Mortgage Termination: An Empirical Hazard Model with Stochastic Term Structure

Yongheng Deng, Ph.D.*

University of California at Berkeley

Department of Economics

Berkeley, CA 94720-3880

October, 1995

* The author's current mailing address is: Office of Federal Housing Enterprise Oversight, 1700 G Street, NW, Washington, DC 20552, Tel: (202) 414-3837, Fax: (202) 414-3826, Email: deng@econ.berkeley.edu

Abstract

Pricing for mortgage and mortgage-backed securities is complicated due to the stochastic and interdependent nature of prepayment and default risks. This paper presents a unified economic model of the contingent claims and competing risks of mortgage termination by prepayment and default. I adopt a proportional hazard framework to analyze these competing and interdependent risks in a model with time-varying covariates. The paper incorporates a stochastic interest rate model into the hazard function for prepayment.

The empirical results reported in the paper provide new evidence about the ruthlessness of default and prepayment behavior and the sensitivity of these decisions to demographic as well as financial phenomena. The results also illustrate that evaluating the interest rate contingent claims with a stochastic term structure has effects not only on predicting the mortgage prepayment behavior but also on predicting the mortgage default behavior.

Key Words: hazard models, mortgage termination, semiparametric estimation, stochastic term structure

Mortgage Termination: An Empirical Hazard Model with Stochastic Term Structure

Pricing for mortgage and mortgage-backed securities is complicated due to the stochastic and interdependent nature of prepayment and default risks. It is widely accepted that mortgages can be viewed as ordinary debt instruments with various options attached to them. Default is a put option; the borrower sells his house back to the lender in exchange for eliminating the mortgage obligation. Prepayment is a call option; the borrower exchanges the unpaid balance on the debt instrument for a release from further obligation.

The contingent claims model, developed by Black and Scholes (1973) and Cox, Ingersoll, and Ross (1985), provides a rationale for borrower behavior, and a number of studies have applied this model to the mortgage market (e.g., Dunn and McConnell 1981, Buser and Hendershott 1984, Brennan and Schwartz 1985, Kau, Keenan, Muller, and Epperson 1992, 1995, Harding 1994, Quigley and Van Order 1995). Hendershott and Van Order (1987) and Kau and Keenan (1995) provide surveys of these models and results.

Several recent empirical studies have applied the Cox proportional hazard model (Cox and Oakes 1984) to evaluate mortgage default or prepayment risk (e.g., Green and Shoven 1986, Schwartz and Torous 1989, Quigley and Van Order 1990, 1995). Instead of solving for the unique critical values of the state variables in the contingent claims model, the proportional hazard model assumes that at each point in time during the mortgage contract period, the mortgage has a certain probability of termination, conditional upon the survival of the mortgage. The hazard function in this model is de…ned as the product of a baseline hazard and a set of time-varying covariates. These covariates need not be limited to the option value itself; they may include other important determinants of behavior. The proportional hazard model can thus incorporate reasonable mortgage prepayment and default behavior that would be considered "sub-optimal" under the pure contingent claims framework.

However, most of the empirical econometric models in mortgage studies evaluate the prepayment option with deterministic term structures. This is appropriate if there are no transactions costs in the mortgage termination, or if the borrowers have perfect information about future interest rate movement. Obviously, neither of the two assumptions is realistic in the mortgage market.

In this paper, I present an algorithm to incorporate a binomial mean-reverting interest rate model into the proportional hazard framework to analyze mortgage prepayment and default risks empirically. In the empirical analysis, I compare the results of the model using this stochastic term structure with similar models using a deterministic term structure.

The paper is organized as follows: section 1 reviews brie‡y the contingent claims model, as well as the proportional hazard model with competing risks. Section 2 describes the data used in this analysis and presents the empirical model for mortgage termination by prepayment and default. Section 3 presents an extensive empirical analysis. Section 4 is the conclusion.

1. The Contingent Claims and Competing Risks

A starting point for studying mortgage termination is the contingent claims model. It has been well accepted that prepayment in the mortgage contract can be viewed as a call option, and default in the mortgage contract can be viewed as a put option. Well-informed borrowers in a perfectly competitive market will exercise either of these two options when they can thereby increase their wealth. Absent either transactions costs or reputation costs which reduce credit ratings, these individuals can increase their wealth by defaulting on a mortgage when the market value of the mortgage exceeds the value of the house. Similarly, by prepaying the mortgage when market value exceeds par, they can increase wealth by re…nancing. However, if a borrower chooses to prepay the mortgage, he forfeits the opportunity to exercise either prepayment or default option in the future.

Likewise, if a borrower decides to default the mortgage, he also forfeits the opportunity to exercise the prepayment or default option in the future.

Pricing these options and also determining when a borrower exercises either option require specifying the underlying state variables and parameters that determine the value of the contract and then deducing the decision rule that maximizes borrower wealth. For residential mortgages, the key variables are the spot interest rate, r, and the value of the house, H. It is commonly assumed (e.g., Kau and Keenan 1995) that r and H are described respectively by the following stochastic processes:

$$dr = \gamma(\theta - r)\,dt + \sigma_r\sqrt{r}\,dz_r; \tag{1.1}$$

$$\frac{dH}{H} = (r - d)\,dt + \sigma_H\,dz_H; \tag{1.2}$$

$$dz_r dz_H = \rho(r, H, t)\,dt. \tag{1.3}$$

Here θ is the mean value of the interest rate. γ is the rate of convergence for the interest rate. d is the imputed rent payout ("dividend") rate. $\sigma_r\sqrt{r}$ and σ_H are the instantaneous standard deviations of the term structure and the house price, respectively. dz_r and dz_H are standard Wiener process with $E[dz] = 0$, and $E\left[dz^2\right] = dt$. ρ is the correlation between the disturbances to the term structure and the disturbances to the house price.

Under the perfect capital market assumption together with the Local Expectations Hypothesis, it has been shown (see Kau, Keenan, Muller, and Epperson 1995) that the value of the mortgage M satisfies

$$\begin{aligned} &\frac{1}{2} r\sigma_r^2 \frac{\partial^2 M}{\partial r^2} + \rho\sqrt{r} H \sigma_r \sigma_H \frac{\partial^2 M}{\partial r \partial H} + \frac{1}{2} H^2 \sigma_H^2 \frac{\partial^2 M}{\partial H^2} + \gamma(\theta - r)\frac{\partial M}{\partial r} \\ &\qquad + (r - d) H \frac{\partial M}{\partial H} + \frac{\partial M}{\partial t} - rM = 0. \end{aligned} \tag{1.4}$$

This follows almost directly from the model of Black and Scholes (1973). From equation (1.4) together with the appropriate boundary conditions, we can solve for the optimal values of the state variables r^* and H^*. This leads to a decision rule about mortgage termination: default when the house value falls to H^*; prepay when interest rate declines to r^*. Now, at each value of the house, H, the homeowner can compute the extent to which the default option is in the money, given the initial contract terms. Similarly, the owner can compute the extent to which the prepayment option is in the money at each interest rate, r.

Thus, the difference between the outstanding mortgage balance and H^* defines the extent to which the put option must be in the money for optimal default, and the difference between the mortgage coupon rate and r^* defines the extent to which the call option must be in the money for optimal prepayment.

Of course, this theory assumes that all observations on mortgage termination behavior are generated by rational, fully-informed mortgage holders who face zero transactions costs and have no other motives for prepayment or default. Clearly other "trigger events," such as job changes and unemployment affect the probability that a mortgage will be terminated. Together with transactions costs, reputation costs, and borrower's expectation for the future interest rates movement, they determine the extent to which the option must be in the money at exercise. The proportional hazard model provides a convenient framework for considering empirically the exercise of these options.

The Cox proportional hazard model (Cox and Oakes 1984) is defined as

$$h(t_{ij}; z) = h_0(t_{ij}) \exp(z(t_{ij})\beta); \tag{1.5}$$

where j denotes types of the competing risks, $z(t)$ is a set of time-varying covariates (which need not be limited to the option value itself; they may include other important determinants of behavior), and $h_0(\cdot)$ is the baseline hazard reflecting the age-related amortization feature of mortgage.

The most popular estimation approach for proportional hazard model is the Cox partial likelihood approach (CPL) (see Cox and Oakes 1984). However, CPL is appropriate for the proportional hazard model with competing risks only if these competing risks are independent. Furthermore, if the data collected are in discrete groups and if there are heavy ties in the discrete index of failure time, then CPL generates a biased estimator for the hazard rate (These limitations of CPL have been discussed by Cox and Oakes 1984, and Kalbfleisch and Prentice 1980).

A major concern in actually estimating hazard models with a large body of economic data is the computational difficulty involved. Computational time can become a real constraint when the model involves competing risks and time-varying covariates. To estimate a useful model for the housing market requires either dramatically limiting sample sizes, arbitrarily and unreasonably aggregating time intervals, or finding a way to aggregate observations on individual behavior.

Deng, Quigley, and Van Order 1995 introduced a semi-parametric estimation (SPE) approach[1] for the proportional hazard model with competing risks and time-varying covariates. To estimate the hazard model with SPE, first we partition the loan level mortgage data set into homogeneous cells based on certain individual characteristics of the loan, e.g., mortgage initial loan-to-value ratio, geographic location of the property, year of mortgage origination, etc. For each cell, we estimate the empirical default and prepayment hazard functions using the Kaplan-Meier approach. Then we map these estimated hazard functions to the original loan level mortgage data set. Finally, we regress the log estimated hazard functions on a set of covariates (e.g., option value covariates and other important determinants of borrower's behavior) and the fixed effects of mortgage duration[2].

The semi-parametric approach has following desirable features when compared to the Cox partial likelihood approach (CPL). First, the SPE can model the interdependent competing risks in a straightforward manner. Second, the SPE can be used to estimate the baseline simultaneously with the covariates. Third, ties in failure time are not a problem in the SPE. Fourth, it allows heterogeneous unobserved error terms to be incorporated into the competing risks hazard model. Last but not least, because the SPE transforms the proportional hazard function into a regression framework, it is far less demanding in computation.

This latter advantage should not be underestimated when dealing with a large sample size of duration data with time-varying covariates.

2. The Model

2.1. The Data

The data set used in the empirical analysis is the individual mortgage history data maintained by the Federal Home Loan Mortgage Corporation (Freddie Mac). This administrative data base contains 1,489,372 observations on single family mortgage loans issued between 1976 to 1983 and purchased by Freddie Mac. All are …xed-rate, level-payment, fully-amortized loans, most with thirty-year terms. The mortgage history period ends in …rst quarter of 1992. For each mortgage loan, the available information includes the year and month of origination and termination (if it has been closed), indicators of prepayment or default, the purchase price of the property, the original loan amount, the initial loan-to-value ratio, the mortgage contract interest rate, the monthly principal and interest payment, the state, the region and the major metropolitan area in which the property is located. The data set also reports the borrower's monthly gross income at origination. Table 1 describes the variables from the Freddie Mac data base used in this analysis.

[Table 1 is about here]

The market rate used in this analysis is the average interest rate charged by lenders on new …rst mortgages reported by Freddie Mac's quarterly market survey. This mortgage interest rate varies by quarter across five major US regions.

The analysis also uses a macro economic variable, unemployment rate, which is measured at the state level. State unemployment data are reported in various issues of: US Department of Labor, "Employment and Unemployment in States and Local Areas (Monthly)" and in the "Monthly Labor Review."

2.2. An Option-based Proportional Hazard Model for Mortgage Termination

The competing risks of mortgage termination consist of two parts: a prepayment risk and a default risk. The function specifying prepayment risk estimates the probability that a mortgage loan is prepaid during any period, conditional upon survival to that particular period. Similarly, the default function estimates the conditional probability of default during any period. The model assumes that borrowers make the prepayment or default decision based upon market conditions to maximize their net wealth. Following the contingent claims model discussed in section 1, the empirical model specifies the probability of exercising these options as a function of the extent to which the options are "in the money" and the "trigger events" that affect the decision about how far the option needs to be into the money in order for it to be optimal to exercise. For instance, an increase in the probability of negative equity will increase the probability that the put option is in the money, hence increase the probability of default. Analogously, the ratio of the present discounted market value of the unpaid balance to the par value of the mortgage measures the extent to which the call option is in the money. The variable, state level unemployment, is an example of trigger event.

The key variables are those measuring the extent to which the put and call options are in the money.

A typical way to value the call option in empirical real estate finance research is to compute the ratio of the present discounted value of the unpaid mortgage balance at the contract interest rate relative to the value discounted at the current market mortgage rate, assuming a deterministic term structure. In this analysis, I define such call option variable as $POPTION$.

Specifically, $POPTION$ for the lth loan observation is defined as

$$
\begin{aligned}
poption_l &= \frac{\displaystyle\sum_{t=1}^{term_l - \tau_i} \frac{mopipmt_l \times 3}{(1 + mktrate_{\omega_l;\,\psi_l+\tau_i}/400)^t} - \sum_{t=1}^{term_l - \tau_i} \frac{mopipmt_l \times 3}{(1 + noterate_l/400)^t}}{\displaystyle\sum_{t=1}^{term_l - \tau_i} \frac{mopipmt_l \times 3}{(1 + mktrate_{\omega_l;\,\psi_l+\tau_i}/400)^t}} \\
&= 1 - \frac{mktrate_{\omega_l;\,\psi_l+\tau_i} \times \left(1 - \left(\dfrac{1}{1 + noterate_l/400}\right)^{term_l - \tau_i}\right)}{noterate_l \times \left(1 - \left(\dfrac{1}{1 + mktrate_{\omega_l;\,\psi_l+\tau_i}/400}\right)^{term_l - \tau_i}\right)};
\end{aligned}
\tag{2.1}
$$

where τ_i is loan age measured in quarters, ω_l is a vector of indices for geographical location, ψ_l is loan origination time, $mopipmt_l$ is monthly principal and interest payment, $noterate_l$ is mortgage contract rate, $mktrate_{\omega_l;\,\psi_l+\tau_i}$ is the current local market mortgage rate, and $term_l$ is mortgage loan term calculated by

$$
term_l = \frac{\log\left(\dfrac{mopipmt_l - origamt_l \times (noterate_l/1200)}{mopipmt_l}\right)}{\log\left(\dfrac{1}{1 + noterate_l/1200}\right) \times 3};
\tag{2.2}
$$

where $origamt_l$ is original loan amount.

To value the put option analogously, we need to measure the market value of each house quarterly and to compute homeowner equity quarterly. Obviously, we do not observe the course of price variation for individual houses in the sample. In this analysis, I use the weighted repeat sales housing price index (WRS) estimated by Abraham and Schauman (1991). The WRS index provides estimates of the course of housing prices in each metropolitan area. It also provides an estimate of the variance in price for each house in the sample, by metropolitan area and elapsed time since purchase.

Based on estimates of the mean and variance of individual house prices, together with the unpaid mortgage balance (computed from the contract terms), I estimate the distribution of homeowner equity quarterly for each observation. In particular, "EQR" is the estimate of equity ratio assuming prices of all houses in the MSA grow at the mean rate, and "PNEQ" is the probability that equity is negative, i.e., the probability that the put option is in the money.

Specifically, equity ratio for the lth loan observation is defined as:

$$\begin{aligned}
eqr_l &= \frac{mktvalue_l - pdvunpblc_l}{mktvalue_l} \\
&= \frac{purprice_l \times \frac{msa_{\omega_l, \tau_l+\tau_i}}{msa_{\omega_l, \tau_l}} - \sum_{t=1}^{term_l - \tau_i} \frac{mopipmt_l \times 3}{(1 + noterate_l/400)^t}}{purprice_l \times \frac{msa_{\omega_l, \tau_l+\tau_i}}{msa_{\omega_l, \tau_l}}} \\
&= 1 - \frac{(LTV/100) \times \left(1 - \left(\frac{1}{1 + noterate_l/400}\right)^{term_l - \tau_i}\right)}{\left(\frac{msa_{\omega_l, \tau_l+\tau_i}}{msa_{\omega_l, \tau_l}}\right) \times \left(1 - \left(\frac{1}{1 + noterate_l/400}\right)^{term_l}\right)};
\end{aligned} \tag{2.3}$$

where $purprice_l$ is the purchasing price of the house at the time of loan initiation, $mktvalue_l$ is the estimated current market price of the house based on the mean value of WRS index in the MSA, and $pdvunpblc_l$ is the present discounted value of the remaining loan balance.

The probability of negative equity, pneq, is thus

$$pneq_l = ncdf\left(\frac{\log(pdvunpblc_l) - \log(mktvalue_l)}{\sqrt{e^2_{\omega_l, \tau_l+\tau_i}}}\right); \tag{2.4}$$

where $ncdf(\cdot)$ is cumulative standard normal distribution function, and $e^2_{\omega_l, \tau_l+\tau_i}$ is the variance of the WRS index.

To estimate the model with SPE, …rst I calculate the call-option and put-option covariates for each individual loan and construct the covariates matrix z_j (¿; ; !), where z_j (¿; ; !) consists of the call-option covariate, POPTION (or LATPOPTION[3]), the put-option covariate, PNEQ, and its variation, LOG(PNEQ), the initial loan-to-value ratio, the payment-to-income ratio at the origination, and the local unemployment rate. Then I map the covariates matrix z_j (¿; ; !) to the estimated "locally-smoothed" hazard function. Finally, I estimate the prepayment and default functions using the approach described in appendix A.

2.3. Evaluating the Prepayment Option with a Binomial Mean-Reverting Interest Rate Process

The prepayment option covariate, POPTION, de…ned in equation (2.1) was computed with a deterministic term structure. This is appropriate only if there are no transactions costs in the mortgage prepayment. However, prepaying a mortgage may hardly be a frictionless practice. For example, there may be points charged for new loans, there may be costs for title insurance, and there may be reappraisal fees, documentation fees, and escrow fees for re…nancing. Furthermore, there are intangible costs such as time costs involved in the re…nancing. Because the …xed rate mortgage has a remarkably lengthy term to maturity, it is inappropriate to assume that borrowers have perfect information about the interest rate term structure. Therefore, it is inappropriate to model the mortgage contingent claims with a deterministic term structure.

Black and Scholes (1973), and Cox, Ingersoll, and Ross (1985) introduced the mean-reverting interest rate di¤usion function with a Brownian motion process to model stock and bond prices. Since then, there have been several discrete-time models developed to approximate the stochastic interest rate term structure movement. Among those discrete-time models are Ho and Lee (1986), Nelson and Ramaswamy (1990), Hull and White (1990), and Tian (1992).

In this paper, I adopt Tian's Simplified Binomial Process (SBP)[4] to derive the prepayment option value. The SBP is a mean-reverting binomial process with upper and lower boundaries to prevent interest rates from having negative values. It is a path independent process and therefore has a significant computational advantage over other models. Tian (1992) also demonstrated that the SBP is superior to other discrete-time models in numerical accuracy in estimating the market interest rate movement.

Following Tian, a simple additive binomial lattice tree of the market interest rate can be constructed as follows:

Partition the time interval $[t_0; T]$ into n equal distance sub-intervals. The length of each sub-interval is $\Delta t = (T - t_0)/n$, e.g., if we define 120 sub-intervals for a 30-year fixed rate mortgage, then Δt is 0:25. Let $\varphi = \sqrt{r}$, where r is the current market rate. Let u and d be the respective distances of φ jumping upward and downward from one sub-interval to the next, p_{ij} be the probability that φ jumps upward at the ith interval and jth node of the lattice tree, and $(1 - p_{ij})$ be the probability that φ jumps downward at the ith interval and jth node of the lattice tree, where $i = 0; 1; ::: ; (n - 1)$, and $j = -i; -(i - 2); ::: ; (i - 2); i$.

Figure 1 illustrates such binomial lattice tree. It starts at t_0 when the observed loan is terminated. Let $\varphi_{0;0} = \sqrt{r_{0;0}}$, where $r_{0;0}$ is the current local market interest rate[5]. In the next sub-interval, φ may jump upward to $\varphi_{1;1}$ with the probability of $p_{0;0}$, or downward to $\varphi_{1;-1}$ with the probability of $(1 - p_{0;0})$, where $\varphi_{1;1} = \varphi_{0;0} + u_{0;0}$, and $\varphi_{1;-1} = \varphi_{0;0} - d_{0;0}$. In sequence, from $\varphi_{1;1}$, φ may jump upward to $\varphi_{2;2}$ with the probability of $p_{1;1}$, or downward to $\varphi_{2;0}$ with the probability of $(1 - p_{1;1})$, where $\varphi_{2;2} = \varphi_{1;1} + u_{1;1}$, and $\varphi_{2;0} = \varphi_{1;1} - d_{1;1}$. Similarly from $\varphi_{1;-1}$, φ may jump upward to $\varphi_{2;0}$ with the probability of $p_{1;-1}$, or downward to $\varphi_{2;-2}$ with the probability of $(1 - p_{1;-1})$, where $\varphi_{2;0} = \varphi_{1;-1} + u_{1;-1}$, and $\varphi_{2;-2} = \varphi_{1;-1} - d_{1;-1}$.

[Figure 1 is about here]

Note that $\varphi_{1,1} - d_{1,1} = \varphi_{1,-1} + u_{1,-1} = \varphi_{2,0}$. Note also that $\Delta\varphi$ and hence Δr is independent of time interval, i, and the node of the jump position, j. However, p_{ij} depends on both i and j, i.e., the probability of interest rate jumping up or down varies from each time interval and node of jump position.

Following Tian, we can derive the upper and lower boundaries of the lattice tree such that

$$r_{min} = \left(\sqrt{\frac{\sigma_r^2 (1 - \lambda \Delta t)}{4\lambda^2 \Delta t}} + \theta - \left(\frac{\sigma_r}{2\lambda \sqrt{\Delta t}} \right) \right)^2 ; \tag{2.5}$$

$$r_{max} = \left(\sqrt{\frac{\sigma_r^2 (1 - \lambda \Delta t)}{4\lambda^2 \Delta t}} + \theta + \left(\frac{\sigma_r}{2\lambda \sqrt{\Delta t}} \right) \right)^2 : \tag{2.6}$$

Therefore, whenever the interest rate reaches the upper boundary, the probability of the interest rate jumping up in the next time interval is zero. Similarly, whenever the interest rate reaches the lower boundary, the probability of the interest rate jumping up in the next time interval is one. Thus, the model eliminates the possibility of interest rates having negative values.

During the implementation process, depending on the length of Δt, φ_{ij} may just jump out of the upper boundary at certain nodes, e.g., in Figure 1, $\varphi_{4,4}$ is going to jump out of upper boundary. However, it can be shown that $p_{3,3}$ is very close to zero. Consequently, we simply set $p_{3,3}$ to be zero. Similarly, at those nodes where φ_{ij} jumps out of the lower boundary (such as $\varphi_{3,-3}$ and $\varphi_{5,-3}$ in Figure 1), we set $p_{i-1,j-1}$ to be one.

Evaluating the prepayment option for each individual loan with this binomial interest rate model is still a computationally intensive process. Instead of computing the POPTION assuming a deterministic term structure as described in equation (2.1), I calculate the ratio, LATPOPTION, using the binomial interest rate lattice tree according to the following procedure:

First, for each individual loan, I calculate the remaining terms of mortgage payment from the time the loan is terminated (or censored), e.g., for a loan prepaid at the 20th quarter, the remaining terms are 100 quarters.

Second, starting from the point when the decision of exercising options is to be made, i.e., the point when the observed individual loan is terminated (or censored), I compute forwardly the interest lattice tree for the remaining contract terms and the associated probabilities at each node of the tree such that

$$r_{ij} = \left(\sqrt{r_{00}} + \left(\frac{i \times \sigma_r \sqrt{\Delta t}}{2} \right) \right)^2 ; \tag{2.7}$$

$$p_{ij} = \frac{1}{2} + \frac{\sqrt{\Delta t}}{\sigma_r} \times \left(\frac{4\lambda\theta - \sigma_r^2}{8\sqrt{r_{ij}}} - \frac{\lambda\sqrt{r_{ij}}}{2} \right) ; \tag{2.8}$$

$$\begin{aligned} i &= 0; 1; 2; \ldots; (120 - \tau - 1) ; \\ j &= -i; -(i-2) ; \ldots; (i-2) ; i; \end{aligned}$$

where r_{00} is the current local market mortgage rate at age τ when the mortgage is terminated.[6]

Third, I calculate the value of the prepayment option backward based on the interest rate lattice tree and the associated probabilities[7] such that

$$voption_{ij} = 1 + p_{ij} \times \left(\frac{voption_{i+1;j+1}}{1 + r_{ij}} \right) + (1 - p_{ij}) \times \left(\frac{voption_{i+1;j-1}}{1 + r_{ij}} \right) ; \tag{2.9}$$

$$\begin{aligned} i &= 0; 1; 2; \ldots; (120 - \tau - 1) ; \\ j &= -i; -(i-2) ; \ldots; (i-2) ; i; \end{aligned}$$

and

$$voption_{120-\tau;j} = 1; \tag{2.10}$$

$$j = -(120 - \tau);\ -(120 - \tau - 2);\ \ldots;\ (120 - \tau - 2);\ (120 - \tau):$$

Finally, I compute the ratio of the prepayment option, LATPOPTION,[8] such that

$$\text{latpoption}_l = \frac{\text{voption}_{00} - \sum_{t=1}^{120-\tau} \frac{1}{(1 + \text{noterate}_l/400)^t}}{\text{voption}_{00}}: \tag{2.11}$$

3. The Empirical Analysis

The empirical analysis is based upon the Freddie Mac individual mortgage history data described in section 2.1. The analysis is confined to mortgage loans issued for owner occupancy, and includes only those loans which were either closed or still active[9] at the first quarter of 1992. In addition, the analysis is confined to loans issued in 30 major metropolitan areas (MSAs). The sub-sample contains a total of 489,372 observations. Loans are observed in each quarter from the quarter of origination through the quarter of termination, maturation, or through 1992:I for active loans.

Figures 2 and 3 summarize the raw data used in the empirical analysis. Figure 2 displays the conditional prepayment rate, separately by loan-to-value ratio (LTV), as a function of duration. Conditional prepayment rates are slightly higher for higher LTV loans. Rates increase substantially after the first fifteen quarters. Figure 3 displays raw conditional default rates by LTV. Again, default rates increase substantially after about fifteen quarters. Note also that the default rates increase dramatically with initial LTV. Default rates for loans with LTV above 95 percent are three or four times higher than default rates for 90 to 95 percent LTV loans. The default rates for these latter loans are, in turn, about five times as high as for those with LTV below 80 percent.

[Figure 2 is about here]

[Figure 3 is about here]

Finally, note that conditional default rates are quite low. Even for the riskiest class of loans, conditional default rates are no higher than four in a thousand in any quarter. Residential mortgages are relatively safe investments (and simple random samples of mortgages are likely to contain very few observations on default).

Tables 2 and 3 present a variety of models estimated by the SPE method. To estimate hazard functions non-parametrically, the entire sample of 489,372 loans has been partitioned into 120 cells, according to 30 major MSAs and 4 LTV groups. For each cell, there are 64 time intervals (measured in quarter, from 76:II to 92:I, and normalized by setting 76:II equals 1). Empirical hazard rates of prepayment and default have been calculated for every time interval in each cell based on the entire sample. Then the estimated empirical hazard rates were mapped to 9,183 mortgage loans which were randomly drawn from the total sample, assuming that the randomly-drawn sub-sample has the same distribution as the population. The mapping is based on the geographic location, the initial LTV, and the age of the observed individual loans.

Table 2 presents three alternative speci…cations using a deterministic term structure in the prepayment function. All three models specify the prepayment and default functions as a seemingly unrelated regression system. Besides the explicitly speci…ed interdependent competing risks in the model, the interdependence between default and prepayment behavior is also re‡ected by the correlations of unobserved error terms between the prepayment and default functions.

[Table 2 is about here]

The results show that …nancial motivation is of paramount importance in governing the prepayment and default behavior. For example, when the call option is in the money, the prepayment hazard increases very substantially. Similarly a higher probability of negative equity increases the default hazard and reduces the prepayment hazard. Note also that the probability of negative equity is highly signi…cant and negatively associated with the prepayment hazard, verifying the interdependence between prepayment and default behavior.

The initial loan-to-value ratio is positive and highly significant, particularly in the default function, across all specifications. This variable, known at the time when mortgages are initiated, may well reveal borrowers' risk preferences.

The results also show that higher unemployment probability is associated with higher default risks. However, it is also associated with lower prepayment rates — indicating that liquidity constraints (which make refinancing more difficult for unemployed households) keep them from exercising in-the-money call options.

Model 2 expands model 1 to include a variable measuring the initial payment-to-income ratio. This variable is negative and significant in the default function. A smaller mortgage payment relative to income generally indicates that housing is a smaller fraction of the borrower's investment portfolio. More sophisticated investors, such as these borrowers, are apparently more likely to behave in a ruthless fashion in the face of equity declines.

Model 3 imposes the constraint that as the probability of negative equity ratio approaches zero, then the probability of default also approaches zero.[10] The result is basically similar to that in model 2, except that unemployment becomes insignificant in the default function.

Models 4 to 6 in Table 3 are similar to models 1 to 3 in Table 2 except that the variable $LATPOPTION$ measures prepayment options using the binomial mean-reverting interest rate model described in section 2.3. All three models assume that the mean value of the interest rate process, μ, is ten percent; the rate of convergence, λ, is eight percent; the interest rate volatility, σ_r, is four percent; and the length of sub-interval, Δt, is 0.25. From equations (B.2) to (B.8), it can be shown that the probability of interest rate jumping up, p, approaches to one when interest rate, r, reaches 0.007633, and p approaches to zero when interest rate, r, jumps up to 1.1824. From equations (B.6) and (B.7), we can calculate that the jump size, $\Delta\varphi$, is 0.01.

[Table 3 is about here]

The results indicate that the prepayment behavior is less sensitive to the changing of estimated value of the prepayment option when such value is computed with a stochastic term structure. The negative correlation between prepayment and default functions becomes more signi…cant in models with stochastic term structure. To observe the di¤erence between these two models, it is useful to compare the simulated unconditional prepayment and default rates based on estimates from these two models.

Figures 4 and 5 illustrate the di¤erence between hazard rates estimated from the model using a stochastic interest rate term structure, and the hazard rates estimated from the model using a deterministic interest rate term structure.[11] The solid lines are the average cumulative unconditional default and prepayment rates simulated from the model where prepayment options are evaluated with a binomial mean-reverting interest rate model. The dotted lines are the average cumulative unconditional default and prepayment rates simulated from the model where prepayment options are evaluated with a deterministic term structure.

[Figure 4 is about here]

[Figure 5 is about here]

Figure 4 shows that the predicted cumulative prepayment rates are higher in the model using a stochastic interest rate term structure.[12] Although the variable measuring the prepayment option seems not playing a role in the default function, Figure 5 indicates that using the binomial mean-reverting interest rate model to compute the value of prepayment option a¤ects predicting default hazard rates as well. This fact also illustrates the interdependence between these two competing risks.[13]

4. Conclusion

This paper has presented a unified model of the contingent claims and competing risks of mortgage termination by prepayment and default. The model considers these two hazards as interdependent competing risks and estimates them jointly. The model is estimated using a semi-parametric estimation approach (SPE) introduced by Deng, Quigley, and Van Order (1995). As shown, the SPE has several important advantages over the more familiar Cox Partial Likelihood approach when applied to problems of this nature.

The computational advantage of the SPE permits the empirical model to incorporate a binomial mean-reverting interest rate model into the prepayment hazard function. Using the stochastic term structure enhances the empirical hazard model in analyzing mortgage holders' rational behavior.

The substantive results of the analysis provide powerful support for the contingent claims model which predicts the exercise of financial options. The financial value of the call option is strongly associated with exercise of the prepayment option, and the probability that the put option is in the money is strongly associated with exercise of the default option. The results illustrate that introducing volatility and uncertainty about future interest rate movement has effects not only on predicting mortgage prepayment behavior, but also on predicting default behavior. The results also provide strong support for the interdependence of the decisions to prepay and to default on mortgage obligations.

In addition, the results indicate that liquidity constraints play an important role in the exercise of options in the mortgage market. Ceteris paribus, mortgage holders who are at greater risk for unemployment (as measured by the unemployment rate in their state of residence) are less likely to exercise in-the-money prepayment options. Those who are more likely to have low levels of equity are also less likely to exercise prepayment options when it is in their financial interest to do so. All three of these results are explicable, not by option theory, but rather by liquidity constraints which arise from quali…cation rules typically enforced by lenders in mortgage re…nance.

Finally, the results suggest that, holding other things constant, those who have chosen high initial LTV loans are more likely to exercise options in the mortgage market – prepayment as well as default. Further, those whose income, wealth, or housing demands permit them to choose low initial payment-to-income levels seem consistently more likely to behave ruthlessly in the exercise of default options. It appears that these factors, known at the time mortgages are issued, also re‡ect investor preferences for risk and investor sophistication in the market for mortgages on owner-occupied housing.

Acknowledgments

This paper is based on part of my Ph.D. dissertation submitted to the University of California at Berkeley. The author is grateful to Professor John Quigley, Professor Daniel McFadden, and Professor Nancy Wallace for their valuable advice. The author also wishes to thank Tony Lancaster, Glenn Sueyoshi, the participants in the National Science Foundation Symposium on Duration and Hazard Models at University of California at Berkeley in August, 1995, and an anonymous reviewer for their comments and helpful suggestions. The …nancial support of the Center for Real Estate and Urban Economics, University of California at Berkeley is gratefully acknowledged.

Appendix

A. A Semi-parametric Estimator for The Proportional Hazard Model with Competing Risks and Time-varying Covariates

This section brie‡y discusses the semi-parametric estimator introduced by Deng, Quigley, and Van Order (1995).

Let T be a continuous random variable, measuring the duration of stay, e.g., the length of time since a mortgage was originated. If each individual enters the state at the same calendar time (i.e., all individuals take out mortgages on the same day), then there is no difference between duration and calendar time. However, in most situations, duration is not the same as calendar time.

Define

$$F(t) = \Pr(T \geq t) \tag{A.1}$$

as the survivor function. The probability density function of the random variable t is:

$$\begin{aligned} f(t) &= \lim_{\Delta t \to 0^+} \frac{\Pr(t \leq T < t + \Delta t)}{\Delta t} \\ &= \frac{-dF(t)}{dt}. \end{aligned} \tag{A.2}$$

Define a hazard function that specifies the instantaneous rate of failure at $T = t$ conditional upon survival to time T such that

$$\begin{aligned} h(t) &= \lim_{\Delta t \to 0^+} \frac{\Pr(t \leq T < t + \Delta t | T \geq t)}{\Delta t} \\ &= \frac{f(t)}{F(t)} \\ &= \frac{-d \ln F(t)}{dt}. \end{aligned} \tag{A.3}$$

The Cox proportional hazard model (Cox and Oakes 1984) is defined as

$$h(t_{ij}; z) = h_0(t_{ij}) \exp(z(t_{ij})\beta); \tag{A.4}$$

where j denotes types of the competing risks, $z(t)$ is a set of time-varying covariates, and $h_0(\cdot)$ is the baseline hazard.

For most empirical applications, duration data are collected in discrete form. Therefore, it is necessary to transform the above continuous model into a grouped discrete duration model.

Define $T \in R^+$ as a duration variable. Let T_i $(i = 1, 2, \ldots, q)$ be the discrete time intervals that partition the support of T. Let

$$h_j(t; Z) = h_{0j}(t) \left[\exp\left(Z_j(t)\beta_j \right) \eta_j \right] \qquad \text{(A.5)}$$

be the hazard rate of duration t, where $j = 1, 2, \ldots, J$ is the type of competing risk, $h_{0j}(\cdot)$ is a baseline hazard function, $\exp\left(Z_j(t)\beta_j \right)$ is a proportionality factor, and η_j is an error term with a non-negative distribution.

A log integrated hazard function for risk type j can be constructed:

$$\log \left[\int_{T_{i-1}}^{T_i} h_j(t; Z_j)\, dt \right] = Z_j(T_i)\beta_j + \gamma_j(T_i) + \varepsilon_j; \qquad \text{(A.6)}$$

where

$$\gamma_j(T_i) = \log \left[\int_{T_{i-1}}^{T_i} h_{0j}(t)\, dt \right]; \qquad \text{(A.7)}$$

and

$$\varepsilon_j = \log \eta_j;$$

$$j = 1, 2, \ldots, J; \quad i = 1, 2, \ldots, q;$$

given that $Z_j(t)$ is constant between T_{i-1} and T_i.

The left-hand side of equation (A.6) is typically not directly observable in micro data. We can, however, use the "local smoothing" technique, developed in the literature on non-parametric methods, to estimate individual hazard functions based on the empirical distribution of the hazard functions. Partition the covariate matrix Z into K distinct matrices $Z_1, \ldots, Z_K$. The kth subgroup contains M_k observations. $M_1 + M_2 + \ldots + M_K = N$, where N is the total sample size. For each subgroup, estimate the hazard rate such that $\hat{h}_{jkt} = \frac{n_{jkt}}{S_{kt}}$, where n_{jkt} is the number of individuals who fail in the tth period with type j in the kth subgroup, and S_{kt} is the total number of individuals surviving to the tth period in the kth subgroup.[14]

Now replacing the left-hand side of equation (A.6) with the smoothed log hazard function, $\log \int_{T_{i-1}}^{T_i} \hat{h}_{jk}(t; Z_{jk})\, dt$, yields

$$\log\left[\int_{T_{i-1}}^{T_i} \hat{h}_{jk}(t; Z_{jk})\, dt\right] = Z_{jk}(T_i)\beta_j + \gamma_j(T_i) + \varepsilon_j + u_{jk}(T_i); \qquad \text{(A.8)}$$

$$j = 1, 2, \ldots, J; \quad k = 1, 2, \ldots, K; \quad i = 1, 2, \ldots, q;$$

where $u_{jk}(T_i) = \log\left[\int_{T_{i-1}}^{T_i} \hat{h}_{jk}(t; Z_{jk})\, dt\right] - \log\left[\int_{T_{i-1}}^{T_i} h_{jk}(t; Z_{jk})\, dt\right]$.

The covariance of the ε captures the correlation among competing risks which is not explicitly specified in the model. Equation (A.8) is a seemingly unrelated regression system which can be estimated by the approach proposed by Zellner (1962).

Deng, Quigley, and Van Order (1995) have shown that the SPE is consistent and the rate of convergence for this non-parametric estimator is $N^{-(2/5)}$.

B. A Simplified Binomial Mean-Reverting Interest Rate Process

This section summarizes the Simplified Binomial Process (SBP) introduced by Tian (1992).

To evaluate the interest rate contingent claim in the mortgage prepayment, the instantaneous spot rate is assumed to contain all information about future interest rates and thus drives the entire term structure. Following Cox, Ingersoll, and Ross (1985), the short-term interest rate, r, is assumed to have the following stochastic process:

$$dr = \kappa(\mu - r)dt + \sigma_r \sqrt{r}\, dz; \qquad \text{(B.1)}$$

where θ is the mean value of interest rate, κ is the rate of convergence for the interest rate, $\sigma_r \sqrt{r}$ measures the volatility of interest rate, dz is a standard Wiener process with $E[dz] = 0$, and $E\left[dz^2\right] = dt$. Both κ and σ_r are positive constants.

The key to constructing the Tian simplified binomial process to approximate a mean-reverting diffusion process is path independence. In other words, the interest rate level must be the same after an upward move followed by a downward move as it is after a downward move followed by an upward move. Such a binomial interest rate lattice tree is computationally efficient since the number of nodes in the tree increases only linearly with the number of time steps.

To construct a path independent binomial lattice tree, it is necessary to transform the interest rate process in equation (B.1) into a form that has a constant volatility. A typical transformation for this process is simply

$$\phi = \sqrt{r}. \tag{B.2}$$

Thus, equation (B.1) becomes

$$d\phi = q\,dt + \nu\,dz, \tag{B.3}$$

where

$$\begin{aligned} q &= \kappa(\theta - r)\frac{\partial \phi}{\partial r} + \frac{\sigma^2 r}{2}\frac{\partial^2 \phi}{\partial r^2} \\ &= \frac{\alpha_1}{\phi} - \alpha_2 \phi, \end{aligned} \tag{B.4}$$

$$\alpha_1 = \frac{4\kappa\theta - \sigma_r^2}{8}, \qquad \alpha_2 = \frac{\kappa}{2}, \tag{B.5}$$

and

$$\nu = \frac{\sigma_r}{2}. \tag{B.6}$$

In general, for a mean-reverting path independent binomial process which converges to the process (B.3), it leads to a solution that

$$\triangle \varphi_{ij} = u_{ij} = d_{ij} = \nu \sqrt{\triangle t}; \qquad \text{(B.7)}$$

and

$$p_{ij} = \frac{1}{2} + \frac{q\left(\varphi_{ij}\right)\sqrt{\triangle t}}{2\nu}: \qquad \text{(B.8)}$$

Since p_{ij} is a probability, we must require that $0 \le p_{ij} \le 1$. This implies that

$$\left|q\left(\varphi_{ij}\right)\right| \le \frac{\nu}{\sqrt{\triangle t}}: \qquad \text{(B.9)}$$

Combining equations (B.2), (B.4), (B.5), (B.6), and (B.9), we can derive the upper and lower boundaries of the lattice tree such that

$$\varphi_{min} \le \varphi \le \varphi_{max}; \qquad \text{(B.10)}$$

where

$$\varphi_{min} = \left\lceil -\left(\frac{\sigma_r}{2\lambda\sqrt{\triangle t}}\right) + \sqrt{\frac{\sigma_r^2(1-\lambda\cdot t)}{4\lambda^2\cdot t}} + \theta \right\rceil; \qquad \text{(B.11)}$$

$$\varphi_{max} = \left\lfloor \left(\frac{\sigma_r}{2\lambda\sqrt{\triangle t}}\right) + \sqrt{\frac{\sigma_r^2(1-\lambda\cdot t)}{4\lambda^2\cdot t}} + \theta \right\rfloor: \qquad \text{(B.12)}$$

Notes

[1]Appendix A presents Deng, Quigley, and Van Order's estimation approach brie‡y.

[2]As shown in Appendix A, these …xed e¤ects are transformed baseline hazard functions.

[3]The de…nition and computation of the $LATPOPTION$ will be discussed in section 2.3.

[4]Appendix B summarizes Tian's SBP approach.

[5]In this analysis, I use the regional average interest rate charged by lenders on new …rst mortgages reported by Freddie Mac's quarterly market survey as the relevant current market rate.

[6]In section 2.2, I used $¿_i$ to denote the age of the ith individual loan. I drop the loan record index i here to avoid confusion with the index i which is used here to denote interest rate jump steps.

[7]During the implementation process, I set up a check point to avoid redundant computation at those branches where the associated probabilities are zero.

[8]Note that SPE only requires evaluating $LATPOPTION$ at the point when the loan is terminated or censored, while CPL requires computing such option value covariates and other time-varying covariates for the whole path of the loan history. That makes the computation impossible if we use CPL to estimate this model.

[9]It excludes those observations which were in delinquency or foreclosure at the time data were collected.

[10]LOG$PNEQ$ is the logarithm of $PNEQ$ and scaled by a factor of 0.01. From equation (1.5), we can see that the default hazard function is now a product of baseline hazard function, $PNEQ$, and an exponential function of other covariates. Therefore, with this specifwhen $PNEQ$ approaches

zero, the probability of default approaches zero too.

[11]The simulations are based on estimates from model 1 in Table 2, and model 4 in Table 3. I use Freddie Mac's interest rate model to generate 300 random paths of mortgage interest rates. The mean value of the house price inflation rate is set at ten percent annually and the mean level of the mortgage contract rate is set at ten percent. The mean value of unemployment rate is set at eight percent. Initial loan-to-value ratio is 80 percent.

[12]Note that results may vary depending on the assumption about the interest rate term structure model, i.e., the assumption of mean level of the interest rate, the rate of convergence toward mean, the volatility of the interest rate movement, and the length of the sub-interval.

[13]These results also persisted in a variety of other specifications not reported here. In general, the predicted cumulative prepayment rates at the end of year 15 is about nine per cent lower in models with a deterministic term structure than that in models with a stochastic term structure. On the other hand, the predicted cumulative default rates at the end of year 15 is about 16 per cent higher in models with a deterministic term structure than that in models with a stochastic term structure.

[14]Note the risk set of the conditional hazard rates includes not only the individuals that have the same failure type and whose durations are greater than the current one, but also all those individuals with a different failure type and whose durations are greater than the current one. Furthermore, the risk set also includes those right-censored observations with durations greater than the current one.

References

Abraham, J. M. and W. S. Schauman (1991). New evidence on home prices from freddie mac repeat sales. Journal of The American Real Estate and Urban Economics Association 19 (3), 333–352.

Black, F. and M. S. Scholes (1973). The pricing of options and corporate liabilities. Journal of Political Economy 81, 637–654.

Brennan, M. J. and E. S. Schwartz (1985). Determinants of GNMA mortgage prices. Journal of The American Real Estate and Urban Economics Association 13, 209–228.

Buser, S. A. and P. H. Hendershott (1984). Pricing default free mortgages. Housing Finance Review 3, 405–429.

Cox, D. R. and D. Oakes (1984). Analysis of Survival Data. Monographs on Statistics and Applied Probability. London: Chapman and Hall.

Cox, J. C., J. E. Ingersoll, and S. A. Ross (1985). A theory of the term structure of interest rates. Econometrica 53, 385–467.

Deng, Y., J. M. Quigley, and R. Van Order (1995). Mortgage terminations. Center for Real Estate Finance and Urban Economics Working Paper 95-230, University of California at Berkeley.

Dunn, K. B. and J. J. McConnell (1981). Valuation of mortgage-backed securities. The Journal of Finance 36, 599–617.

Green, J. and J. B. Shoven (1986). The e¤ect of interest rates on mortgage prepayments. Journal of Money, Credit and Banking 18, 41–50.

Harding, J. P. (1994). Rational mortgage valuation using optimal intertemporal re…nancing strategies and heterogeneous borrowers. Ph.D. Dissertation, Haas School of Business, University of California at Berkeley.

Hendershott, P. H. and R. Van Order (1987). Pricing mortgages: An interpretation of the models and results. Journal of Financial Services Research 1, 77–111.

Ho, T. S. Y. and S. B. Lee (1986). Term structure movements and pricing of interest rate claims. The Journal of Finance 41, 1011–1029.

Hull, J. and A. White (1990). Pricing interest-rate derivative securities. Review of Financial Studies 3 (4), 573–592.

Kalbfleisch, J. D. and R. L. Prentice (1980). The Statistical Analysis of Failure Time Data. Wiley Series in Probability and Mathematical Statistics. New York: John Wiley and Sons.

Kau, J. B. and D. C. Keenan (1995). An overview of the option-theoretic pricing of mortgages. Journal of Housing Research 6 (2), 217–244.

Kau, J. B., D. C. Keenan, W. J. Muller, and J. F. Epperson (1992). A generalized valuation model for fixed-rate residential mortgages. Journal of Money, Credit and Banking 24 (3), 279–299.

Kau, J. B., D. C. Keenan, W. J. Muller, and J. F. 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.

Nelson, D. B. and K. Ramaswamy (1990). Simple binomial processes as diffusion approximations in financial models. Review of Financial Studies 3 (3), 393–430.

Quigley, J. M. and R. Van Order (1990). Efficiency in the mortgage market: The borrower's perspective. Journal of The American Real Estate and Urban Economics Association 18 (3), 237–252.

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

Schwartz, E. S. and W. N. Torous (1989). Prepayment and the valuation of mortgage-backed securities. The Journal of Finance 44 (2), 375–392.

Tian, Y. (1992). A simpli…ed binomial approach to the pricing of interest-rate contingent claims. Faculty of Administrative Studies Working Paper, York University, North York, Ontario, Canada, M3J 1P3.

Zellner, A. (1962). An e¢ cient method of estimating seemingly unrelated regressions and tests for aggregation bias. Journal of The American Statistical Association 57, 348–368.