Woodhead Behavior and the Pricing of Residential Mortgages

by

Yongheng Deng
School of Policy, Planning and Development
University of Southern California
Los Angeles
CA 90089-0626
Phone: 213-821-1030
Fax: 213-740-6170
E-mail: ydeng@usc.edu

John M. Quigley
Department of Economics
University of California
Berkeley
CA 94720-3880
Phone: 510-643-7411
Fax: 510-643-7357
E-mail: quigley@econ.berkeley.edu

December 2000

Paper prepared for the AFA-AREUEA Joint Session, January 5-7, 2001, New Orleans, LA.

Wood-Head Behavior and the Pricing of Residential Mortgages

Abstract

Option theory which has dominated residential mortgage prepayment and default research implies that a borrower will exercise prepayment or default options if the call option or put option, respectively, is "in the money" by some optimal amount. Empirical research provides evidence that 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. Nevertheless, evidence also shows that in general borrowers do not behave as "ruthlessly" as the theory predicts. This paper investigates the irrational behavior of those borrowers who do not call their mortgages even when the option is deeply into the money.

We develop an option-based empirical model of prepayment and default in a competing risks hazard framework to analyze this phenomenon -- the behavior of irrational "wood-heads." Of course we do not observe "woodheads" explicitly in any body of data. Instead, we analyze the heterogeneity and correlates of unobserved heterogeneity within a large sample of mortgage holders. We utilize recent simulation techniques in a three-stage estimation approach to estimate a competing risks hazard model of mortgage prepayment and default with unobserved heterogeneity which is due in part to the behavior of "woodheads."

We then analyze the impact of this behavior on the market value of mortgage and mortgage-backed securities using the industry-standard Monte Carlo simulation pricing approach. The results indicate the empirical importance of heterogeneity and non optimizing behavior in the valuation and pricing of mortgages.

Keywords: Mortgage prepayment, heterogeneity, mortgage pricing.
JEL Codes: G2, R2, D2

I. Introduction

Recent research on the economic behavior of mortgage holders yields three well-known insights. First, the contingent claims model provides a coherent and useful framework for analyzing borrower behavior. Default and prepayment are options to put and call the contract respectively, and other aspects of the mortgage (including interest rate caps and other details of adjustable rate mortgages) are usefully viewed as options. (See, for example, Kau and Keenan, 1995 for a recent survey). Second, the jointness of the prepayment and default options is important in explaining behavior. A homeowner who exercises a default option today gives up the option to default tomorrow, but she also gives up the option to prepay tomorrow. Kau, et al., (1995) have outlined the theoretical relationships among the options, and Schwartz and Torous (1993) have demonstrated their practical importance. Third, duration or competing risks models provide a convenient analytical tool for analyzing borrower behavior. Models of this sort were first applied to borrower behavior in the mortgage market almost fifteen years ago (See Green and Shoven, 1986), and they have increased in realism and sophistication in the past decade. (See, Deng, 1997 for a recent application.)

This paper analyzes a fourth issue in making this approach useful in empirical applications, namely the heterogeneity of mortgages holders. In the original applications of duration models to bio-statistics problems, the unobserved heterogeneity of subjects was clearly recognized. For example, in early models analyzing the survival times of patients after medical treatment, it was pointed out that those who are least physically fit are more likely to succumb and to exit the sample of subjects (Kalbfleisch and Prentice, 1980). In later work applying these models to labor markets, the same issue of selectivity was emphasized (Heckman and Singer, 1985).

An analogous complication arises in duration models of mortgage terminations. After a mortgage is issued, those who are most financially astute are those most likely recognize, and thus to exercise, in-the-money options to terminate. This means that any sample of surviving mortgage holders is successively more likely to include disproportionate fractions of those less financially astute. This fact can have important implications for the pricing of pools of mortgages.

The empirical importance of heterogeneity of mortgage borrowers is demonstrated empirically in our companion paper (See Deng, et al., 2000). The estimated parameters of failure time models of the behavior of mortgage holders are very different when unobserved heterogeneity is accounted for. In particular, the magnitude and significance of variables measuring the values of options are much larger when unobserved heterogeneity is accounted for.

Appropriate methods for controlling for completely unobserved heterogeneity among borrowers may include assumptions about discrete groupings of heterogeneous agents (Deng, et al., 2000) or assumptions about mixture distributions of agents with different underlying hazards (Hall, 2000). In contrast, Stanton (1995, 1996) and others (e.g., Richard and Roll, 1989) have specified heterogeneity among pools of mortgage securities, not individual mortgage holders. Stanton applies a mixture distribution to analyze mortgage pool prepayment risks by mixing a prepayment hazard function which is homogeneous across agents with pool-specific transaction cost functions. The exogenous transaction cost function is assumed to follow a beta distribution which varies across individual mortgage pools. Of course, completely unobserved heterogeneity of agents provides little help in the pricing of mortgage pools (although models recognizing completely unobserved heterogeneity do identify unbiased coefficient estimates).

This paper presents a model of borrower behavior in the mortgage market in which some correlates of the unobserved heterogeneity of individual borrowers are observed. We use this information to develop a simulated maximum likelihood estimator of the proportional hazard model in the presence of unobserved heterogeneity among mortgage holders. The model we develop is completely general in that we can specify any continuous distribution of unobserved heterogeneity in the population.

Significantly, the model can be used to price mortgage pools in real time, that is, it does not depend upon information obtained only after observing the behavior of an entire cohort of borrowers over the lifetimes of their contracts. In contrast, the model developed here permits spot prices to be updated continuously with the information revealed by the behavior of borrowers holding mortgage contracts. This feature may have direct application in the secondary mortgage market for the pricing of mortgage-backed securities composed of seasoned loans.

In section II below we sketch out the basic model and the estimation strategy employed. In section III we estimate the model using a sample of individual mortgages. We compare the results of this estimation procedure with those obtained assuming completely unobserved heterogeneity among borrowers. In section IV we consider the pricing implications of these models.

II. The Model

The proportional hazard model introduced by Cox and others (Cox and Oakes, 1984) provides a framework for considering the contingent claims model empirically and for measuring the effect of financial options on the behavior of mortgage holders.

Let T_p and T_d be discrete random variables representing the duration of a mortgage until it is terminated by the mortgage holder in the form of prepayment or default, respectively. Following the Cox model, the joint survivor function conditional on ξ_p, ξ_d, r, H, Y, and X can be expressed in the following form:

$$
\begin{aligned}
&S\left(t_p, t_d \mid r, H, Y, X, \xi_p, \xi_d, \theta\right) \\
&= \exp\left\{-\xi_p \sum_{k=1}^{t_p} \exp\left(\gamma_{pk} + \beta_{p_1}^{'} g_{pk}\left(r, H, Y\right) + \beta_{p_2}^{'} X\right)\right. \qquad (1) \\
&\qquad \left. -\xi_d \sum_{k=1}^{t_d} \exp\left(\gamma_{dk} + \beta_{d_1}^{'} g_{dk}\left(r, H, Y\right) + \beta_{d_2}^{'} X\right)\right\}.
\end{aligned}
$$

In this formulation $g_{jk}\left(r, H, Y\right)$ are time-varying variables measuring the financial values of the prepayment and default options ($j = p$, d). r and H are the relevant interest rates and property values, respectively, and Y is a vector of other variables that are also relevant to describing the market values of the options empirically. X is a vector of other non-option-related variables, which may include indicators reflecting a borrower's credit risk or financial strength, as well as other trigger events, such as unemployment and divorce. X may include time varying covariates. ξ_p and ξ_d are unobserved error terms associated with the hazard functions for prepayment and default respectively. θ is a vector of parameters (*e.g.*, γ and β) of the hazard function. γ_{jk} are parameters of the baseline hazard function. The baseline may be estimated nonparametrically, following Han and Hausman (1990):

$$\gamma_{jk} = \log\left[\int_{k-1}^{k} h_{0j}(s)\,ds\right], \quad j = p, d. \tag{2}$$

Alternatively, the form of the baseline may be imposed employing by some standard such as "PSA experience." [1]

As noted above, a major impediment to analyzing economic behavior of mortgage holders is the unobserved borrower-specific heterogeneity embedded in the empirical data we observe. In other words, ξ_i in equation (1) may be decomposed into two parts: μ and η_i. μ is a fixed effect error term representing for example, a proportionate shift in the non parametric or PSA experience baseline. η_i is an unobserved borrower-specific error term. The joint survivor function for the *i*th borrower can be rewritten in following form:

$$\begin{aligned} &S\left(t_{pi}, t_{di} \mid r_i, H_i, Y_i, X_i, \mu_p, \mu_d, \eta_{pi}, \eta_{di}, \theta\right) \\ &= \exp\left\{-\eta_{pi}\mu_p \sum_{k=1}^{t_{pi}} \exp\left(\gamma_{pk} + \beta_{p_1}^{'} g_{pki}\left(r_i, H_i, Y_i\right) + \beta_{p_2}^{'} X_i\right)\right. \qquad (3) \\ &\qquad \left. -\eta_{di}\mu_d \sum_{k=1}^{t_{di}} \exp\left(\gamma_{dk} + \beta_{d_1}^{'} g_{dki}\left(r_i, H_i, Y_i\right) + \beta_{d_2}^{'} X_i\right)\right\}. \end{aligned}$$

Due to the nature of the competing risks between prepayment and default, only the duration associated with the type which terminates first is observed, i.e. $t_i = \min\left(t_{pi}, t_{di}\right)$. Define $F_p\left(t_i \mid \eta_{pi}, \eta_{di}\right)$ as the probability of mortgage termination by prepayment of the *i*th borrower in period *t*, $F_d\left(t_i \mid \eta_{pi}, \eta_{di}\right)$ as the probability of mortgage termination by default of the *i*th borrower in period *t*, $F_u\left(t_i \mid \eta_{pi}, \eta_{di}\right)$ as the probability of mortgage termination by the *i*th borrower in period *t* but with missing information on the cause of the termination, and $F_c\left(t_i \mid \eta_{pi}, \eta_{di}\right)$ as the probability that mortgage duration

[1] The Public Securities Association (PSA) has introduced a widely adopted prepayment measurement standard. This is a series of 360 monthly prepayment rates expressed as constant annual prepayment rates. The series begins at 0.2 percent in the first month and increases by 0.2 percent in each successive month until month 30, when the series levels out at 6 percent per year until maturity. Prepayments are measured as simple linear multiples of this schedule. Therefore, by applying this PSA schedule as the baseline, the proportional factor estimated from the hazard model can be simply expressed as a percentage of the "PSA experience."

data are censored for the ith borrower in period t due to the ending of the data collecting period, such that [2]

$$F_p\left(t_i \mid \eta_{pi},\eta_{di}\right) = S\left(t_i,t_i \mid \eta_{pi},\eta_{di}\right) - S\left(t_i+1,k_i \mid \eta_{pi},\eta_{di}\right) - \frac{1}{2}\Big\{S\left(t_i,t_i \mid \eta_{pi},\eta_{di}\right) + S\left(t_i+1,t_i+1 \mid \eta_{pi},\eta_{di}\right) - S\left(t_i,t_i+1 \mid \eta_{pi},\eta_{di}\right) - S\left(t_i+1,t_i \mid \eta_{pi},\eta_{di}\right)\Big\}, \tag{4}$$

$$F_d\left(t_i \mid \eta_{pi},\eta_{di}\right) = S\left(t_i,t_i \mid \eta_{pi},\eta_{di}\right) - S\left(t_i,t_i+1 \mid \eta_{pi},\eta_{di}\right) - \frac{1}{2}\Big\{S\left(t_i,t_i \mid \eta_{pi},\eta_{di}\right) + S\left(t_i+1,t_i+1 \mid \eta_{pi},\eta_{di}\right) - S\left(t_i,t_i+1 \mid \eta_{pi},\eta_{di}\right) - S\left(t_i+1,t_i \mid \eta_{pi},\eta_{di}\right)\Big\}, \tag{5}$$

$$F_u\left(t_i \mid \eta_{pi},\eta_{di}\right) = S\left(t_i,t_i \mid \eta_{pi},\eta_{di}\right) - S\left(t_i+1,t_i+1 \mid \eta_{pi},\eta_{di}\right), \tag{6}$$

and

$$F_c\left(t_i \mid \eta_{pi},\eta_{di}\right) = S\left(t_i,t_i \mid \eta_{pi},\eta_{di}\right). \tag{7}$$

The unconditional probability of termination is obtained by conditioning on the unobserved η_{pi},η_{di} and then integrating over their distribution such that:

$$F_j\left(t_i\right) = \int_0^\infty \int_0^\infty F_j\left(t_i\right) dG\left(\eta_{pi},\eta_{di}\right), \qquad j = p,d,u,c. \tag{8}$$

where $G\left(\eta_{pi},\eta_{di}\right)$ is the c.d.f. of unobserved heterogeneous error terms for borrower i.

The log likelihood function of the competing risks model is given by

$$\log L = \sum_{i=1}^{N} \delta_{pi} \log\left(F_p\left(T_i\right)\right) + \delta_{di} \log\left(F_d\left(T_i\right)\right) + \delta_{ui} \log\left(F_u\left(T_i\right)\right) + \delta_{ci} \log\left(F_c\left(T_i\right)\right), \tag{9}$$

where $\delta_{ji},\ j = p,d,u,c$, are indicator variables that take the value of one if the ith loan is terminated by prepayment, default, unknown type, or censoring, respectively and take a value of zero otherwise.

---

[2] The dependence of these functions on r, H, Y, X, μ_p, μ_d, and θ has been omitted for notational simplicity. The term $\frac{1}{2}\Big\{S\left(t_i,t_i \mid \eta_{pi},\eta_{di}\right)+S\left(t_i+1,t_i+1 \mid \eta_{pi},\eta_{di}\right)-S\left(t_i,t_i+1 \mid \eta_{pi},\eta_{di}\right)-S\left(t_i+1,t_i \mid \eta_{pi},\eta_{di}\right)\Big\}$ which appears in both equations (4) and (5) is an adjustment for mortgage duration data measured in discrete rather than continuous time.

Without loss of generality, assume the unobserved heterogeneity is only present in the prepayment hazard function.[3] The true population of η_{pi} which characterizes $G(\cdot)$ is typically not observed (and is perhaps not observable). However Stinebrickner (1999) has noted that consistent evaluation of expression (8) can be obtained by substituting a consistent estimate $\hat{\eta}_{pi}$ for the unobserved η_{pi}.

Due to both the nonlinear nature of equation (8) and the dependence of the elements in η_{pi}, no analytic solution for equation (8) exists. However, Stinebrickner (1999) suggested that the integral can be simulated as

$$
\begin{aligned}
&L_i^s\left(r_i, H_i, Y_i, X_i, \theta, \mu_p, \mu_d, \hat{\eta}_{pi}\right) \\
&\quad = \frac{1}{D}\sum_{d=1}^{D} L_i\left(r_i, H_i, Y_i, X_i, \theta, \mu_p, \mu_d, \eta_{pi}\left(\eta_{pi}^{*d}\right)\right),
\end{aligned}
\tag{10}
$$

where D is the number of simulations and η_{pi}^{*d} represents the d^{th} simulation draw of η_{pi}^{*} from its distribution conditional on the observed data.

More concretely, consider a three-stage approach to obtain a consistent estimate of $\hat{\eta}_{pi}$ and to replace equation (8) with the simulated likelihood function as described in equation (10):

Assume that the unobserved heterogeneous component in the prepayment hazard function follows a log normal distribution[4] with a density function of $\phi\left(\log \eta_{pi} \mid \varpi_{pi}\right)$, where $\phi(\cdot)$ is a normal density function, ϖ_{pi} is the parameter of the normal density function. The η_{pi} component is independently distributed across borrowers.

First, a probit model of prepayment is estimated. The prepayment function is specified as $\Phi\left(r_i, H_i, Y_i, X_i, \psi, \varepsilon_i\right)$, where $\Phi(\cdot)$ is a cumulative normal distribution function, r_i, H_i, Y_i, and X_i are specified in equation (1), ψ is a set of parameters of the probit function and ε_i is an error term following a standard normal distribution.

---

[3] This assumption is consistent with the empirical finding from Deng et al (2000). We make this assumption here for convenience only.

[4] The assumption of log normal distribution is not necessary here. It can be any continuously differentiable, nonnegative aggregator function.

Second, the residuals $\hat{\varepsilon}_i$ obtained from the first stage are regressed on the correlates of heterogeneity Z_i such that

$$\hat{\varepsilon}_i = f\left(Z_i, \upsilon_i\right) \tag{11}$$

where υ_i is a random error term following a standard normal distribution.

Third, the likelihood function specified in equation (9) is estimated by replacing $\hat{\eta}_{pi}$ with repeated draws of η_{pi}^{*d}, such that

$$\begin{aligned} \log F_i^s\left(t_i\right) = -\frac{1}{D}\sum_{d=1}^{D}\beta_{p_3}^{'}\eta_{pi}^{*d}\mu_p\sum_{k=1}^{t_i}\exp\left(\gamma_{pk}+\beta_{p_1}^{'}g_{pk}\left(r_i,H_i,Y_i\right)+\beta_{p_2}^{'}X_i\right) \\ -\mu_d\sum_{k=1}^{t_i}\exp\left(\gamma_{dk}+\beta_{d_1}^{'}g_{dk}\left(r_i,H_i,Y_i\right)+\beta_{d_2}^{'}X_i\right) \end{aligned} \tag{12}$$

where $\eta_i^{*d} = \exp\left(\hat{f}\left(Z_i, \upsilon_i^d\right)\right)$, β_{p_3} is a parameter to be estimated jointly with the rest of parameters, θ, of the hazard function, and $\hat{f}$ is the function estimated in equation (11). υ_i^d is the d^{th} draw from a standard normal distribution, $d = 1, \ldots, D$.

III. Empirical Application

We implement this strategy using a rich sample of individual mortgage loan histories 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 the 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 an interest payment, the state, the region and the major metropolitan area in which the property is located. 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 at the fist quarter of

1992. The analysis is confined to loans issued in 30 major metropolitan areas (MSAs)–a total of 47,042 observations. Loans are observed in each quarter form the quarter of origination through the quarter of termination, maturation, or through 1992:I for active loans.

Our analysis is based upon a five percent random sample of these 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 *"Call Option"* (i.e., an element of $g_{pk}\left(r,H,Y\right)$ in section II) measuring the ratio of the present discounted value of the unpaid mortgage balance at the current quarterly mortgage interest rate relative to the value discounted at the contract interest rate.[5]

We also 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 a weighted repeat sales house price index (WRS) separately for each of the 30 MSAs. The WRS index (See Case and Shiller, 1987) provides estimates of the course of house prices in each metropolitan area. Assuming that house prices follow a random walk, the WRS index also provides an estimate of the variance in price for each house in the sample, by metropolitan area and elapsed time since purchase (Deng, et al., 2000).

Estimates of the mean and variance of individual house prices, together with unpaid mortgage balance (computed for the contract terms), permit us to estimate the distribution of homeowner equity quarterly of reach observation. In particular, the variable *"Put Option"* (i.e., an element of $g_{pk}\left(r,H,Y\right)$ in section II) measures the probability that homeowner equity is negative, i.e., the probability that the put option is in

[5] Specifically, for fixed-rate level-payment mortgage i with a mortgage note rate of r_i , and the mortgage term in quarters of TM_i, at each quarter k_i after origination at time τ_i, the local market interest rate is m_{j,τ_i+k_i} , where j indexes the local region, the "*Call Option*" is defined as: $Call\_Option_{i,k} = 1 - V_{i,r}^{*} \big/ V_{i,m}$, where $V_{i,r}^{*} = \sum_{S=1}^{TM_i - k_i} 1 \big/ \left(1 + r_i\right)^{s}$, and $V_{i,m} = \sum_{S=1}^{TM_i - k_i} 1 \big/ \left(1 + m_{j,\tau_i+k_i}\right)^{s}$.

the money.[6] The details of the calculation of the options variables are reported in Deng et al. (2000).

As proxies for other "trigger events," we include measure of the quarterly unemployment rate and the annual divorce rate by state (i.e., *X* in section II).

Our principal correlates of the unobserved heterogeneity across borrowers (i.e. *Z* in equation (11)) are computed in the following way. At each quarter since origination, we calculate whether the call option is in the money (this merely indicates whether interest rates on new first mortgages[7] are lower than the contract interest rate by 50 basis points). We then compute a time varying covariate for each borrower reflecting the number of times since origination that an in-the-money call was not exercised. A "woodhead" is a borrower who systemically passes up profitable opportunities to prepay the mortgage.

Table 1 reports the average values of these variables for the five percent sample of mortgages. On average at origination, the call option is out of the money and the put option is about zero. However, for loans that ultimately defaulted, the value of the put option at origination was more than 5 times as large as it was for loans that ultimately prepaid. At termination, the value of the put option on subsequently defaulted loans was 16 times the value for prepaid loans.

On average, the holders of prepaid mortgages passed up 1.2 potentially profitable refinance opportunities. The holders of defaulted mortgages passed up more than twice as many opportunities.

Table 2 presents two variants of the competing risks model of mortgage termination. Each model includes the value of each option (and its squared value) in both risk equations. The results confirm the theoretical prediction that the value of both options is important in governing the exercise of either option.

---

[6] Specifically, the market value M_i of property *i*, purchased at a cost of C_i at time τ_i and evaluated k_i quarters thereafter is $M_{i,k} = C_i\left(I_{j,\tau_i+k_i} \big/ I_{j,\tau_i}\right)$, where the term in parentheses follows a log normal distribution. The "*Put Option*" variable is defined as: $Put\_Option_{i,k} = \Phi\left(\left(\log V_{i,m} - \log M_{i,k}\right) \big/ \sqrt{\omega^2}\right)$, where $\Phi(\cdot)$ is cumulative standard normal distribution function, ω^2 is an estimated variance, and $V_{i,m}$ is defined in footnote 6.

[7] Interest rates on new mortgage contracts are available by quarter and region. See, for example, the Freddie Mac web site, http://www.freddiemac.com.

In addition to the variables measuring the value of the options, Model 1 includes state average unemployment and divorce rates. Model 2 also includes the initial loan-to-value ratio (LTV), in four categories. In both models, we use the PSA experience as the baseline for the prepayment function. The baseline for the default function is estimated non parametrically.

None of these models controls for unobserved heterogeneity in any way. The results confirm the importance of the option values in the exercise of prepayment and default by mortgage holders. They also provide some evidence that trigger events (unfortunately measured only at the state level) are important in governing exercise. The results also suggest that LTV ratios may reveal information on attitudes towards risk. In any event, the results indicate that, ceteris paribus, those with higher LTVs are more likely to exercise options.

Table 3 presents the same two specifications in a model of unobserved heterogeneity. In Models 3 and 4, we allow for the possibility that there are two distinct groups of borrowers: "ruthless players" and "woodheads." Each borrower belongs to one or the other group, but we do not observe directly the group to which any individual belongs. In Models 5 and 6 we assume there are three distinct groups of borrowers. For each model, we estimate the distribution of unobserved heterogeneity jointly with the competing risk functions.

For both models the magnitude of the option values increases substantially when heterogeneity is accounted for. The magnitudes of the other variables change very little. A comparison of the models which specify two groups of borrowers (Models 3 and 4) with those specifying three groups (Models 5 and 6) reveals no advantage to the more complex specification. The mass point associated with the third group is never significant, and there is no significant improvement in the log likelihood function.

There is a substantial difference between the two groups, however, in exercising the prepayment and the default options. For the prepayment option, those in the high risk group are almost 10 times riskier (e.g., $0.532 \div 0.056$) than borrowers in the low risk group. This difference is highly significant. For the default option, those in the high risk group are about 4.5 times riskier (i.e., $0.366 \div 0.080$) than borrowers in the low risk group. However, this latter difference is not statistically significant at the 5 percent level.

For models 3 and 4 almost 95 percent (e.g., 1/[1+0.058]) of all borrowers are classified into the high risk group.

We now exploit additional information in the estimation of this model, the presumed correlation between our measure of "missed opportunities" and the unobserved heterogeneity among mortgage borrowers.

We begin by estimating a probit model of mortgage prepayment using loan-level event history data. For each quarter after origination, we observe the value of the two options and the contemporaneous state divorce and unemployment rates. We also observe whether each mortgage was prepaid during that quarter. The sample of 22,293 mortgages yields 804,582 loan-level event histories. Table 4 reports the results of the probit estimation using these quarterly observations on prepayment. The results are qualitatively similar to those reported in tables 2 and 3 for the prepayment function. Importantly, however, these probit estimates provide a set of residuals for the simulated maximum likelihood estimation.

We regress the residuals upon our measure of the "number of missed opportunities" each borrower has had up to the current quarter year. We explore two specifications of equation (11):

Specification 1:

$$\hat{\varepsilon}_{it} = \alpha + \beta_1 M_{it} + \beta_2 T_{it} + \nu_{it}$$

Specification 2:

$$\hat{\varepsilon}_{it} = \alpha + \beta_1 M_{it} + \beta_2 T_{it} + \frac{\nu_{it}}{\sqrt{M_{it}/T_{it}}}$$

where $\hat{\varepsilon}_{it}$ is the residual from the probit model (reported in table 4), M_{it} is the number of missed opportunities (i.e. the number of times that individual *i* has failed to exercise the prepayment option when it was in the money from origination to time *t*), T_{it} is age of the mortgage measured in quarters, and ν_i is a random error term which follows a standard normal distribution. Specification 1 assumes that the unobserved error components of the prepayment hazard function follow a normal random distribution with a potentially different mean for each individual, while specification 2 presents the unobserved error components of the hazard function to follow a normal random distribution with not only a different mean but also heterogeneous variance-covariance matrices across each individual.

The estimated coefficients are highly significant in each specification:

Specification 1:

$$\hat{\varepsilon}_{it} = -\underset{(21.75)}{0.006084} - \underset{(31.19)}{0.002440}\, M_{it} + \underset{(29.65)}{0.000309}\, T_{it} + \nu_{it}$$

Specification 2:

$$\hat{\varepsilon}_{it} = -\underset{(62.10)}{0.013623} - \underset{(96.00)}{0.003584}\, M_{it} + \underset{(57.27)}{0.000776}\, T_{it} + \frac{\nu_{it}}{\sqrt{M_{it}/T_{it}}}$$

where *t* ratios are in parentheses.

Tables 5 uses these two specifications in the estimation of the competing risks models with unobserved heterogeneity. In each of these models, heterogeneity is modeled as a continuous measure and is estimated jointly with the parameters of the competing risk equations. A comparison of Tables 5 and 6 indicates that the results are quite consistent across error specifications.

A comparison with Table 3, in which heterogeneity is specified as two or three distinct groups, indicates that the magnitude of the coefficients of the option values is larger for the continuously varying specification of heterogeneity. The coefficients of the other variables are largely unchanged.

The values of the log likelihood function in Table 5 are substantially higher than those of similar models reported in Table 3. For the simplest model, the specification of unobserved heterogeneity as a continuous measure yields a log likelihood of –74,117 (Model 7) as compared with a value of –74,471 for the discrete measure (Model 3). For the models including LTV (Models 8 and 4), the value of the log likelihood is also much larger for the continuous measure (-74,039) than for the discrete representation (-74,391).

The results based upon specification 2, models 9 and 10 reconfirm the above finding.

IV. Conclusion

Table 6 provides a summary of the importance of unobserved heterogeneity in the analysis of borrower behavior and also of the importance of the technique used in this paper relative to conventional MLE methods. Model 11 present MLE estimates of the competing risks of mortgage termination without controlling for the unobserved heterogeneity. Model 12 present MLE estimates with controlling for the unobserved heterogeneity among borrowers. In these models, we also include M, the "woodhead" variable, as a covariate in the risk equations. This specification is thus similar to those used by Richard and Roll (1989) and Schwartz and Torous (1993) in their analyses of prepayments for pools of mortgages. As reported by Stanton (1995), this specification is similar to representations of borrower "burnout" in the practitioner literature.

In models 11 and 12, the option variables are highly significant and the proxies for trigger events, unemployment and divorce, are also highly significant. There is continuing evidence that the propensity to exercise options varies with initial LTV. The "woodhead" variables are significant as predators of prepayment, not default. This finding persists even when unobserved heterogeneity is accounted for, in Model 12.

The results from the SMLE method introduced in this paper are qualitatively similar to those noted in the MLE models. In Models 13 and 14, the coefficients of the options variables are marginally higher. The coefficients of the trigger events are unchanged. The LTV variables and the "woodhead" variables are also similar.

However, the SMLE models fit the data much better than the model estimated by maximum likelihood. For a comparably specified model, the value of the likelihood function is –74,393 for MLE estimation, and it improves to –74,307 when unobserved heterogeneity is accounted for. When the "woodhead" variables are specified as correlates of the unobservable differences across individuals, the model improves quiet substantially. For both specifications, the likelihood improves to around –74,020. In either case, the method of specifying the heterogeneity among borrowers really matters in fitting the model to prepayment and default experience.

Presumably, these differences can be taken into account in calculating the equilibrium price for an individual contract. More importantly, however, these differences can be taken into account in the pricing of securities and pools of seasoned mortgages. Our further research will address this question.

REFERENCE

Case, K. E., and R. J. Shiller (1987): "Prices of Single-Family Homes Since 1970: New Indexes for Four Cities," *New England Economic Review*, September, 45-56.

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

Deng, Yongheng (1997): "Mortgage Termination: An Empirical Hazard Model with Stochastic Term Structure," *Journal of Real Estate Finance and Economics*, 14, 309-331.

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

Green, J., and J.B. Shoven (1986): "The Effect of Interest Rates on Mortgage Prepayment," *Journal of Money, Credit and Banking*, 18, 41-50.

Hall, Arden (2000): Controlling for Burnout in Prepayment Model Estimation, Fair, Isaacs, Co., mimeo January 29.

Heckman, James J. and B. Singer (1984): "Econometric Duration Analysis," *Journal of Econometrics*, 32, 59-84.

Kalbfleisch, J. D., and R. L. Prentice (1980): *The Statistical Analysis of Failure Time Data*. New York: Wiley.

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 (1995): "The Valuation at Origination of Fixed Rate Mortgages with Default and Prepayment," *Journal of Real Estate Finance and Economics,* 11(1), 5-36.

Richard, S. F. and R. Roll (1989): "Prepayments on Fixed Rate Mortgage-Backed Securities," *Journal of Portfolio Management* 15, 73-82.

Schwartz, E.S., and W.N. Torous (1993): "Mortgage Prepayment and Default Decisions: a Poisson Regression Approach," *Journal of the American Real Estate and Urban Economics Association*, 21(4), 431-449.

Stanton, R. (1995): "Rational Prepayment and the Valuation of Mortgage-Backed Securities," *The Review of Financial Studies*, 8(3), 677-708.

Stanton, R. (1996): "Unobservable Heterogeneity and Rational Learning: Pool Specific vs. Generic Mortgage-Backed Security Prices," , *Journal of Real Estate Finance and Economics* 12, 243-263.

Stinebrickner, Todd R. (1999): "Estimation of a Duration Model in the Presence of Missing Data," *Review of Economics and Statistics*, 81(3), 529-542.

TABLE I.

DESCRIPTIVE STATISTICS ON MORTGAGE LOANS MEAN VALUES AT ORIGINATION AND TERMINATION

Variable	At Origination				At Termination	
	All Loans	Prepaid	Defaulted	Other*	Prepaid	Defaulted
Call Option (fraction of contract value)	-0.0514 (0.0066)	-0.0536 (0.0070)	-0.0728 (0.0130)	-0.0439 (0.0049)	-0.0221 (0.0311)	0.0364 (0.0305)
Put Option (probability of negative equity)	0.0069 (0.0005)	0.0067 (0.0005)	0.0350 (0.0023)	0.0058 (0.0004)	0.0067 (0.0015)	0.1080 (0.0379)
Squared Term of Call Option	0.0092 (0.0005)	0.0099 (0.0005)	0.0183 (0.0017)	0.0069 (0.0005)	0.0315 (0.0031)	0.0319 (0.0026)
Squared Term of Put Option	0.0005 (0.0000)	0.0005 (0.0000)	0.0035 (0.0001)	0.0004 (0.0000)	0.0016 (0.0007)	0.0496 (0.0225)
State Unemployment Rate (percent)	6.8722 (2.5582)	6.9737 (2.6185)	6.7838 (3.4710)	6.5788 (2.2237)	6.6791 (2.8396)	7.5629 (3.0008)
State Divorce Rate (percent)	5.4457 (0.7988)	5.4353 (0.7715)	5.8359 (0.5072)	5.4550 (0.8805)	4.8546 (0.5683)	5.2625 (0.3117)
Initial Loan-To-Value Ratio (LTV)	0.7665 (0.0239)	0.7647 (0.0245)	0.8878 (0.0063)	0.7650 (0.0226)	-	-
Freq. of Call Option In-the-Money	-	-	-	-	1.2359 (12.085)	2.5238 (21.818)
No. of Observations	22,293	16,391	315	5,587	16,391	315

Note: Standard deviations are in parentheses.

* Other includes matured mortgages as well as those outstanding at the end of the observation period.

TABLE II.

MAXIMUM LIKELIHOOD ESTIMATES FOR COMPETING RISKS OF MORTGAGE PREPAYMENT AND DEFAULT WITHOUT UNOBSERVED HETEROGENEITY

	Model 1		Model 2	
	Prepay	Default	Prepay	Default
Call Option (fraction of contract value)	5.058 (126.05)	6.776 (19.23)	5.070 (124.62)	6.813 (18.94)
Put Option (probability of negative equity)	-5.805 (12.89)	13.956 (15.90)	-6.011 (12.55)	9.310 (9.42)
Squared Term of Call Option	1.780 (13.64)	1.124 (1.01)	1.808 (13.75)	1.040 (0.94)
Squared Term of Put Option	6.194 (10.25)	-14.347 (11.69)	6.486 (10.45)	-9.572 (7.41)
State Unemployment Rate (percent)	-0.046 (9.35)	0.035 (0.88)	-0.046 (9.17)	0.062 (1.52)
State Divorce Rate (percent)	-0.055 (5.24)	0.401 (3.76)	-0.057 (5.38)	0.395 (3.66)
0.6<LTV≤0.75			0.080 (3.02)	2.144 (2.85)
0.75<LTV≤0.8			0.074 (3.14)	2.490 (3.43)
0.8<LTV≤0.9			0.132 (5.16)	3.416 (4.74)
LTV>0.9			0.033 (1.05)	3.805 (5.27)
LOC1	0.411 (19.29)	0.293 (1.42)	0.385 (17.98)	0.014 (0.97)
Log Likelihood	-74,608		-74,523	

Note: T-ratios are in parentheses. All models are estimated by ML approach. Prepayment and default functions are considered as correlated competing risks and they are estimated jointly. LOC1 is the location parameter of the error distribution with a mass point of one.

TABLE III.

MAXIMUM LIKELIHOOD ESTIMATES FOR COMPETING RISKS

of Mortgage Prepayment and Default with Unobserved Heterogeneity

	Model 3		Model 4		Model 5		Model 6	
	Prepay	Default	Prepay	Default	Prepay	Default	Prepay	Default
Call Option (fraction of contract value)	5.755 (95.82)	7.256 (19.14)	5.750 (95.62)	7.255 (18.47)	5.764 (94.32)	7.131 (18.07)	5.759 (95.61)	7.146 (17.71)
Put Option (probability of negative equity)	-6.166 (13.23)	14.019 (15.67)	-6.471 (13.06)	9.350 (9.23)	-6.174 (13.15)	14.168 (15.62)	-6.422 (12.96)	9.646 (9.39)
Squared Term of Call Option	3.246 (18.72)	2.607 (2.13)	3.237 (18.67)	2.416 (1.95)	3.263 (18.70)	2.327 (1.87)	3.253 (18.75)	2.221 (1.77)
Squared Term of Put Option	6.627 (10.25)	-14.319 (11.04)	7.059 (10.65)	-9.504 (7.00)	6.639 (10.26)	-14.390 (11.12)	6.862 (10.65)	-9.750 (7.15)
State Unemployment Rate (percent)	-0.054 (10.69)	0.027 (0.67)	-0.053 (10.38)	0.056 (1.34)	-0.054 (10.68)	0.027 (0.66)	-0.053 (10.39)	0.057 (1.33)
State Divorce Rate (percent)	-0.078 (6.65)	0.378 (3.54)	-0.080 (6.81)	0.374 (3.44)	-0.078 (6.63)	0.385 (3.54)	-0.080 (6.77)	0.388 (3.49)
0.6<LTV≤0.75			0.068 (2.21)	2.125 (2.81)			0.068 (2.19)	2.216 (2.79)
0.75<LTV≤0.8			0.067 (2.44)	2.477 (3.38)			0.066 (2.39)	2.594 (3.34)
0.8<LTV≤0.9			0.124 (4.19)	3.396 (4.67)			0.122 (4.11)	3.523 (4.53)
LTV>0.9			0.053 (1.44)	3.807 (5.23)			0.048 (1.30)	3.960 (5.03)
LOC1	0.532 (16.08)	0.366 (1.41)	0.500 (15.09)	0.018 (0.96)	0.534 (15.52)	0.343 (1.33)	0.501 (14.91)	0.013 (0.87)
LOC2	0.056 (4.86)	0.080 (1.10)	0.053 (4.67)	0.004 (0.90)	0.056 (0.81)	0.008 (0.00)	0.052 (2.08)	0.000 (0.00)
LOC3					0.057 (0.31)	0.542 (0.55)	0.053 (0.50)	0.059 (0.66)
MASS2	0.058 (6.56)		0.058 (6.36)		0.040 (1.26)		0.044 (2.31)	
MASS3					0.020 (0.52)		0.015 (0.67)	
Log Likelihood	-74,471		-74,391		-74,470		-74,389	

Note: T-ratios are in parentheses. All models are estimated by ML approach. Prepayment and default functions are considered as correlated competing risks and they are estimated jointly. A bivariate distribution of unobserved heterogeneous error terms is also estimated simultaneously with the competing risks hazard functions. LOC1 and LOC2 are the location parameters of the error distribution. MASS1 and MASS2 are the mass points associated with LOC1 and LOC2, respectively. MASS1 is normalized to 1 during the estimation.

TABLE IV.

FIRST-STAGE ESTIMATES OF MORTGAGE PREPAYMENT RISK BASED ON A PROBIT MODEL

Variables	Coefficients
Call Option (fraction of contract value)	2.320 (88.24)
Put Option (probability of negative equity)	-2.360 (11.77)
Squared Term of Call Option	1.589 (26.76)
Squared Term of Put Option	2.589 (8.86)
State Unemployment Rate (percent)	-0.026 (12.81)
State Divorce Rate (percent)	-0.046 (10.22)
0.6<LTV≤0.75	0.030 (2.52)
0.75<LTV≤0.8	0.034 (3.17)
LTV>0.8	0.047 (4.21)
Constant	-1.473 (60.25)
Log Likelihood	-73,285

Note: The Probit Model is estimated using an event-history mortgage loan data set created by authors. Total numbers of events includes 804,582 records. The Probit Model is estimated by ML approach. T-ratios are in parentheses.

SECOND-STAGE ESTIMATES OF DISTRIBUTION OF UNOBSERVED HETEROGENEITY

Specification 1:

$$\hat{\varepsilon}_{it} = -\underset{(21.75)}{0.006084} - \underset{(31.19)}{0.002440} M_{it} + \underset{(29.65)}{0.000309} T_{it} + \nu_{it}$$

Specification 2:

$$\hat{\varepsilon}_{it} = -\underset{(62.10)}{0.013623} - \underset{(96.00)}{0.003584} M_{it} + \underset{(57.27)}{0.000776} T_{it} + \frac{\nu_{it}}{\sqrt{M_{it}/T_{it}}}$$

where $\hat{\varepsilon}_{it}$ is residual from the first-stage probit model estimation, M_{it} is the number of missed opportunities (i.e. the number of times that individual *i* has failed to exercise the prepayment option when it was in the money from origination to time *t*), T_{it} is age of the mortgage measured in quarters, ν_{it} is random variable following a standard normal distribution, T-ratios are in parentheses.

TABLE V.

SIMULATED MAXIMUM LIKELIHOOD ESTIMATES FOR COMPETING RISKS OF MORTGAGE PREPAYMENT AND DEFAULT WITH UNOBSERVED HETEROGENEITY

	$\hat{\varepsilon}_{it} = \alpha + \beta_1 M_{it} + \beta_2 T_{it} + v_{it}$				$\hat{\varepsilon}_{it} = \alpha + \beta_1 M_{it} + \beta_2 T_{it} + \frac{v_{it}}{\sqrt{M_{it}/T_{it}}}$			
	Model 7		Model 8		Model 9		Model 10	
	Prepay	Default	Prepay	Default	Prepay	Default	Prepay	Default
Call Option (fraction of contract value)	6.567 (95.14)	6.776 (19.25)	6.564 (94.96)	6.813 (18.98)	6.304 (97.99)	6.775 (19.23)	6.303 (97.78)	6.813 (18.95)
Put Option (probability of negative equity)	-5.212 (11.55)	13.960 (15.79)	-5.285 (11.06)	9.309 (9.39)	-4.776 (10.69)	13.959 (15.84)	-4.811 (10.14)	9.309 (9.40)
Squared Term of Call Option	5.082 (25.76)	1.127 (1.02)	5.080 (25.73)	1.040 (0.94)	4.653 (25.07)	1.127 (1.01)	4.657 (25.07)	1.039 (0.94)
Squared Term of Put Option	5.608 (8.93)	-14.354 (11.51)	5.746 (8.98)	-9.572 (7.35)	5.018 (8.09)	-14.351 (11.61)	5.121 (8.05)	-9.571 (7.38)
State Unemployment Rate (percent)	-0.040 (8.13)	0.035 (0.88)	-0.041 (8.10)	0.062 (1.52)	-0.032 (6.42)	0.035 (0.88)	-0.032 (6.47)	0.062 (1.52)
State Divorce Rate (percent)	-0.017 (1.54)	0.402 (3.77)	-0.018 (1.62)	0.396 (3.66)	0.006 (0.59)	0.401 (3.76)	0.006 (0.51)	0.396 (3.66)
0.6<LTV≤0.75			0.045 (1.60)	2.144 (2.85)			0.047 (1.74)	2.144 (2.85)
0.75<LTV≤0.8			0.033 (1.30)	2.490 (3.43)			0.033 (1.36)	2.490 (3.43)
0.8<LTV≤0.9			0.085 (3.17)	3.416 (4.74)			0.087 (3.36)	3.416 (4.74)
LTV>0.9			-0.004 (0.13)	3.805 (5.27)			-0.012 (0.37)	3.805 (5.27)
$\hat{\varepsilon}$	31.656 (28.34)		31.506 (28.08)		10.301 (28.49)		10.259 (28.28)	
LOC1	0.258 (17.67)	0.293 (1.42)	0.250 (16.70)	0.014 (0.98)	0.187 (16.72)	0.293 (1.42)	0.182 (15.99)	0.014 (0.97)
Log Likelihood	-74,117		-74,039		-74,136		-74,057	

Note: T-ratios are in parentheses. All models are estimated by a three-stage Simulated Maximum Likelihood Estimation (SMLE) approach. The simulated unobserved borrower heterogeneities were drawn from error distribution specification 1. Prepayment and default functions are considered as correlated competing risks and they are estimated jointly. LOC1 is the location parameter of the error distribution with a mass point of one.

TABLE VI.

COMPARISON OF MLE WITH SMLE

	MLE				SMLE			
	Model 11		Model 12		Model 13		Model 14	
	w/o Het.		w/ Het.		Error Spec. 1		Error Spec. 2	
	Prepay	Default	Prepay	Default	Prepay	Default	Prepay	Default
Call Option (fraction of contract value)	5.784 (78.85)	6.811 (12.25)	6.249 (70.76)	7.505 (12.67)	6.278 (72.39)	6.812 (12.27)	6.005 (73.66)	6.812 (12.26)
Put Option (probability of negative equity)	-6.051 (12.55)	9.338 (9.41)	-6.295 (12.81)	9.564 (9.50)	-5.268 (11.01)	9.336 (9.38)	-4.811 (10.12)	9.336 (9.39)
Squared Term of Call Option	3.223 (18.83)	0.989 (0.71)	4.169 (19.71)	2.989 (1.98)	4.676 (21.81)	0.991 (0.72)	4.289 (21.21)	0.990 (0.72)
Squared Term of Put Option	6.646 (10.55)	-9.617 (7.39)	6.904 (10.86)	-9.749 (7.32)	5.724 (8.92)	-9.615 (7.33)	5.145 (8.03)	-9.614 (7.36)
State Unemployment Rate (percent)	-0.049 (9.75)	0.063 (1.51)	-0.055 (10.60)	0.054 (1.27)	-0.039 (7.73)	0.063 (1.50)	-0.031 (6.07)	0.063 (1.51)
State Divorce Rate (percent)	-0.056 (5.19)	0.395 (3.60)	-0.072 (6.23)	0.374 (3.39)	-0.017 (1.53)	0.395 (3.61)	0.006 (0.57)	0.395 (3.61)
0.6<LTV≤0.75	0.070 (2.58)	2.147 (2.85)	0.064 (2.12)	2.130 (2.82)	0.045 (1.60)	2.147 (2.85)	0.046 (1.68)	2.147 (2.85)
0.75<LTV≤0.8	0.065 (2.68)	2.494 (3.43)	0.062 (2.33)	2.487 (3.40)	0.032 (1.29)	2.494 (3.43)	0.031 (1.27)	2.494 (3.43)
0.8<LTV≤0.9	0.121 (4.65)	3.418 (4.74)	0.118 (4.08)	3.397 (4.69)	0.084 (3.14)	3.418 (4.74)	0.084 (3.23)	3.418 (4.74)
LTV>0.9	0.030 (0.95)	3.807 (5.27)	0.049 (1.36)	3.814 (5.25)	-0.004 (0.11)	3.806 (5.27)	-0.011 (0.33)	3.807 (5.27)
Call_In-the-Money_1 (dummy)	0.130 (3.78)	0.157 (0.65)	0.078 (2.12)	0.052 (0.22)	0.145 (4.22)	0.157 (0.65)	0.256 (7.52)	0.157 (0.65)
Call_In-the-Money_2 (dummy)	0.051 (1.26)	-0.339 (1.10)	-0.025 (0.58)	-0.483 (1.55)	0.151 (3.68)	-0.339 (1.10)	0.279 (6.78)	-0.339 (1.10)
Call_In-the-Money_3 (dummy)	-0.466 (12.61)	0.021 (0.09)	-0.428 (10.54)	-0.063 (0.26)	0.149 (3.21)	0.021 (0.09)	0.104 (2.37)	0.021 (0.09)
$\hat{\varepsilon}$					32.856 (24.26)		10.577 (25.22)	
LOC1	5.784 (78.85)	6.811 (12.25)	0.484 (15.49)	0.018 (0.95)	0.238 (15.95)	0.014 (0.96)	0.171 (14.88)	0.014 (0.96)
LOC2			0.040 (3.08)	0.000 (0.00)				
MASS2			0.033 (4.96)					
Log Likelihood	-74,393		-74,307		-74,026		-74,016	

Note: T-ratios are in parentheses. Models 11 and 12 are estimated by MLE approach. Model 11 is estimated without heterogeneous error term. In Model 12, a bivariate distribution of unobserved heterogeneous error terms is estimated simultaneously with the competing risks hazard functions. Models 13 and 14 are estimated by SMLE. Model 13 assumes that the unobserved heterogeneous error term follows specification 1, and model 14 assumes that the unobserved heterogeneous error term follows specification 2.