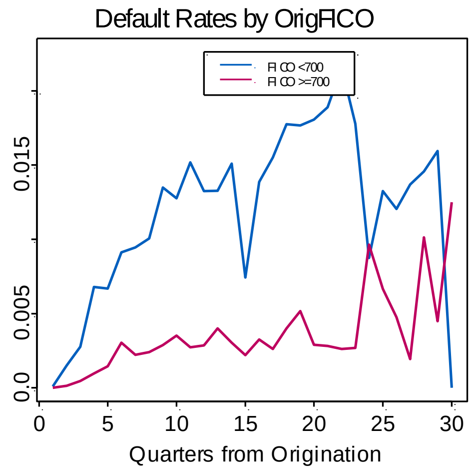


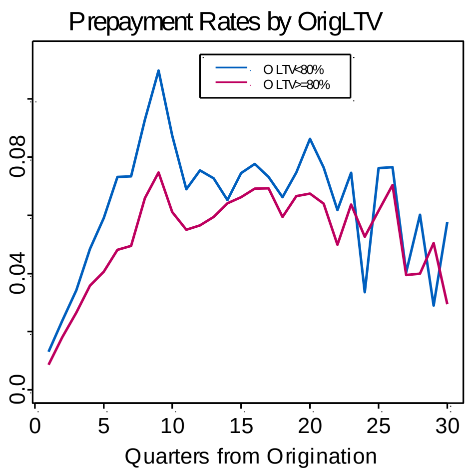
Appendix 1. Geographical Distribution by State

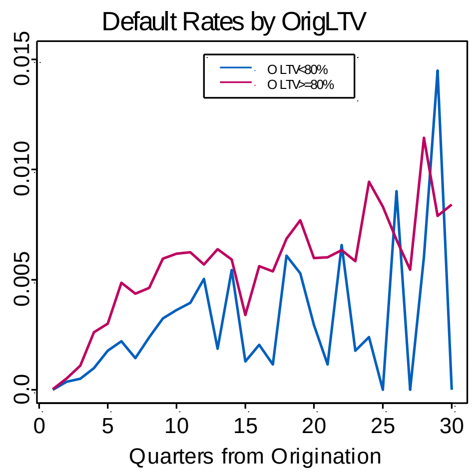
State	State Name	Loan Count	(%)
AL	Alabama	356	1.0%
AR	Arkansas	309	0.8%
AZ	Arizona	1171	3.1%
CA	California	4337	11.6%
CO	Colorado	631	1.7%
CT	Connecticut	566	1.5%
DC	District of Columbia	65	0.2%
DE	Delaware	96	0.3%
FL	Florida	2359	6.3%
GA	Georgia	1577	4.2%
IA	Iowa	305	0.8%
ID	Idaho	267	0.7%
IL	Illinois	1455	3.9%
IN	Indiana	606	1.6%
KS	Kansas	390	1.0%
KY	Kentucky	358	1.0%
LA	Louisiana	379	1.0%
MA	Massachusetts	819	2.2%
MD	Maryland	971	2.6%
ME	Maine	194	0.5%
MI	Michigan	1074	2.9%
MN	Minnesota	565	1.5%
MO	Missouri	633	1.7%
MS	Mississippi	172	0.5%
MT	Montana	65	0.2%
NC	North Carolina	2476	6.6%
ND	North Dakota	16	0.0%
NE	Nebraska	175	0.5%
NH	New Hampshire	332	0.9%
NJ	New Jersey	2145	5.7%
NM	New Mexico	252	0.7%
NV	Nevada	442	1.2%
NY	New York	1890	5.1%
OH	Ohio	1696	4.5%
OK	Oklahoma	241	0.6%
OR	Oregon	582	1.6%
PA	Pennsylvania	1840	4.9%
RI	Rhode Island	185	0.5%
SC	South Carolina	509	1.4%
SD	South Dakota	58	0.2%
TN	Tennessee	480	1.3%
TX	Texas	89	0.2%
UT	Utah	533	1.4%
VA	Virginia	1662	4.5%
VT	Vermont	71	0.2%
WA	Washington	1039	2.8%
WI	Wisconsin	633	1.7%
WV	West Virginia	230	0.6%
WY	Wyoming	46	0.1%
	Total	37342	100.0%

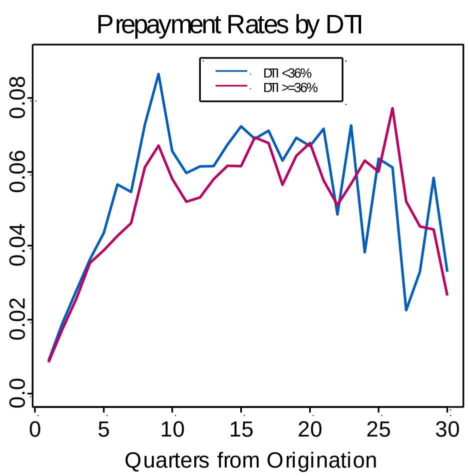
Appendix 2. Empirical Hazard Rates for Selected Covariates

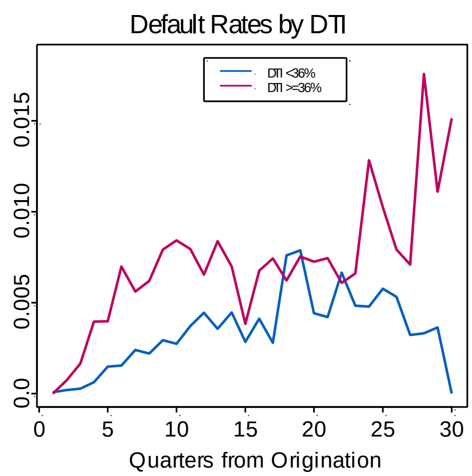








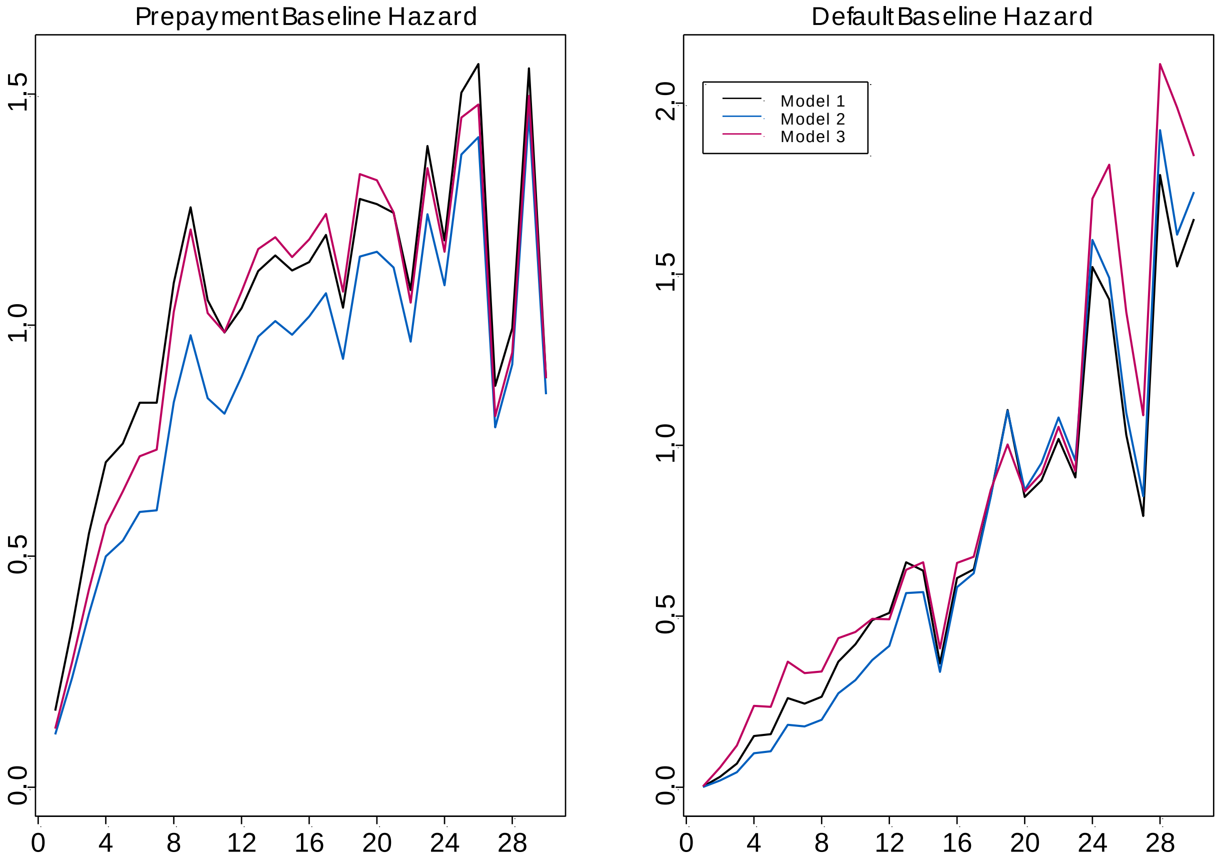




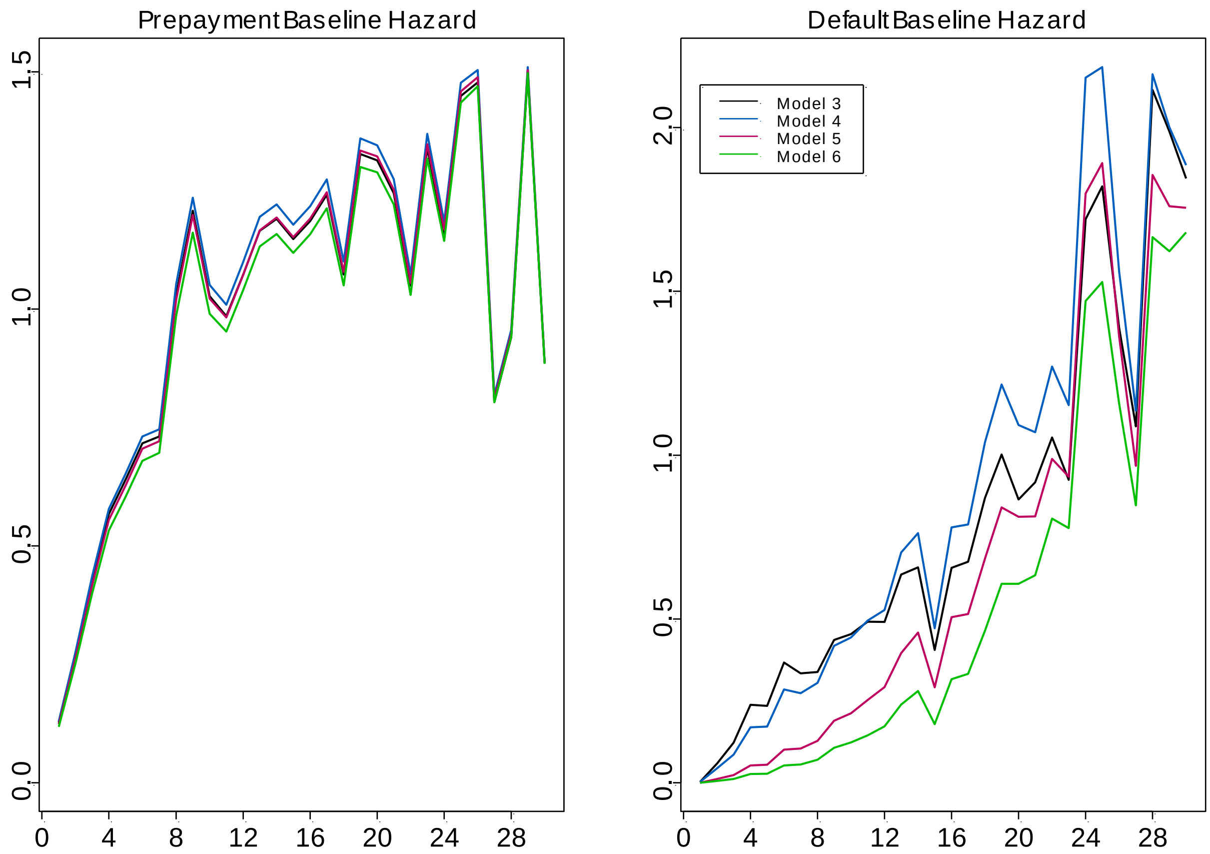
Appendix 3. Summary Statistics of Explanatory Variables at Origination and Termination

Variables	At Origination				At Termination (or at Last Observed Period)			
	All	Defaulted	Prepaid	Censored	All	Defaulted	Prepaid	Censored
Number of Observations	37,342	1,229	14,953	21,160	37,342	1,229	14,953	21,160
Call Option	0.01	0.05	0.01	0.01	0.09	0.11	0.06	0.11
Put Option	0.15	0.27	0.14	0.15	0.15	0.26	0.05	0.22
Original LTV	0.78	0.89	0.77	0.79	0.78	0.89	0.77	0.79
Mark-to-Market LTV	0.78	0.88	0.76	0.78	0.72	0.84	0.63	0.77
DTI (%)	37.6	41.9	37.0	37.8	37.6	41.9	37.0	37.8
Origination FICO	728	688	730	729	728	688	730	729
Updated FICO					719	555	726	723
Change in FICO Score					-9	-133	-3	-6
State Unemployment Rate (%)	5.10	4.68	5.19	5.07	7.26	6.17	5.41	8.64
Borrower Age	45	45	45	45	47	47	47	47
Years at Current Residency	8	9	8	8	8	9	8	8
Points	0.49	0.82	0.64	0.37	0.49	0.82	0.64	0.37
Percent Equity Cashout (%)	6.35	6.55	6.31	6.37	6.35	6.55	6.31	6.37

Appendix 4. Estimated Baseline Hazards for Models in Table 4.



Appendix 5. Estimated Baseline Hazards for Models in Table 5.



Appendix 6. Normalized Distributions of Heterogeneity Parameters in Models 3-6.