Title

Mortgage Terminations

Authors
Deng, Yongheng
Quigley, John M.
Van Order, Robert

Publication Date
1995-04-01

Peer reviewed

Institute of Business and Economic Research
University of California at Berkeley

FISHER CENTER FOR REAL ESTATE AND URBAN ECONOMICS

WORKING PAPER SERIES

WORKING PAPER NO. 95-230

MORTGAGE TERMINATIONS

By

YONGHENG DENG
JOHN M. QUIGLEY
ROBET VAN ORDER

WALTER A. HAAS SCHOOL OF BUSINESS

FISHER CENTER FOR REAL ESTATE AND URBAN ECONOMICS
UNIVERSITY OF CALIFORNIA AT BERKELEY
Kenneth T. Rosen, Chair
Robert H. Edelstein, Co-Chair
Dwight M. Jaffee, Co-Chair

The Center was established in 1950 to examine in depth a series of major changes and issues involving urban land and real estate markets. The Center is supported by both private contributions from industry sources and by appropriations allocated from the Real Estate Education and Research Fund of the State of California.

INSTITUTE OF BUSINESS AND ECONOMIC RESEARCH
Carl Shapiro, Director

The Institute of Business and Economic Research is an organized research unit of the University of California at Berkeley. It exists to promote research in business and economics by University faculty. These working papers are issued to disseminate research results to other scholars. The authors welcome comments; inquiries may be directed to the author in care of the Center.

Mortgage Terminations*

Yongheng Deng
University of California, Berkeley

John M. Quigley
University of California, Berkeley

Robert Van Order
Freddie Mac

April 1995

WORKING PAPER NO. 95-230

* A previous version of this paper was presented at the NBER Summer Institute, Cambridge MA, August 1994. Financial support for this research has been provided by the Center for Real Estate and Urban Economics, University of California, Berkeley.

Abstract

Pricing mortgage contracts is complicated by the behavior of homeowners who may choose to exercise their options to default or to prepay. These options are distinct, but not independent. In this paper, we present and estimate a unified economic model of the competing risks of mortgage termination by prepayment and default. We adopt a proportional hazards framework to analyze these competing and dependent risks in a model with time-varying covariates applied to a large sample of individual loans. The empirical model is computationally feasible only through the development of an alternative estimation technique based on semi-parametric methods (SPE). The SPE has several other advantages over more familiar approaches when applied to this class of problem.

The substantive results of the analysis provide powerful support for the contingent claims model which predicts the exercise of financial options. The results also provide strong support for the interdependence of the decisions to prepay and to default on mortgage obligations. Furthermore, the results indicate that liquidity constraints, investor preferences for risk, and investor sophistication also play important roles in the exercise of options in the mortgage market.

1. Introduction

The mortgage market is quite large and is increasing in importance. The outstanding volume of residential mortgages is currently over $ 3 trillion, and volume has doubled in the past decade. In comparison, the stock of outstanding U.S. government debt is currently about $ 5 trillion. Almost half of the stock of mortgages is held in "mortgage-backed securities," and about two-thirds of new mortgages are "securitized." The rise of securitization, the trading of these securities, and the growing use of mortgage securities as collateral for "derivatives" (*e.g.*, collateralized mortgage obligations) has generated a great deal of interest in the pricing of mortgages and mortgage-backed securities.

Pricing mortgage contracts is complicated, primarily by the options available to the borrower to default or to prepay. These option are distinct, but not independent. Thus, one cannot accurately calculate the economic value of the default option without considering the financial incentive for prepayment. Appropriately valuing these prepayment and default risks is crucial to the pricing of mortgages and to understanding the economic behavior of homeowners.

The contingent claims models, developed by Black and Scholes [4], Cox, Ingersoll, and Ross [10], and others, provide a coherent motivation for borrower behavior, and a number of studies have applied this model to the mortgage market.[1]

All empirical applications, however, ignore the interdependence between the prepayment and default options. Indeed, all these empirical analyses, typically based upon hazard models, focus on the exercise of a single option. In early work, for example, Schwartz and Torous [29] analyzed mortgage prepayment by employing a model with fixed covariates and assuming that mortgages were simply free of default risk. More general models using time-varying covariates by Green and Shoven [13] and Quigley and Van Order [25] made analogous assumptions in the analysis of prepayment behavior. Quigley and Van Order [26] studied default behavior using the model of a single hazard and assuming prepayments followed some given pattern.

In this paper we present a unified economic model of the competing risks of mortgage termination by prepayment and default. We adopt a proportional hazard framework to analyze these competing and dependent risks in a model with time-varying covariates applied to a large sample of individual loans. The empirical model is computationally feasible only through the development of an alternative estimation technique based on semi-parametric methods. In the em-

[1]See, for example, Dunn and McConnell [11], Buser and Hendershott [6], Brennan and Schwartz [5], Kau *et al* [17][18], Quigley and Van Order [26], Hendershott and Van Order [15], and Kau and Keenan [19] provide surveys of these results.

pirical analysis below, we derive the semi-parametric estimator (SPE) for the proportional hazard model with competing risks and compare it with the Cox partial likelihood estimator (CPL).

The properties of the CPL estimator are well known, as are the computational difficulties, at least in principle. The SPE estimator derived below offers several important advantages in the application of the model to the mortgage market.

The paper is organized as follows: section 2 reviews briefly the contingent claims model and the proportional hazard model. It also clarifies some econometric issues when applying the proportional hazard model to the study of mortgage terminations. Section 3 derives the semi-parametric approach for the proportional hazard model, specified with competing risks and time-varying covariates. Section 4 presents an extensive empirical analysis. Section 5 is a brief conclusion.

2. Mortgage Terminations and Competing Risks

Well-informed borrowers in a perfectly competitive market will exercise financial 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 refinancing.

The problem of pricing these options and also of determining when to exercise either option requires specifying the underlying state variables and parameters that determine the value of the contract and then deducing the rule for exercise that maximizes borrower wealth. For residential mortgages, the key state variables are interest rates and house values. The value of a mortgage $M(c, r, H, \tau, \Gamma)$ depends upon the coupon rate, c, a vector of relevant interest rates, r, property value, H, the age of the mortgage, τ, and the remaining time to maturity, Γ. With continuous time, a standard arbitrage argument is sufficient to derive an equilibrium condition for M (a second order partial differential equation) such that the value of the mortgage equals the risk-adjusted expected present value of its net cash flows.

Assume that house price changes are continuous with an instantaneous mean μ and a standard deviation σ_h. Let d be the imputed rent payout ("dividend") rate. For simplicity, assume there is only one interest rate, the instantaneous short rate r, which determines the yield curve. Let θ be the mean value of the short rate, γ be the rate of convergence for the short rate, σ_r be the volatility of the short rate, and ρ be the correlation between interest rate changes and house price changes. Then it has been shown (Kau *et al* [18]) that the value of the mortgage

M satisfies

$$\tfrac{1}{2} r\sigma_r^2 \frac{\partial^2 M}{\partial r^2} + \rho\sqrt{r} H \sigma_r \sigma_h \frac{\partial^2 M}{\partial r \partial H} + \tfrac{1}{2} H^2 \sigma_h^2 \frac{\partial^2 M}{\partial H^2} + \gamma(\theta - r)\frac{\partial M}{\partial r} + (r-d) H \frac{\partial M}{\partial H} + \frac{\partial M}{\partial \tau} - rM = 0. \tag{2.1}$$

This follows almost directly from the model of Black and Scholes [4]. From equation (2.1) together with the appropriate boundary conditions, one 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 rates decline to r^*, where both r^* and H^* depend upon the same parameters (*i.e.*, c, τ and Γ). The borrower does not have to solve (2.1) and the boundary conditions in order to know when to exercise either option. For instance, the prepayment option should be exercised whenever the borrower can refinance the loan at par at an interest rate less than the coupon on the loan. The decision to default depends simply on the value of the house compared to the market value of the mortgage.

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, complete capital markets, and have no other motives for prepayment or default. If this is not true, then other "trigger events," such as job changes, unemployment, or household dissolution can affect the probability that a mortgage will be terminated. Hence, we should not expect the frictionless model to hold exactly[2], and we should not expect all consumers to exercise options at the same values of r^* and H^*. We can, however, estimate the extent to which the option is "in the money"[3] and model the probability of exercise as a function of the extent to which it is financially profitable. The proportional hazard model provides a convenient framework for considering empirically the probability of exercise of these options and the importance of other trigger events.

Let T be a continuous random variable which measures the duration of stay, *i.e.*, the length of time since a mortgage was originated. If each individual enters the state at the same calendar time (*i.e.,* all take out mortgages on the same day), then there is no difference between duration and calendar time. In general, however duration is not the same as calendar time.

Define

$$F(t) = \Pr(T \geq t) \tag{2.2}$$

[2]In any event, we cannot observe either the market value of an old mortgage (due to the changing of its riskiness and term during the mortgage contract period) or (for the same reason) the coupon at which it could be refinanced at par.

[3]For the default option, this is measured by the house price minus mortgage balance. For prepayment, it is measured by the present value of remaining payments discounted at the current coupon rate minus the remaining balance.

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

$$\begin{aligned} f(t) &= \lim_{\Delta t \to 0^-} \frac{\Pr(t \leq T < t + \Delta t)}{\Delta t} \\ &= \frac{-dF(t)}{dt}. \end{aligned} \tag{2.3}$$

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

$$\begin{aligned} h(t) &= \lim_{\Delta t \to 0^+} \frac{\Pr(t \leq T < t + \Delta t | T \geq t)}{\Delta t} \\ &= \frac{f(t)}{F(t)} \\ &= \frac{-d \ln F(t)}{dt}. \end{aligned} \tag{2.4}$$

The relationships among the density function $f(t)$, the survivor function $F(t)$, and the hazard function $h(t)$ are thus:

$$F(t) = \exp\left(-\int_0^t h(u)\, du\right), \tag{2.5}$$

$$f(t) = h(t) \exp\left(-\int_0^t h(u)\, du\right). \tag{2.6}$$

2.1. The Cox Partial Likelihood Estimator

In the Cox proportional hazard model [9], the hazard function is defined as

$$h(t_{ij}; z) = h_0(t_{ij}) \exp(z(t_{ij})\beta), \tag{2.7}$$

where j denotes failure types, $z(t)$ is a set of time-varying covariates (for example, measures of the extent to which the default or prepayment option is in the money) as well as some trigger events. $h_0(\cdot)$ is the hazard for an individual under baseline conditions, *i.e.*, $z = 0$.

The partial likelihood function for the independent competing risks model with time-varying covariates and censoring, provided by Cox [9], is

$$L_P(\beta_1, ..., \beta_J) = \prod_{j=1}^{J} \prod_{i=1}^{n_j} \left(\frac{\exp[z_{ij}(t_{ij})\beta_j]}{\sum_{l \in R(t_{ij})} \exp[z_l(t_{ij})\beta_j]} \right), \tag{2.8}$$

where $j = 1, ..., J$ denotes failure types, n_j is the number of failure time observations of the jth failure type, and $R(\cdot)$ is the risk set such that, $R\left(t_{(ij)}\right) = \left\{(l) : T_{(l)} \geq t_{ij}\right\}$.

This partial likelihood function is derived from the rank statistic of a data set containing observations on durations. It is not a full likelihood function, and the derivation of the function does not depend on the error distribution of the duration variables. With a Cox partial likelihood specification, one cannot specify the correlation among competing risks through the error terms.

Note, further, that the partial likelihood function defined above does not allow interaction among competing risks through its functional form. This is because the log partial likelihood function has an additive form such that

$$\log L_P(\beta_1, ..., \beta_J) = F_1(\beta_1) + ... + F_J(\beta_J). \qquad (2.9)$$

Thus the Hessian is assumed to be block diagonal, and statistical inferences about the β's estimated simultaneously will be the same as inferences obtained from separate maximum likelihood estimations.

Second, it is worth noting that if the data collected are in discrete groups, and if there are heavy ties in the discrete index of failure time, then the Cox partial likelihood estimation generates a biased estimator for the hazard rate (This is discussed by Cox [9]).

Finally, as noted above, the Cox partial likelihood estimation concentrates the baseline hazard function out of its estimation procedure. In order to estimate the shape of the baseline hazard, one may either arbitrarily assume some functional form for the baseline hazard,[4] or use a two step semi-parametric approach to estimate the baseline functions, for example following Kaplan-Meier.[5]

3. A Semi-parametric Estimator

Prentice and Gloeckler [24] and Meyer [23] (PGM, for short) suggested a maximum likelihood estimation approach for the proportional hazard model with grouped duration variables. The PGM approach integrated the baseline hazard function into grouped intervals so that the baseline could be estimated simultaneously with the coefficients of the covariates in a semi-parametric way. Compared to the Cox partial likelihood approach, the PGM approach is a consistent estimator even with many ties in the discrete duration variables. In addition, the PGM method allows the proportional hazard model to incorporate heterogeneous error terms with the assumption that such unobserved error terms follow a Gamma distribution. However, the PGM method is valid only for the proportional hazard model with a single risk.

As we document below, a major concern in applying the competing risks hazard model to analyze mortgage terminations is the computational requirements.

[4] Schwartz and Torous [29] assumed a log logistic function form for the baseline hazard for their prepayment hazard model.

[5] See Quigley, Van Order, and Deng[27], Appendix A.2.

More generally, even with modern technology, computational time can be a constraint when the model involves time-varying covariates. Ryu [28] suggested a minimum chi-square approach (MCS) to estimate the proportional hazard model with a single hazard and categorical time-invariant covariates. Ryu's MCS transforms the proportional hazard model into a regression-based model: thus it reduces computational time significantly. However, with continuous time-varying covariates, such as option values, it is hardly feasible to group them into categories. Furthermore, Ryu's MCS is based on a chi-square statistic derived from the survivor function. With competing risks, the survivor function cannot identify the specific risk it has survived. Thus, the MCS cannot be applied directly, but the idea behind the MCS can be generalized.

In this section, we derive a semi-parametric estimator (SPE) for the proportional hazard model with competing risks and time-varying covariates.

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

$$h_j\left(t, Z\right) = h_{0j}\left(t\right)\left[\exp\left(Z_j\left(t\right)\beta_j\right)\right]\eta_j \tag{3.1}$$

be the hazard rate of duration t, where $j = 1, 2, ..., J$ is the type of competing risk, $h_{0j}(\cdot)$ is a baseline hazard function, $\exp(Z_j(t)\beta_j)$ 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:

$$\log\left[\int_{T_{i-1}}^{T_i} h_j(t, Z_j)\, dt\right] = Z_j(T_i)\beta_j + \gamma_j(T_i) + \varepsilon_j, \tag{3.2}$$

where

$$\gamma_j(T_i) = \log\left[\int_{T_{i-1}}^{T_i} h_{0j}(t)\, dt\right], \tag{3.3}$$

and

$$\varepsilon_j = \log \eta_j,$$

$$j = 1, 2, ..., J, \quad i = 1, 2, ..., q,$$

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

The left-hand side of equation (3.2) is 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 $Z_1, ..., Z_K$. The kth subgroup contains M_k observations. $M_1 + M_2 + ... + M_K = N$, where N is the total sample size. For each subgroup, estimate the hazard rate such that $\widehat{h}_{jkt} = \frac{n_{jkt}}{S_{kt}}$, 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.[6]

Selection of the value of the smoothing parameter is analogous to parameter selection in non-parametric estimation. M_k is chosen to reduce the noise as well as to keep the approximation error (bias) low. Noise will be reduced by letting M_k approach infinity as a function of the sample size. Approximation error will be eliminated if the neighborhood around z shrinks asymptotically to zero. Unfortunately, these prescriptions conflict. A standard proposition in the non-parametric literature[7] suggests that as $M_k \to \infty$, $\frac{M_k}{N} \to 0$, and $N \to \infty$, a balance between these two goals can be achieved in an asymptotic sense by setting $M_k \sim N^{4/5}$. A consequence is that the mean squared error itself converges to zero at a rate of $M_k^{-1} \sim N^{-(4/5)}$. In other words, the rate of convergence for this non-parametric estimation is $N^{-(2/5)}$.

Now replacing the left-hand side of equation (3.2) with the smoothed log haz-

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

[7]See Härdle [14].

ard function, $\log \int_{T_{i-1}}^{T_i} \widehat{h}_{jk}(t, Z_{jk})\,dt$, yields

$$\log\left[\int_{T_{i-1}}^{T_i} \widehat{h}_{jk}(t, Z_{jk})\,dt\right] = Z_{jk}(T_i)\,\beta_j + \gamma_j(T_i) + \varepsilon_j + u_{jk}(T_i), \qquad (3.4)$$

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

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

It is easy to show that the coefficient vector $\widehat{\beta}$ estimated from equation (3.4) is consistent. To simplify, consider the case with one period and one risk type. Denote $y = \log\left[\int_{T_{i-1}}^{T_i} h_{jk}(t, Z_{jk})\,dt\right]$, and $\widehat{y} = \log\left[\int_{T_{i-1}}^{T_i} \widehat{h}_{jk}(t, Z_{jk})\,dt\right]$, $M = \max(M_1, M_2, ..., M_K)$. The least squares estimate of β is

$$\widehat{\beta} = \frac{\sum_{k=1}^{K}\sum_{i=1}^{M_k}(z_i - \overline{z})\,\widehat{y}_k}{\sum_{k=1}^{K}\sum_{i=1}^{M_k}(z_i - \overline{z})^2} = \frac{\sum_{k=1}^{K}\sum_{i=1}^{M_k}(z_i - \overline{z})\,(y_i + u_{ik})}{\sum_{k=1}^{K}\sum_{i=1}^{M_k}(z_i - \overline{z})^2}$$

$$= \frac{\sum_{k=1}^{K}\sum_{i=1}^{M_k}(z_i - \overline{z})\,y_i}{\sum_{k=1}^{K}\sum_{i=1}^{M_k}(z_i - \overline{z})^2} + \frac{\frac{1}{M}\sum_{k=1}^{K}\sum_{i=1}^{M_k}(z_i - \overline{z})\,u_{ik}}{\frac{1}{M}\sum_{k=1}^{K}\sum_{i=1}^{M_k}(z_i - \overline{z})^2} \qquad (3.5)$$

$$= B + C$$

It is clear that B is the standard least squares estimator of β, and $\lim_{M\to\infty} C = 0$. Thus $\widehat{\beta}$ is a consistent estimator of β. This result generalizes in an obvious way to the multivariate case.

The covariance of the ε_j's captures the correlation among competing risks. Equation (3.4) is a seemingly unrelated regression system which can be estimated by the approach proposed by Zellner [30].

Note that the log hazard transformation of the proportional hazard model defined by equations (3.4) also permits us to specify the competing risks as endogenous in the log hazard regression system. For example,

$$\log\left[\int_{T_{i-1}}^{T_i} \widehat{h}_{jk}\left(t, Z_{jk}\right) dt\right] = \widehat{H}_{\bar{j}k}\left(T_i, Z_{\bar{j}k}\right)\alpha_j + Z_{jk}\left(T_i\right)\beta_j + \gamma_j\left(T_i\right) + u_{jk}\left(T_i\right),$$
$$j = 1, 2, ..., J, \quad k = 1, 2, ..., K, \quad i = 1, 2, ..., q. \tag{3.6}$$

where $\widehat{H}_{\bar{j}k}\left(T_i, Z_{\bar{j}k}\right) = \log\left[\int_{T_{i-1}}^{T_i} \widehat{h}_{\bar{j}k}\left(t, Z_{\bar{j}k}\right) dt\right]$, and $\bar{j}$ denotes all failure types other than j. Since $\widehat{H}_{\bar{j}k}\left(T_i, Z_{\bar{j}k}\right)$ is endogenous, this is a linear simultaneous equations system, and standard techniques can be applied.

The semi-parametric approach has several desirable features when compared to the Cox partial likelihood approach (CPL). First, the SPE can model the dependent 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.

4. The Empirical Analysis

The empirical analysis is based upon individual mortgage history data maintained by the Federal Home Loan Mortgage Corporation (Freddie Mac). The data base contains 1,489,372 observations on single family mortgage loans issued between 1976 to 1983 and purchased by Freddie Mac. All are fixed-rate, level-payment, fully-amortized loans, most of them with thirty-year terms. The mortgage history period ends in first 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. For the mortgage default and prepayment model. censored observations include all matured loans as well as the loans active at the end of the period.

The analysis is confined to mortgage loans issued for owner occupancy, and includes only those loans which were either closed or still active[8] at the first quarter of 1992. The analysis is confined to loans issued in 30 major metropolitan areas (MSAs) — a total of 780,443 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.

The key variables are those measuring the extent to which the put and call options are in the money. To value the call option, the current mortgage interest rate and the initial contract terms are sufficient. We compute a variable "*POPTION*" measuring the ratio of the present discounted value of the unpaid mortgage balance at the current quarterly mortgage interest rate[9] relative to the

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

[9]The rate used is the average interest rate charged by lenders on new first mortgages reported by Freddie Mac's quarterly market survey. This mortgage interest rate varies by quarter across five major US regions.

value discounted at the contract interest rate.[10]

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. We do, however, have access to a large sample of repeat (or paired) sales of single family houses in these 30 metropolitan areas (MSAs). This information is sufficient to estimate rather precisely a weighted repeat sales housing price index (WRS) separately for each of the 30 MSAs. The WRS index provides estimates of the course of housing prices in each metropolitan area. It also provides an

---

[10]Specifically, POPTION for the lth loan observation is defined as

$$\begin{aligned} poption_l &= \frac{\sum_{t=1}^{term_l-\tau_i} \frac{mopipmt_l \times 3}{(1+mktrate_{\omega_l,\kappa_l+\tau_i}/400)^t} - \sum_{t=1}^{term_l-\tau_i} \frac{mopipmt_l \times 3}{(1+noterate_l/400)^t}}{\sum_{t=1}^{term_l-\tau_i} \frac{mopipmt_l \times 3}{(1+mktrate_{\omega_l,\kappa_l+\tau_i}/400)^t}} \\ &= 1 - \frac{mktrate_{\omega_l,\kappa_l+\tau_i} \times \left(1 - \left(\frac{1}{1+noterate_l/400}\right)^{term_l-\tau_i}\right)}{noterate_l \times \left(1 - \left(\frac{1}{1+mktrate_{\omega_l,\kappa_l+\tau_i}/400}\right)^{term_l-\tau_i}\right)}, \end{aligned} \tag{4.1}$$

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

$$term_l = \frac{\log\left(\frac{mopipmt_l}{origamt_l \times (noterate_l/1200) + mopipmt_l}\right)}{\log(1+noterate_l/1200) \times 3}, \tag{4.2}$$

where $origamt_l$ is original loan amount.

estimate of the variance in price for each house in the sample, by metropolitan area and elapsed time since purchase.[11]

Estimates of the mean and variance of individual house prices, together with

[11] Housing price indices and their volatilities are estimated according to the three stage procedure suggested by Case and Shiller [7] and modified by Quigley and Van Order[26]. The model assumes that log price for ith house at time t is given by

$$P_{it} = I_t + H_{it} + N_{it} \tag{4.3}$$

where I_t is the logarithm of the regional housing price level, H_{it} is a Gaussian random walk, such that,

$$\begin{aligned} E\left[H_{i,t+\tau} - H_{it}\right] &= 0, \\ E\left[H_{i,t+\tau} - H_{it}\right]^2 &= \tau\sigma_{\eta_1}^2 + \tau^2\sigma_{\eta_2}^2; \end{aligned}$$

and N_{it} is white noise, such that,

$$\begin{aligned} E\left[N_{it}\right] &= 0, \\ E\left[N_{it}\right]^2 &= \tfrac{1}{2}\sigma_\nu^2. \end{aligned}$$

The model is estimated on paired sales of owner occupied housing. In the first stage, the log price of the second sale minus the log price of the first sale is regressed on a set of dummy variables, one for each time period in the sample except the first period. The dummy variables have values of zero in every quarter except the quarter in which the sales occurred. For the quarter of the first sale, the dummy is -1, and for the quarter of the second sale, the dummy is $+1$. (This follows Bailey, Muth, and Nourse [3] exactly.)

In the second stage, the squared residuals $\left(\epsilon^2\right)$ from each observation in the first stage are regressed upon τ and τ^2

$$\epsilon^2 = A + B\tau + C\tau^2, \tag{4.4}$$

where τ is the interval between the first and second sale. The coefficients A, B, and C are estimates of σ_ν^2, $\sigma_{\eta_1}^2$, and $\sigma_{\eta_2}^2$ respectively.

In the third stage, the stage one regression is reestimated by GLS with weights $\sqrt{A + B\tau + C\tau^2}$.

The estimated log price level difference $\left(\widehat{I}_{t+\tau} - \widehat{I}_t\right)$ is normally distributed with mean $(I_{t+\tau} - I_t)$, and variance $\left(\tau\sigma_{\eta_1}^2 + \tau^2\sigma_{\eta_2}^2 + \sigma_\nu^2\right)$. Denote $msa_\tau = \exp\left(\widehat{I}_\tau\right)$ as the estimated regional housing price index; then $\log\left(\frac{msa_{\kappa+\tau}}{msa_\kappa}\right)$ is normally distributed with mean $(I_{\kappa+\tau} - I_\kappa)$ and variance $\left(\tau\sigma_{\eta_1}^2 + \tau^2\sigma_{\eta_2}^2 + \sigma_\nu^2\right)$.

Means and Variances are estimated for each of 30 major MSA regions using samples of paired sales. There are about four million paired sales in the Freddie Mac data base.

the unpaid mortgage balance (computed from the contract terms), permit us to estimate the distribution of homeowner equity quarterly for each observation. In particular, "EQY" is the estimate of equity, "EQR" is the equity ratio, and "$PNEQ$" is the probability that equity is negative, *i.e.*, the probability that the put option is in the money.[12]

As proxies for other "trigger events," we include measures of the quarterly unemployment rate and the annual divorce rate.[13]

---

[12]Specifically, equity for the lth loan observation is defined as:

$$\begin{aligned} eqy_l &= \frac{mktvalue_l - pdvunpblc_l}{mktvalue_l} \\ &= \frac{purprice_l \times \frac{msa_{\omega_l,\kappa_l+\tau_i}}{msa_{\omega_l,\kappa_l}} - \sum_{t=1}^{term_l-\tau_i} \frac{mopipmt_l \times 3}{(1+noterate_l/400)^t}}{purprice_l \times \frac{msa_{\omega_l,\kappa_l+\tau_i}}{msa_{\omega_l,\kappa_l}}} \\ &= 1 - \frac{(LTV/100) \times \left(1 - \left(\frac{1}{1+noterate_l/400}\right)^{term_l-\tau_i}\right)}{\left(\frac{msa_{\omega_l,\kappa_l+\tau_i}}{msa_{\omega_l,\kappa_l}}\right) \times \left(1 - \left(\frac{1}{1+noterate_l/400}\right)^{term_l}\right)}, \end{aligned} \tag{4.5}$$

where $purprice_l$ is the purchasing price of the house at the time of loan initiation, and $pdvunpblc_l$ is the present discounted value of the remaining loan balance.

The probability of negative equity, $pneq$, is thus

$$pneq_l = \text{ncdf}\left(\frac{\log(pdvunpblc_l) - \log(mktvalue_l)}{\sqrt{e^2_{\omega_l,\kappa_l+\tau_i}}}\right), \tag{4.6}$$

where $pdvunpblc_l$ and $mktvalue_l$ are defined above, ncdf$(\cdot)$ is cumulative standard normal distribution function, and $e^2_{\omega_l,\kappa_l+\tau_i}$ is defined in footnote 8.

[13]Unemployment and divorce rates are 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*". State divorce data are reported in various issues of U.S. National Center for Health Statistics, "*Vital Statistics of*

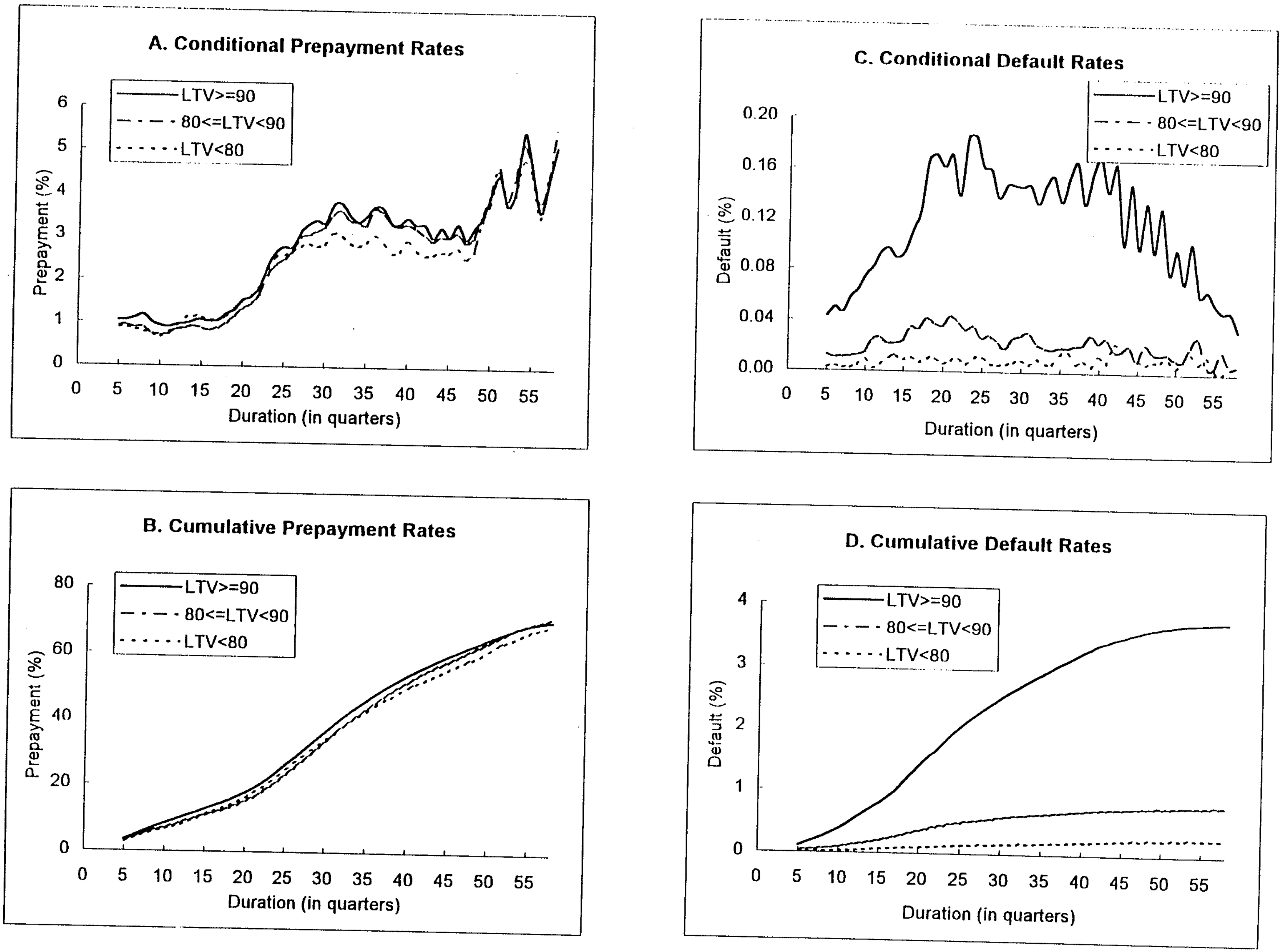
Figure 1 summarizes the raw data used in the empirical analysis. Panel A of Figure 1 displays the conditional prepayment rate, separately by loan-to-value ratio (LTV), as a function of duration. Conditional prepayment rates are slightly higher for higher LTV loans. Rates increase substantially after the first fifteen quarters. Panel C of Figure 1 displays raw conditional default rates by LTV. Note again that default rates increase substantially after about fifteen quarters. Note also that the default rates increase dramatically with initial LTV. Default rates for 90 percent LTV loans are four or five times higher than default rates for 80 to 90 percent LTV loans. The default rates for these latter loans are, in turn, about twice as high as for those with LTV below 80 percent.

Finally, note that conditional default rates are quite low. Even for the riskiest class of loans, conditional default rates are no higher than two 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).

The competing risks CPL model is estimated using a sub-sample of these loans drawn using a choice-based sampling technique which greatly over samples defaulted loans. A sample of 1455 mortgage loans was selected, of which 567 were randomly drawn from prepaid loans, 614 from defaults and 274 from cen-

the United States, Volume III, Marriage and Divorce", and in "*Statistical Abstract of the U.S.*".

Figure 1


A. Conditional Prepayment Rates
LTV>=90
80<=LTV<90
LTV<80
Prepayment (%)
Duration (in quarters)
B. Cumulative Prepayment Rates
LTV>=90
80<=LTV<90
LTV<80
Prepayment (%)
Duration (in quarters)
C. Conditional Default Rates
LTV>=90
80<=LTV<90
LTV<80
Default (%)
Duration (in quarters)
D. Cumulative Default Rates
LTV>=90
80<=LTV<90
LTV<80
Default (%)
Duration (in quarters)

sored observations. Weights appropriate to this choice-based sample were applied to the likelihood function, following Manski and Lerman[21], and the coefficient estimates were obtained by maximum likelihood procedures.

The size of the sub-sample and the covariates included in the analysis are constrained by the computational intensity of the CPL model. The model reported below includes three covariates in the prepayment function and four covariates in the default function.

For the SPE model. the entire sample of 780,443 loans has been partitioned into 120 groups, according to 30 major MSAs and 4 LTV groups. For each group, there are 64 cells. reflecting failure time periods (measured in quarter. from 76:II to 92:I). Empirical hazard rates of prepayment and default have been calculated for every period in each cell based on the entire sample. Then the estimated empirical hazard rates were mapped to 10,977 mortgage loans which were randomly drawn from the total sample. We assume that the randomly-drawn sub-sample has the same distribution as the population.

Table 1 presents coefficients of the competing risks model estimated by the two methods. Each model uses the same set of covariates and the same 1,455 observations. The results of both models confirm the importance of financial factors in motivating prepayment and default. For example, when market interest rates drop, the value of the prepayment option increases, and the prepayment hazard increases. Similarly, when housing prices decline, the probability of negative equity increases, and the default hazard increases. Both of these effects are large in magnitude, but the statistical significance is much higher for the results from the SPE procedure.

Note also that the probability of negative equity is highly significant in the prepayment model, verifying the dependence between prepayment and default behavior. The coefficients of the SPE model provide weak evidence on the link between initial payment-to-income levels and subsequent default behavior. The default probabilities are also sensitive to state variations in marital stability.

The model also suggests that unemployment probabilities are negatively related to default probabilities, *i.e.*, higher unemployment rates are associated with lower default rates. This seems surely to be incorrect and may be attributable to the parsimonious specification of the model.

The process of estimating the parameters in Table 1 also reveals the computational limitations of the CPL method. The estimation of model 1, with 1455 individual mortgages evaluated quarterly for up to 64 quarters took approximately two weeks on a SPARC-2 workstation. Further, it took almost three weeks to estimate the parameters of model 2 using the CPL method. Finally, it

Table 1. Competing Risks of Mortgage Termination Estimated by Cox Partial Likelihood and Semi-parametric Methods

(t ratios in parentheses)

	Model 1				Model 2			
	Prepayment		Default		Prepayment		Default	
	CPL	SPE	CPL	SPE	CPL	SPE	CPL	SPE
POPTION	3.722 (5.18)	3.117 (3.10)			3.951 (7.34)	3.735 (25.79)		
PNEQ			2.954 (2.95)	1.695 (9.02)	-8.467 (2.81)	-1.106 (2.45)	3.063 (1.07)	3.003 (9.67)
Pmnt/Income					0.668 (0.58)	-0.251 (0.84)	0.924 (0.23)	1.452 (2.69)
Unemploymnt							-0.016 (0.05)	-0.169 (5.43)
Divorce							-0.126 (0.14)	0.301 (3.14)

Note: CPL models are estimated by maximum likelihood criteria using observations on 1455 mortgages. SPE models are estimated by minimum distance criteria using the same sample. Baseline hazards (not reported) are estimated simultaneously under the SPE method and are estimated sequentially using the Kaplan-Meier technique under the CPL method.

proved computationally infeasible to expand the specification of the model using the CPL method.

Table 2 presents a variety of models estimated by the SPE method, specifying the prepayment and default functions as a seemingly unrelated regression system. These models are estimated using the sample of 10,977 mortgages. This sample is about seven and a half times as large as that underlying the estimates in table 1. The dependence between prepayment and default behavior is reflected by the correlations of unobserved error terms between the prepayment and default functions.

The results again show that financial motivation is of paramount importance 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.

Model 3 expands model 2 to include unemployment and divorce variables in the prepayment function. Both variables are negative and highly significant — indicating that liquidity constraints (which make refinancing more difficult for unemployed and divorced households) keep them from exercising in-the-money call options. The model also includes a variable measuring the initial payment-to-income ratio. This variable is negative and significant in the default function and negative and less significant in the prepayment function. These general results persist across other specifications reported in Tables 2 and 3.[14] 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 4 expands model 3 to include the initial loan-to-value ratio. The variable is positive and highly significant, particularly in the default function. This may well reveal borrowers' risk preferences.

In Models 5 and 6, we impose the constraint that as the probability of negative equity ratio approaches zero, then the probability of default also approaches zero. The results are basically similar, but the statistical significance of the variable is substantially increased.

Models 7 and 8 investigate the effects of different thresholds in the equity ratio and the initial LTV upon prepayment and default behavior. We include two dummy variables with values of one if the probability of negative equity is less than ten percent, or if the probability of negative equity lies between ten and

[14] These results also persisted in a variety of other specifications not reported here.

Table 2. Competing Risks of Mortgage Prepayment and Default

(t ratios in parentheses)

	Model 3		Model 4		Model 5		Model 6		Model 7		Model 8	
	Prepmt	Default	Prepmt	Default	Prepmt	Default	Prepmt	Default	Prepmt	Default	Prepmt	Default
POPTION	0.067		0.066		0.076		0.075		0.068		0.073	
	(4.71)		(4.64)		(5.37)		(5.29)		(4.89)		(5.10)	
PNEQ	-1.047	6.562	-1.066	5.385	-1.159		-1.173					
	(25.22)	(36.93)	(25.25)	(32.00)	(28.03)		(27.90)					
Log(PNEQ) × 0.01						1.255		0.998				
						(57.94)		(41.19)				
EQR≤0.1 (dummy)									-0.557	0.807	-0.557	0.822
									(20.97)	(9.39)	(20.97)	(9.58)
0.1<EQR≤0.3 (dummy)									-0.137	0.173	-0.136	0.182
									(15.23)	(5.98)	(15.22)	(6.29)
LTV			0.439	2.663			0.052	1.594				
			(2.68)	(40.49)			(3.17)	(22.14)				
LTV≤0.8 (dummy)									-0.072	-1.592	-0.072	-1.588
									(11.31)	(76.62)	(11.29)	(76.56)
LTV≥0.95 (dummy)									0.021	0.428	0.020	0.413
									(2.25)	(14.13)	(2.14)	(13.61)
Pmnt/Income	-0.114	-1.387	-0.122	-1.881	-0.121	-1.582	-0.130	-1.850			-0.056	-0.835
	(3.08)	(8.91)	(3.28)	(12.91)	(3.29)	(10.95)	(3.51)	(13.04)			(1.52)	(7.09)
Unemploymnt	-0.047	-0.012	-0.046	0.003	-0.046	0.041	-0.046	-0.023	-0.043	0.055	-0.043	0.052
	(27.29)	(1.72)	(27.08)	(0.50)	(26.88)	(6.06)	(26.66)	(3.40)	(24.47)	(9.82)	(24.49)	(9.39)
Divorce	-0.085	0.041	-0.086	0.002	-0.083	0.133	-0.084	0.101	-0.100	0.165	-0.100	0.169
	(20.79)	(2.33)	(20.94)	(0.09)	(20.19)	(8.31)	(20.42)	(6.41)	(24.94)	(12.71)	(24.75)	(13.06)
Var. of Res.	0.071	1.334	0.071	1.160	0.071	1.150	0.071	1.101	0.071	0.760	0.071	0.757
Cov. of Res.	0.027		0.024		0.029		0.027		-0.004		-0.005	
R^2	0.866	0.239	0.866	0.341	0.866	0.343	0.866	0.373	0.866	0.567	0.866	0.569

Note: All models are estimated by SPE using 10,977 observations. Baseline hazard estimates are not reported.

thirty percent, respectively. We also include two dummy variables with values of one if the initial LTV is less than 80 percent, or if the initial LTV is at least 95 percent, respectively. Clearly the hazard of prepayment declines as the equity ratio is reduced, while the hazard of default increases as the equity ratio is reduced. The higher LTV group is substantially more likely to exercise both prepayment and default options.

In Table 3, we report the results when a fully simultaneous relationship between household prepayment and default is postulated. The probability of default is included in the prepayment function, and the probability of prepayment is included in the default function. Coefficients of the two-equation system are estimated by three stage least squares. Each of the models illustrates that the probability of prepayment hazard has a negative effect on default behavior, but that the probability of default hazard is insignificant in the prepayment function. Note also that these two endogenous variables are highly correlated with the variables measuring option value, "*POPTION*" and "*PNEQ*". This suggests that the simultaneous equation system may not be a good empirical specification in modeling mortgage prepayment and default behavior.

Figure 2 presents the estimated baseline hazards using the two different estimation approaches. Panels A and B illustrate the baseline hazards from the CPL

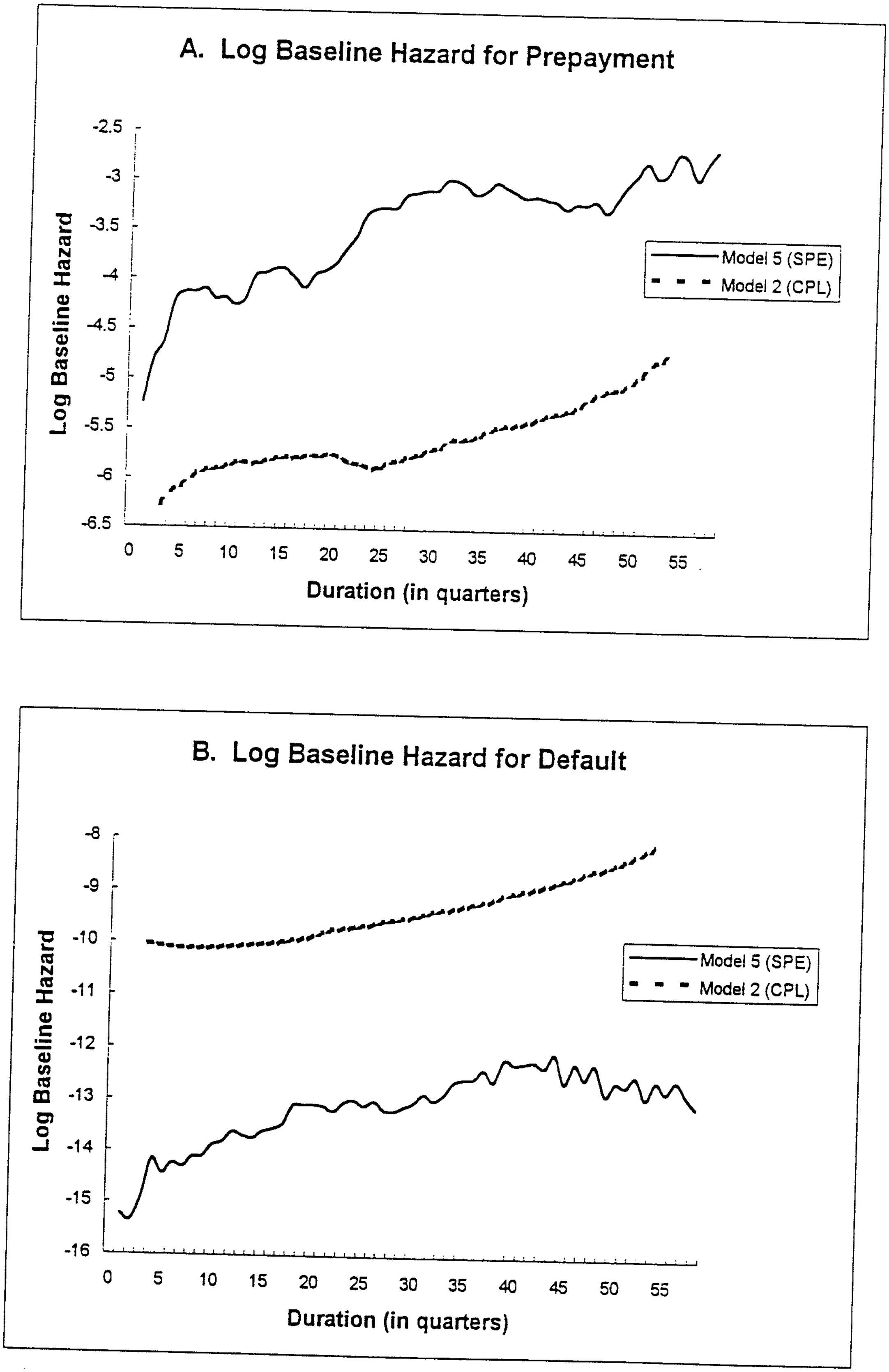
Table 3. 3SLS Estimates of Competing Risks of Mortgage Prepayment and Default

(t ratios in parentheses)

	Model 9		Model 10		Model 11	
	Prepmt	Default	Prepmt	Default	Prepmt	Default
POPTION	0.057		0.065		0.068	
	(4.11)		(4.53)		(4.78)	
PNEQ	-1.057	4.278	-1.046	5.057	-1.067	6.967
	(24.18)	(3.91)	(23.95)	(4.74)	(24.81)	(7.52)
LTV	0.039	2.631			0.044	2.605
	(2.30)	(33.38)			(2.66)	(34.34)
Pmnt/Income			-0.113	-1.499	-0.123	-1.755
			(3.07)	(8.07)	(3.32)	(10.45)
Unemploymnt	-0.046	-0.048	-0.047	-0.084	-0.046	0.077
	(27.09)	(0.96)	(27.28)	(1.67)	(27.06)	(1.79)
Divorce	-0.087	-0.114	-0.085	-0.089	-0.086	0.134
	(21.23)	(1.20)	(20.82)	(0.96)	(20.92)	(1.68)
Prob Prepmnt		-1.149		-1.476		-1.527
		(1.10)		(1.43)		(1.74)
Prob Default	0.000		0.000		0.000	
	(0.00)		(0.00)		(0.00)	
Var. of Res.	0.071	1.330	0.071	1.568	0.071	1.254
Cov. of Res.	0.107		0.132		-0.085	
R^2	0.866	0.279	0.866	0.169	0.866	0.306

Note: All models are estimated by SPE using 10,977 observations. Baseline hazard estimates are not reported.

Figure 2



A. Log Baseline Hazard for Prepayment
Log Baseline Hazard
-2.5
-3
-3.5
-4
-4.5
-5
-5.5
-6
-6.5
0
5
10
15
20
25
30
35
40
45
50
55
Duration (in quarters)
Model 5 (SPE)
Model 2 (CPL)
B. Log Baseline Hazard for Default
Log Baseline Hazard
-8
-9
-10
-11
-12
-13
-14
-15
-16
0
5
10
15
20
25
30
35
40
45
50
55
Duration (in quarters)
Model 5 (SPE)
Model 2 (CPL)

model. using the two-step procedure.[15] In general, the baseline hazard functions have an increasing slope during the first fifteen years of the mortgage life.[16] This indicates that the proportional hazard specification is more appropriate for the mortgage prepayment and default model than the logit model or simple regression models.

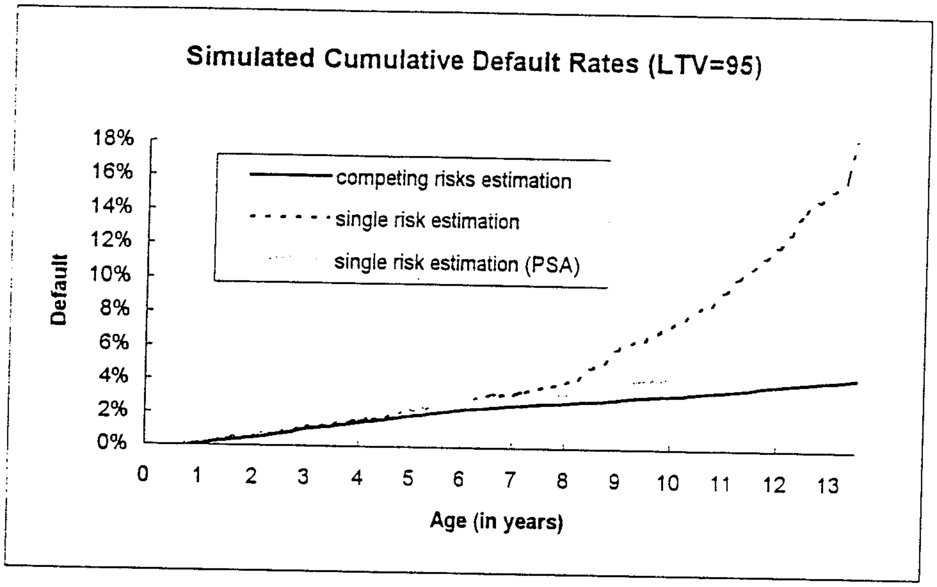
Figures 3 and 4 illustrate the difference between hazard rates estimated jointly from the competing risks model and those estimated separately. We use Freddie Mac's interest rate and house price inflation model to generate 300 random paths of mortgage interest rates and 300 random paths of house price inflation rates.[17] All simulations use the SPE approach. The solid lines in both figures are the average cumulative unconditional default and prepayment rates simulated from the joint (SPE) estimation of default and prepayment functions. The dotted lines in both figures are the average cumulative unconditional default and prepayment rates estimated separately. The shadowed lines in figure 3 are the average cumulative unconditional default rates estimated from the default function by assuming

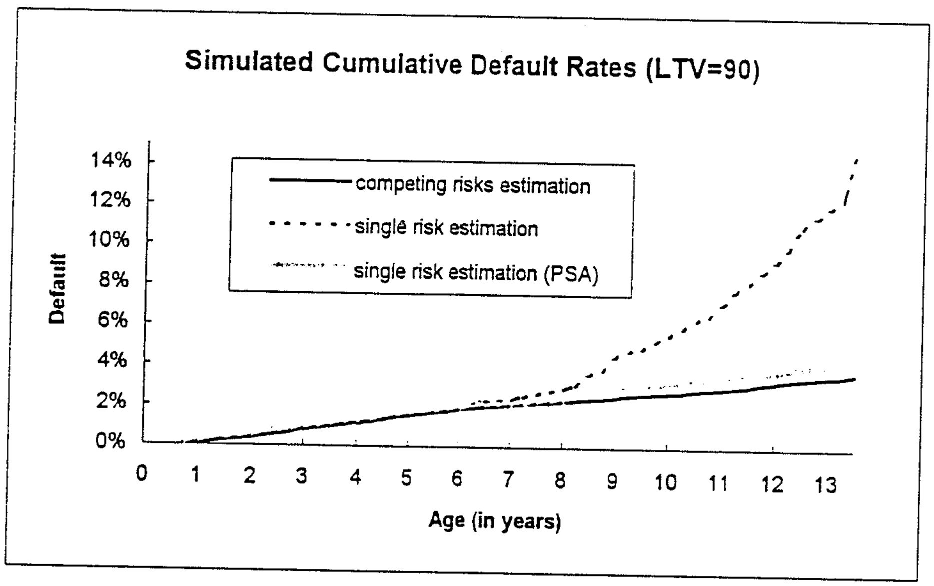
[15]For the CPL model, baseline hazards are estimated quarterly, given the coefficient estimates, using the procedure estimated by Kaplan and Meier. For the SPE model, quarterly baseline hazards are estimated simultaneously with the coefficients.

[16]The data set is not sufficient to investigate the shape of the baseline for the second half of the life of 30 year mortgages.

[17]The mean value of the house price inflation rate is set at 10 percent annually and the mean level of the mortgage contract rate is set at 8 percent. Series are generated from mean reverting processes with positive correlation. The mean value of unemployment and divorce rates are set at 8 and 6 percent respectively.

Figure 3


Simulated Cumulative Default Rates (LTV=95)
competing risks estimation
single risk estimation
single risk estimation (PSA)
Default
18%
16%
14%
12%
10%
8%
6%
4%
2%
0%
0 1 2 3 4 5 6 7 8 9 10 11 12 13
Age (in years)


Simulated Cumulative Default Rates (LTV=90)
competing risks estimation
single risk estimation
single risk estimation (PSA)
Default
14%
12%
10%
8%
6%
4%
2%
0%
0 1 2 3 4 5 6 7 8 9 10 11 12 13
Age (in years)

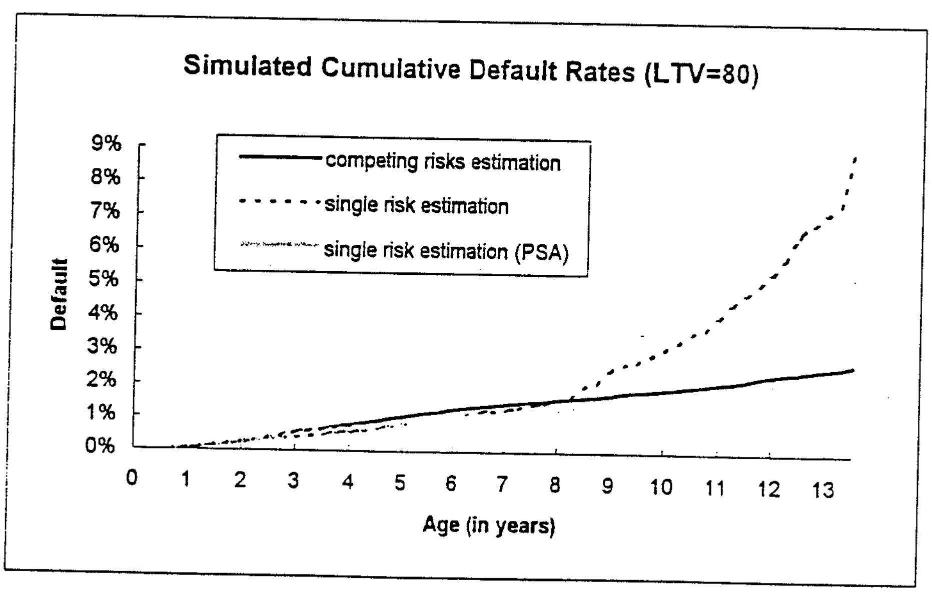
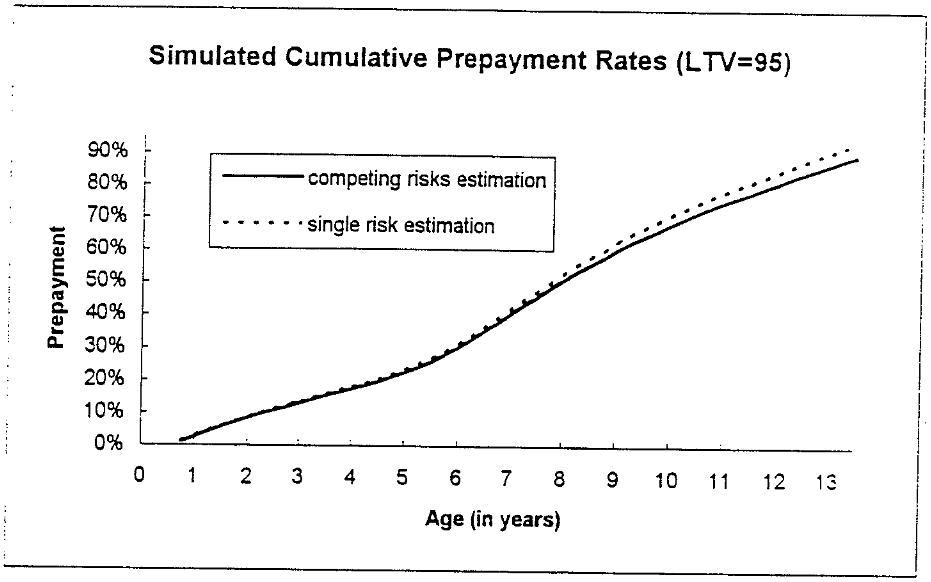
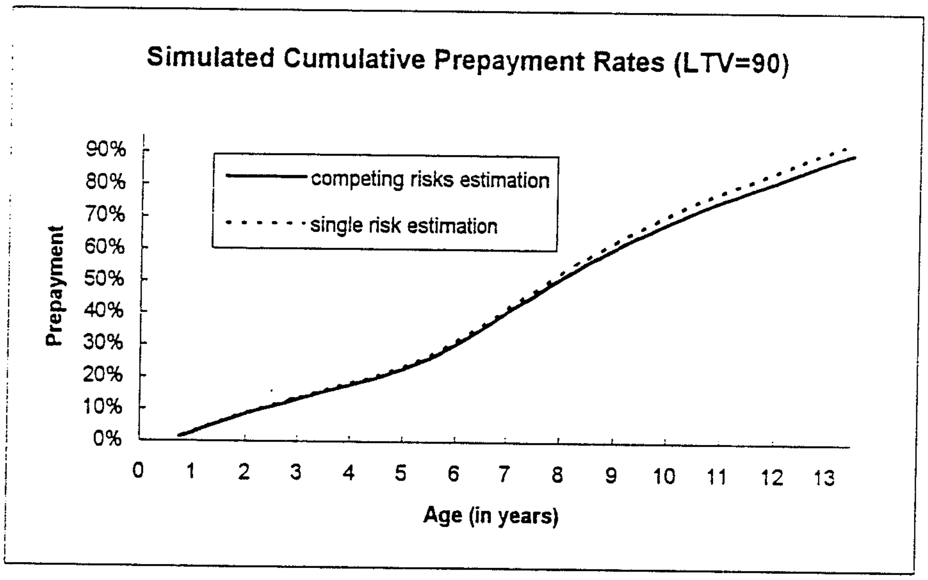

Simulated Cumulative Default Rates (LTV=80)
competing risks estimation
single risk estimation
single risk estimation (PSA)
Default
9%
8%
7%
6%
5%
4%
3%
2%
1%
0%
0 1 2 3 4 5 6 7 8 9 10 11 12 13
Age (in years)

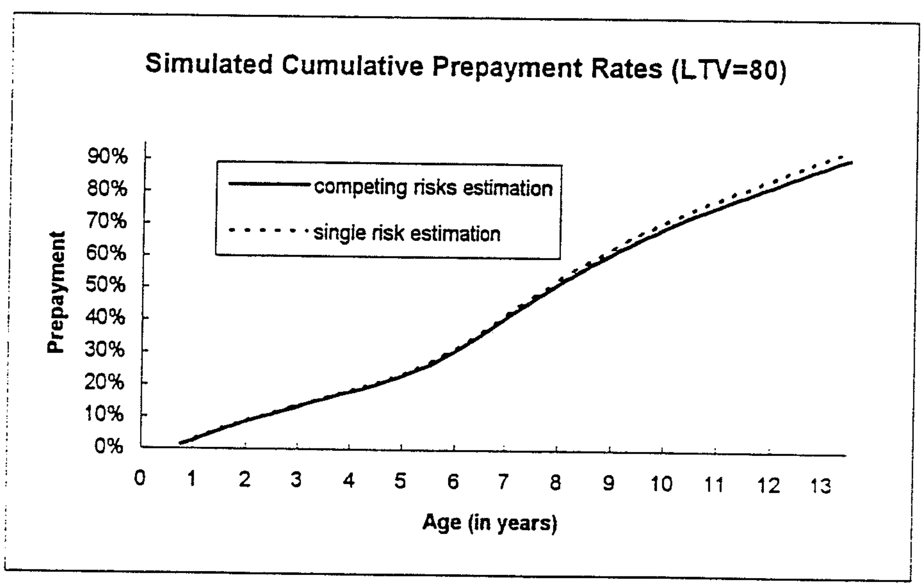
Figure 4



Simulated Cumulative Prepayment Rates (LTV=95)
competing risks estimation
single risk estimation
Prepayment
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
0 1 2 3 4 5 6 7 8 9 10 11 12 13
Age (in years)



Simulated Cumulative Prepayment Rates (LTV=90)
competing risks estimation
single risk estimation
Prepayment
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
0 1 2 3 4 5 6 7 8 9 10 11 12 13
Age (in years)



Simulated Cumulative Prepayment Rates (LTV=80)
competing risks estimation
single risk estimation
Prepayment
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
0 1 2 3 4 5 6 7 8 9 10 11 12 13
Age (in years)

that prepayment follows an exogenously specified industry rule of thumb.[18] The figure indicates that the single risk model over-predicts the cumulative default rate dramatically when the prepayment risk is ignored. The prediction error in the single risk model increases with LTV as well as the mortgage age. (This arises because the single risk default model ignores the negative effect of the prepayment option upon borrowers' default behavior.) However, if we assume a widely used mechanical pattern of prepayment (the "PSA standard"), the single risk estimation still greatly over-predicts the cumulative default rate at high LTV,[19] and it under-predicts the cumulative default rate when LTV equals 80. In contrast, the prediction error from the single risk model of prepayment is a good bit smaller.

5. Conclusion

This paper has presented a unified model of the competing risks of mortgage termination by prepayment and default. The model considers the two hazards as dependent competing risks and estimates them jointly. The paper also presents a semi-parametric estimation approach for the dependent competing risks hazard

[18]This rule of thumb, the "PSA Standard," is defined as follows: 0 percent constant prepayment rate (CPR) in month zero, increasing by .2 percent CPR monthly, rising to 6 percent CPR in month 30, and remaining at 6 percent CPR thereafter through maturity.

[19]For LTV = 95, the average cumulative default rate is 31 percent higher from single risk estimation after 54 quarters (the end of the sample). For LTV = 90, the average cumulative rate is 17 percent higher.

model with time-varying covariates. As shown, the SPE has several important advantages over the CPL when applied to mortgage prepayment and default.

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 also provide strong support for the interdependence of the decisions to prepay and to default on mortgage obligations.

The results also 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, as are those who are at greater risk for divorce (at least as measured by the divorce rate in their state). 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 qualification rules typically enforced by lenders in mortgage refinance.

The results also 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 reflect investor preferences for risk and investor sophistication in the market for mortgages on owner-occupied housing.

References

[1] Abraham. J. M., and W. S. Schauman. "New Evidence on Home Prices from Freddie Mac Repeat Sales," *AREUEA Journal* 19(3): 333-352. Fall 1991.

[2] Amemiya. T. *Advanced Econometrics*. Cambridge, Massachusetts: Harvard University Press, 1985.

[3] Bailey, M. J., R. F. Muth, and H. O. Nourse. "A Regression Method for Real Estate Price Index Construction." *Journal of the American Statistical Association*, 58: 933-942, 1963.

[4] Black, F.. and M. S. Scholes. "The Pricing of Options and Corporate Liabilities," *Journal of Political Economy* 81: 673-659, 1972.

[5] Brennan, M. J., and E. S. Schwartz. "Determinants of GNMA Mortgage Prices," *AREUEA Journal* 13: 209-228, 1985.

[6] Buser, S. A., and P. H. Hendershott. "Pricing Default Free Mortgages," *Housing Finance Review* 3:, 1984.

[7] Case, K. E., and R. J. Shiller. "Prices of Single-Family Homes Since 1970: New Indexes for Four Cities," *New England Economic Review*, Sept/Oct: 45-56, 1987.

[8] Cosslett, S. "Maximum Likelihood Estimator for Choice-Based Samples," *Econometrica* 9: 1289-1316, 1981.

[9] Cox, D. R., and D. Oakes. *Analysis of Survival Data*, New York: Chapman and Hall, 1984.

[10] Cox, J. C., J. E. Ingersoll and S. A. Ross. "A Theory of the Term Structure of Interest Rates," *Econometrica* 53: 385-467, 1985.

[11] Dunn, K. B., and J. J. McConnell. "Valuation of Mortgage-Backed Securities." *Journal of Finance* 36: 599-617, 1981.

[12] Foster, C., and R. Van Order. "FHA Mortgage Termination: A Prelude to Rational Mortgage Pricing," *AREUEA Journal* 13 (3):273-291, 1985.

[13] Green, J., and J. B. Shoven. "The Effect of Interest Rates on Mortgage Prepayments," *Journal of Money, Credit and Banking* 18: 41-50, February 1986.

[14] Härdle, W. *Applied Nonparametric Regression*, Cambridge: Cambridge University Press,1990.

[15] Hendershott, P. H., and R. Van Order. "Pricing Mortgages: An Interpretation of the Models and Results," *Journal of Financial Services Research* 1: 77-111, 1987.

[16] Kalbfleisch, J. D., and R. L. Prentice. *The Statistical Analysis of Failure Time Data,* New York: John Wiley, 1980.

[17] Kau, J., D. Keenan, W. Muller III, and J. Epperson. "Option Theory and Fixed-Rate Mortgages," University of Georgia, Mimeo, 1986.

[18] ———. " The Valuation at Origination of Fixed Rate Mortgages with Default and Prepayment," Working paper, 1993.

[19] Kau, J., and D. Keenan. "An Overview of the Option-Theoretic Pricing of Mortgages," Working paper, 1993.

[20] Lancaster, T. *The Econometric Analysis of Transition Data*, Cambridge: Cambridge University Press, 1990.

[21] Manski, C., and S. Lerman. "The Estimation of Choice Probabilities from Choice-Based Samples," *Econometrica* 45: 1977-1988, 1977.

[22] Manski, C. F., and D. McFadden. *Structural Analysis of Discrete Data with Econometric Applications*. Cambridge, Mass.. London, England: The MIT Press, 1981.

[23] Meyer, B. D. "Unemployment Insurance and Unemployment Spells." *Econometrica*, 58: 757-82, 1990.

[24] Prentice, R. L., and L. A. Gloeckler. "Regression Analysis of Grouped Survival Data with Application to Breast Cancer Data." *Biometrics*, 34: 57-67, 1978.

[25] Quigley, J. M., and R. Van Order. "Efficiency in the Mortgage Market: The Borrower's Perspective," *AREUEA Journal* 18(3): 1990.

[26] ———. "Some Tests of Contingent Claims Models of Mortgage Default," *Journal of Real Estate Finance and Economics*, forthcoming.

[27] Quigley, J. M., R. Van Order, and Y. Deng. "The Competing Risks for Mortgage Terminations by Default and Prepayment: A semi-parametric Estimator," Paper, presented at the NBER Summer Institute 1994, Cambridge, Mass.

[28] Ryu, K. "Group Duration Analysis of the Proportional Hazard Model: Minimum Chi-square Estimators and Specification Tests with Two Empirical Applications," Working Paper, Department of Economics, University of California, Los Angeles, CA 90024, 1993.

[29] Schwartz, E. S., and W. N. Torous. "Prepayment and the Valuation of Mortgage-Backed Securities," *Journal of Finance* 44: 375-392, 1989.

[30] Zellner. A. "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias." *Journal of the American Statistical Association* 57:348-368, 1962.