Woodhead Behavior and the Pricing of Residential Mortgages*

by

Yongheng Deng
University of Southern California
Los Angeles, CA 90089-0626
email: ydeng@usc.edu

John M. Quigley
University of California
Berkeley, CA 94720-3880
email: quigley@econ.berkeley.edu

August 2001

Abstract

Mortgage terminations arise because borrowers exercise options. Empirically the extent to which the call is in the money 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 borrowers do not behave as "ruthlessly" as the theory predicts. This paper investigates the apparently irrational behavior of those borrowers who do not terminate their mortgages even when the option is deeply into the money.

We develop an option-based empirical model to analyze this phenomenon -- the behavior of irrational "woodheads." Of course we do not observe "woodheads" explicitly in any body of data. Instead, we analyze the correlates of unobserved heterogeneity within a large sample of mortgage holders. We extend SMLE techniques proposed by Stinebrickner (1999) to estimate the competing risks of mortgage prepayment and default, recognizing unobserved heterogeneity, which is due in part to the behavior of "woodheads." The extended model is clearly superior to alternatives on statistical grounds.

We then analyze the economic implications of this more powerful model. We analyze the predictions of the model for the valuation and pricing of mortgage pools and mortgage-backed securities. Based upon an extensive Monte Carlo simulation, we find that the SMLE model yields prices for seasoned mortgage pools that vary by as much as about forty basis points from more primitive estimates.

The results indicate the empirical importance of heterogeneity and the implications of non-optimizing behavior for the valuation and pricing of mortgages and mortgage-backed securities.

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

*A previous version of this paper was presented at the joint AFA/AREUEA session at the Allied Social Science Meetings, New Orleans, January 2001. We are grateful for the comments of Robert Van Order. Deng acknowledges financial support from Lusk Center for Real Estate at USC.

1. Introduction

The growth in the scale and complexity of the U. S. mortgage market since the securitization revolution of the 1980s has been enormous. The volume of residential mortgages outstanding nearly doubled during the 1990s, to over $5.6 trillion, and during the past five years originations of residential mortgages averaged more than $1.2 trillion annually. More than half of all new mortgages are securitized, and the volume of outstanding home mortgage securities currently exceeds $2.8 trillion. In comparison, the total U. S. Treasury debt held by the public is currently about $3.4 trillion.

This growth has generated enormous interest in the economics of mortgage markets. 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 many 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 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 mortgage holders. In the original applications of duration models to biostatistics 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, 1984).

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 the 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.

Methods for controlling for completely unobserved heterogeneity among borrowers 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 combining a prepayment hazard function which is homogeneous across agents with pool-specific transactions cost functions. An exogenous transactions cost function is assumed to follow a beta distribution which varies across individual mortgage pools. Of course, the ex post recognition of unobserved heterogeneity among agents provides no help at all in the pricing of mortgage pools (although models recognizing completely unobserved heterogeneity provide more efficient 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, together with recent results of Stinebrickner (1999), to develop a simulated maximum likelihood estimator (SMLE) of the proportional hazard model in the presence of 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, the model does not depend upon information obtained only after observing the behavior of an entire cohort of borrowers over the lifetimes of their contracts. Our 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 from more primitive models. In section IV we consider the pricing implications of these models.

2. 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:

$$S\left(t_p, t_d \mid r, H, Y, X, \xi_p, \xi_d, \theta\right)$$

(1)
$$= \exp\left\{ -\xi_p \sum_{k=1}^{t_p} \exp\left(\gamma_{pk} + \beta'_{p_1} g_{pk}\left(r, H, Y\right) + \beta'_{p_2} X\right) - \xi_d \sum_{k=1}^{t_d} \exp\left(\gamma_{dk} + \beta'_{d_1} g_{dk}\left(r, H, Y\right) + \beta'_{d_2} X\right) \right\}.$$

In this formulation $g_{jk}\left(r, H, Y\right)$ are time-varying measures of 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 by employing some standard such as "PSA experience." [1]

As noted above, a major impediment to analyzing the 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 ith borrower can be rewritten in following form:

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

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.

[1] The Public Securities Association (PSA) has defined a prepayment measurement standard which has been widely adopted by fixed-income securities analysts. This is a series of 360 monthly prepayment rates expressed as constant annual 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. (See Fabozzi, 1995, for details.) Prepayments are often reported as simple linear multiples of this schedule. Therefore, by adopting this PSA schedule as the baseline, the factors of proportionality estimated from the hazard model can be expressed simply as a percentage of the "PSA experience."

$t_i = \min\left(t_{pi}, t_{di}\right)$. 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*, and $F_c\left(t_i \mid \eta_{pi}, \eta_{di}\right)$ as the probability that mortgage duration data are censored for the *i*th borrower in period *t* due to the end of the data collection period, such that [2]

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

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

and

(6)
$$F_c\left(t_i \mid \eta_{pi}, \eta_{di}\right) = S\left(t_i, t_i \mid \eta_{pi}, \eta_{di}\right).$$

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

(7)
$$F_j\left(t_i\right) = \int_0^\infty \int_0^\infty F_j\left(t_i\right) dG\left(\eta_{pi}, \eta_{di}\right), \qquad j = p, d, c.$$

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

(8)
$$\log L = \sum_{i=1}^{N} \delta_{pi} \log\left(F_p\left(T_i\right)\right) + \delta_{di} \log\left(F_d\left(T_i\right)\right) + \delta_{ci} \log\left(F_c\left(T_i\right)\right),$$

---

[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 $\frac{1}{2}\left\{S\left(t_i, t_i \mid \eta_{pi}, \eta_{di}\right) + S\left(t_i+1, t_i+1 \mid \eta_{pi}, \eta_{di}\right) - S\left(t_i, t_i+1 \mid \eta_{pi}, \eta_{di}\right) - S\left(t_i+1, t_i \mid \eta_{pi}, \eta_{di}\right)\right\}$, which appears in both equations (4) and (5) is an adjustment for mortgage duration data measured in discrete rather than continuous time.

where N is the sample size and δ_{ji}, $j = p,d,c$, are indicator variables that take the value of one if the *i*th loan is terminated by prepayment, default, or censoring, respectively, and zero otherwise.

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 (7) can be obtained by substituting a consistent estimate $\hat{\eta}_{pi}$ for the unobserved η_{pi}.

Due to both the nonlinear nature of equation (7) and the dependence of the elements in η_{pi}, no analytic solution for equation (7) exists. However, Stinebrickner (1999) suggested that the integral can be simulated as

$$\begin{aligned} &L_i^s\left(r_i, H_i, Y_i, X_i, \theta, \mu_p, \mu_d, \hat{\eta}_{pi}\right) \\ &= \frac{1}{D}\sum_{d=1}^{D} L_i\left(r_i, H_i, Y_i, X_i, \theta, \mu_p, \mu_d, \eta_{pi}\left(\eta_{pi}^{*d}\right)\right), \end{aligned} \tag{9}$$

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 (7) with the simulated likelihood function as described in equation (9):

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, with parameter ϖ_{pi}. 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

---

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

[4] The assumption of log normal distribution is not at all necessary. The distribution can include any continuously differentiable, nonnegative aggregator function.

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.

Second, the residuals $\hat{\varepsilon}_i$ obtained from the probit estimates are regressed on the correlates of heterogeneity M_i such that

(10) $$\hat{\varepsilon}_i = f\left(M_i, \upsilon_i\right)$$

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

Third, the likelihood function specified in equation (8) is estimated by replacing $\hat{\eta}_{pi}$ with repeated draws of η_{pi}^{*d}, such that

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

where $\eta_i^{*d} = \exp\left(\hat{f}\left(M_i, \upsilon_i^d\right)\right)$, 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 (10). υ_i^d is the d^{th} draw from a standard normal distribution, $d = 1, \ldots, D$.

3. 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).

3.1. *The Data*

The database contains 1,489,372 observations on single family mortgage loans issued between 1976 and 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 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 those 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 first quarter of 1992. The analysis is confined to loans issued in 30 major metropolitan areas (MSAs)–a total of 447,042 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 in our analysis are those measuring the extent to which the put and call options are in the money and those reflecting the astuteness of borrowers. The current mortgage interest rate and the initial contract terms are sufficient to compute the extent to which the option is in the money. 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 another 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 the unpaid mortgage balance (computed from the contract terms), permit us to estimate the distribution of homeowner equity quarterly for each observation. In particular, the variable *"Put Option"* (i.e., an element of $g_{pk}\left(r,H,Y\right)$ in section II) measures the

[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, when 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 $V_{i,m}=\sum_{S=1}^{TM_i-k_i}1\big/\left(1+m_{j,\tau_i+k_i}\right)^S$.

probability that homeowner equity is negative, i.e., the probability that the put option is in the money.[6] The details of the calculation of these 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).

The correlate of the unobserved heterogeneity across borrowers (i.e., M in equation (10)) is computed in the following way. At each quarter since origination, we calculate whether the call option is in the money (this merely indicates whether current market interest rates on new first mortgages[7] are lower than the contract interest rate). We then compute a time-varying covariate for each borrower reflecting the number of quarters since origination that an in-the-money call was not exercised. A borrower who systemically passes up profitable opportunities to prepay the mortgage is more likely to be a "woodhead." Our measure, M, treats differences in "astuteness" among borrowers, in their "costs of calculation," and in their "transactions costs" as observationally equivalent.

We have computed the measure M at each quarter for each mortgage in the sample. There are a total of 16,044,963 of these event histories in our sample of mortgages. Table I summarizes some of this information.

Panel A presents the distribution of M among mortgages, separately for the full sample and for differently seasoned mortgage pools[8]. As the table indicates, for more than half of the mortgages in the sample, the borrower missed at least one profitable exercise of the call option. For mortgage pools seasoned ten years, about 55 percent of borrowers missed at least one opportunity. About 10 percent of borrowers in the sample

---

[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 $H_{i,k} = C_i \left(I_{j,\tau_i + k_i} \big/ I_{j,\tau_i} \right)$, where $I_{j,t}$ is the price index in metropolitan area j at time t and where the term in parentheses follows a log normal distribution. The "*Put Option*" variable is defined as: $Put\_Option_{i,k} = \Phi\left(\left(\log V_{i,m} - \log H_{i,k} \right) \big/ \sqrt{\omega^2} \right)$, where $\Phi(\cdot)$ is cumulative standard normal distribution function, ω^2 is an estimated variance, and $V_{i,m}$ is defined in footnote 5. The term ω^2 is defined more precisely in Deng et al., (2000).

[7] Interest rates on new mortgage contracts are available by quarter and region at http://www.freddiemac.com.

TABLE I.

NUMBER OF LOANS AND PAYABLE EVENTS BY NUMBER OF MISSED CALL OPTIONS

(Column Percentages in Parentheses)

	Full Sample	5-Year Seasoned Pool	10-Year Seasoned Pool
PANEL A – NUMBER OF LOANS			
M = 0	211,294 (47.3)	170,214 (46.5)	87,697 (45.7)
M = 1-2	75,449 (16.9)	63,021 (17.2)	34,352 (17.9)
M = 3-4	47,619 (10.7)	40,384 (11.0)	25,212 (13.1)
M = 5-8	42,052 (9.4)	30,671 (8.4)	10,887 (5.7)
M = 9-12	25,486 (5.7)	19,323 (5.3)	7,340 (3.8)
M ≥ 13	45,142 (10.1)	42,649 (11.6)	26,290 (13.7)
Total	447,042	366,262	191,778
PANEL B – NUMBER OF PAYABLE EVENTS			
M = 0	7,789,407 (47.2)	7,351,236 (47.3)	4,817,306 (47.3)
M = 1-2	2,825,602 (17.1)	2,707,112 (17.4)	1,786,059 (17.5)
M = 3-4	1,908,598 (11.6)	1,815,629 (11.7)	1,379,532 (13.5)
M = 5-8	1,315,863 (8.0)	1,150,199 (7.4)	569,271 (5.6)
M = 9-12	843,125 (5.1)	740,884 (4.8)	377,717 (3.7)
M ≥ 13	1,809,410 (11.0)	1,765,046 (11.4)	1,263,422 (12.4)
Total	16,492,005	15,530,106	10,193,307

[8] The five (ten) year seasoned pool is a sub-sample of mortgage loans which have durations greater than five (ten) years. Loosely speaking, the full sample may be interpreted as a pool containing the newly issued mortgage loans.

missed more than twelve profitable opportunities, while for ten-year seasoned mortgage pools, about 14 percent of borrowers missed more than twelve profitable opportunities. More seasoned mortgages are associated with larger numbers of missed opportunities to exercise profitable options. Panel B presents the number of payable events, separately for the full sample and for mortgage pools of different seasoning. The results are similar to those reported in Panel A.

Figure I presents the cumulative frequency of M among mortgages in these different pools. It shows again that more seasoned mortgages are associated with larger numbers of missed opportunities to exercise profitable options.

For more seasoned mortgages, at the time payable events occur, borrowers are more likely to have passed up a profitable opportunity to exercise options, and they are likely to have passed up more of these opportunities.

Table II presents the average top deciles of in the money calls (in percent) in the distribution of call options by different seasoning of mortgage pools. These averages are reported separately for borrowers who never passed up a profitable prepayment opportunity (M = 0) and for those who passed up one or two, three or four, five to eight, nine to twelve, and more than twelve profitable prepayment opportunities (M = 1-2, 3-4, 5-8, 9-12, 12+). The table reports two striking regularities.

First, for mortgages of given duration, the averages increase monotonically with M. Larger values of this variable are associated with much larger potential gains from exercise. The average gain from exercise in the top deciles is 15 percent for those who passed up one or two profitable prepayment opportunity, 22 percent for those who passed up three or four opportunities, 26 percent for those who passed up five to eight opportunities, 28 percent for those who passed up nine to twelve opportunities, and 36 percent for those who passed up more than twelve profitable opportunities to refinance. The pattern of average values is similar for mortgages of differing seasonings.

Second, the average values of the call option in the top deciles associated with a given non-zero value of M declines with mortgage seasoning. Foregoing one or two profitable refinance opportunities is associated with an average call option value of 15 percent in the full sample, and with an average value of 12 percent for five year seasoned

FIGURE I.
CUMULATIVE FREQUENCY OF MISSED CALL OPPORTUNITIES

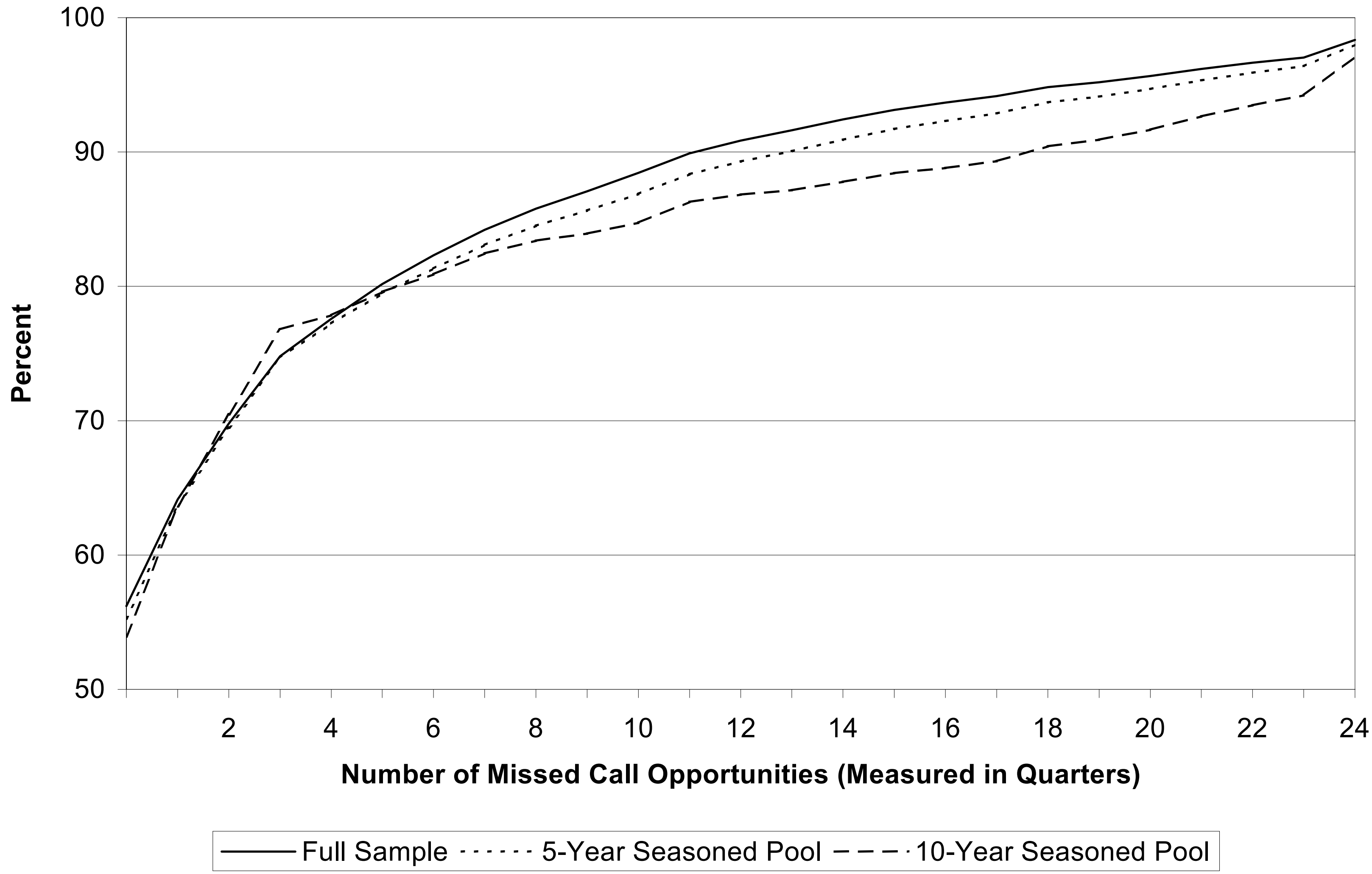


TABLE II.

AVERAGE TOP DECILES OF IN THE MONEY CALLS (IN PERCENT)

BY NUMBER OF MISSED CALL OPPORTUNITIES AND SEASONING OF MORTGAGE POOLS

	Full Sample	5-Year Seasoned Pool	10-Year Seasoned Pool
M = 0	2.89	2.89	2.89
M = 1-2	14.50	12.43	9.05
M = 3-4	21.60	20.85	12.60
M = 5-8	25.77	25.16	16.76
M = 9-12	27.79	27.15	18.70
M ≥ 13	36.00	35.45	33.15

Note: Table entries are the mean values of the extent to which the call-options in the top deciles of the distribution of call options are in the money at the time of termination. The sample consists of 447,042 mortgages. M measures number of quarters that the call option has been in-the-money but the borrower has not exercised the call option.

mortgage pools, and with an average value of 9 percent for ten year seasoned mortgage pools.

These regularities persist for other deciles of the distribution of call values and for other stratifications of mortgage duration.

As they season, mortgage pools are likely to contain larger proportions of borrowers who have foregone profitable refinance opportunities. The number of missed opportunities for profitable exercise is larger at the time of payable events in moreseasoned mortgages. The average value of the call option for those remaining in the pool is higher.

3.2. *Competing Risks Analysis*

Our competing risks analysis is based upon a five percent random sample of these mortgages–22,293 observations on mortgages in 30 MSAs.

Table III presents several 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.

TABLE III.

MAXIMUM LIKELIHOOD ESTIMATES FOR COMPETING RISKS OF MORTGAGE PREPAYMENT AND DEFAULT

	Model 1		Model 2		Model 3		Model 4	
	Prepay	Default	Prepay	Default	Prepay	Default	Prepay	Default
Call Option (fraction of contract value)	5.070	6.813	6.198	6.813	5.750	7.255	6.383	7.233
	(124.62)	(18.94)	(91.95)	(18.94)	(95.62)	(18.47)	(85.29)	(19.12)
Put Option (probability of negative equity)	-6.011	9.310	-6.646	9.310	-6.471	9.350	-6.716	9.511
	(12.55)	(9.42)	(13.80)	(9.41)	(13.06)	(9.23)	(13.77)	(9.53)
Call Option Squared	1.808	1.040	3.566	1.040	3.237	2.416	4.090	2.561
	(13.75)	(0.94)	(22.71)	(0.94)	(18.67)	(1.95)	(22.21)	(2.15)
Put Option Squared	6.486	-9.572	7.249	-9.572	7.059	-9.504	7.324	-9.679
	(10.45)	(7.41)	(11.65)	(7.38)	(10.65)	(7.00)	(11.66)	(7.34)
State Unemployment Rate (percent)	-0.046	0.062	-0.057	0.062	-0.053	0.056	-0.060	0.056
	(9.17)	(1.52)	(11.45)	(1.52)	(10.38)	(1.34)	(11.76)	(1.36)
State Divorce Rate (percent)	-0.057	0.395	-0.073	0.395	-0.080	0.374	-0.083	0.382
	(5.38)	(3.66)	(6.76)	(3.66)	(6.81)	(3.44)	(7.29)	(3.53)
0.6<LTV≤0.75	0.080	2.144	0.079	2.144	0.068	2.125	0.076	2.133
	(3.02)	(2.85)	(2.89)	(2.85)	(2.21)	(2.81)	(2.57)	(2.83)
0.75<LTV≤0.8	0.074	2.490	0.081	2.490	0.067	2.477	0.077	2.476
	(3.14)	(3.43)	(3.30)	(3.43)	(2.44)	(3.38)	(2.91)	(3.39)
0.8<LTV≤0.9	0.132	3.416	0.151	3.416	0.124	3.396	0.144	3.390
	(5.16)	(4.74)	(5.71)	(4.74)	(4.19)	(4.67)	(5.06)	(4.69)
LTV>0.9	0.033	3.805	0.062	3.805	0.053	3.807	0.071	3.798
	(1.05)	(5.27)	(1.91)	(5.27)	(1.44)	(5.23)	(2.00)	(5.24)
M			-0.034				-0.029	
			(18.75)				(13.80)	
Baseline Intercept	0.385	0.014	0.514	0.014				
	(17.98)	(0.97)	(17.27)	(0.97)				
Baseline Intercept (“ruthless”)					0.500	0.018	0.558	0.016
					(15.09)	(0.96)	(15.69)	(0.97)
Baseline Intercept (“woodheads”)					0.053	0.004	0.048	0.000
					(4.67)	(0.90)	(2.20)	(0.00)
Fraction “woodheads”					0.055		0.020	
					(6.36)		(3.26)	
Log Likelihood	-74,523		-74,352		-74,391		-74,300	

Note: T-ratios are in parentheses. All models are estimated by ML approach. Prepayment and default functions are modeled as correlated competing risks estimated jointly. In models 3 and 4, a bivariate distribution of unobserved heterogeneous error terms is also estimated simultaneously with the competing risks hazard functions. This distribution identifies separately the baselines of the two groups and estimates the fraction of the population in each group.

In addition to the variables measuring the value of the options, Model 1 includes state average unemployment and divorce rates as well as the initial loan-to-value ratio (LTV), in four categories. We use the PSA experience as the baseline function for the prepayment equation. The baseline function for the default equation is estimated non parametrically. Thus, the row labeled "Baseline" in the table reports the shift in the baseline prepayment consistent with the parametric form imposed by the PSA standard for prepayment. For the default function, the column reports that the non parametric form of the baseline, estimated according to equation (2), requires no shift.

The results confirm the importance of in-the-money options 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; ceteris paribus, those with higher LTVs are more likely to exercise options.

Model 2 adds the number of missed opportunities, M, a time-varying covariate, to the prepayment function in this model. This specification is analogous to those used by financial analysts in estimating prepayment rates for mortgage pools. [9] The variable is highly significant statistically. Accounting for heterogeneity among borrowers in this way increases the magnitude of the options-related variables and improves the overall fit of the model.

Models 3 and 4 present the same two specifications in a model of unobserved heterogeneity. In each model, we allow for the possibility that there are two distinct groups of borrowers, call them "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. For each model, we estimate the distribution of unobserved heterogeneity and the average behavior of the two groups jointly with the competing risk functions. Model 3 is a generalization of Model 1. Model 4 adds the variable M to the analysis.

The magnitude of the option values increases substantially when unobserved heterogeneity is accounted for. The magnitudes of the other variables change very little.

The variable M is highly significant, even in the model which accounts for unobserved heterogeneity by classifying borrowers into distinct groups.

There is a substantial difference between the two groups, however, in their exercise of the prepayment option. Those in the high risk group are about 10 times riskier (e.g., 0.500 versus 0.053 or 0.558 versus 0.048) than borrowers in the low risk group. This difference is highly significant. For the default option, there is little difference. For models 3 and 4, 95 to 98 percent of all borrowers are classified into the high risk "ruthless" group.

We now exploit additional information in the estimation of this model, namely 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 payment event 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. The results of the probit estimation using these quarterly observations on prepayment are reported in the appendix. The results are qualitatively similar to those reported in Table III 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 (10):

Specification 1: $$\hat{\varepsilon}_{it} = \alpha + \beta_1 M_{it} + \beta_2 T_{it} + \nu_{it}$$

Specification 2: $$\hat{\varepsilon}_{it} = \alpha + \beta \frac{M_{it}}{T_{it}} + \frac{\nu_{it}}{\sqrt{M_{it}/T_{it}}}$$

where $\hat{\varepsilon}_{it}$ is the residual from the probit model (reported in Appendix Table A), M_{it} is the number of missed opportunities (i.e. the number of times that individual i has failed to

[9] In some models employed by financial analysts, a variable measuring the spread between contract and current interest rates is employed, as a measure of the "burnout" of prepayment in pools of mortgages. See, for example, Richard and Roll (1989) and Hall (2000).

exercise the prepayment option when it was in the money from origination up to time *t*), T_{it} is the age of the mortgage 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 distribution with a potentially different mean for each individual, while specification 2 permits the unobserved error components of the hazard function to follow a normal distribution with a different variance as well as a different mean across individuals.

The estimated coefficients are highly significant in each specification:

Specification 1: $$\hat{\varepsilon}_{it} = \underset{(22.01)}{-0.006157} - \underset{(23.06)}{0.000897} M_{it} + \underset{(32.76)}{0.000357} T_{it} + \nu_{it}$$

Specification 2: $$\hat{\varepsilon}_{it} = \underset{(5.42)}{0.001291} - \underset{(51.23)}{0.026748} \frac{M_{it}}{T_{it}} + \frac{\nu_{it}}{\sqrt{M_{it}/T_{it}}}$$

where *t* ratios are in parentheses.

Table IV 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 by SMLE. The table reports the coefficients of the likelihood function, equation (8), estimated using the simulated likelihood function, equation (11) based upon 10,000 draws. Model 5 reports the results of specification 1; model 6 reports the more general specification.

A comparison with Table III, in which heterogeneity is specified as two 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 4 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 –74,074 (Model 5) as compared with a value of –74,391 for the discrete measure (Model 3). For specification 2, the value of the log likelihood is even larger for the continuous measure (-73,987).

TABLE IV.

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

	Model 5		Model 6	
	Prepay	Default	Prepay	Default
Call Option (fraction of contract value)	6.463 (102.44)	6.813 (18.91)	6.402 (117.60)	6.813 (18.83)
Put Option (prob. of negative equity)	-5.131 (10.83)	9.310 (9.40)	-5.761 (12.27)	9.310 (9.38)
Call Option Squared	4.593 (27.53)	1.040 (0.93)	3.804 (26.33)	1.040 (0.93)
Put Option Squared	5.466 (8.67)	-9.572 (7.38)	6.315 (10.14)	-9.572 (7.35)
State Unemployment Rate (percent)	-0.035 (7.06)	0.062 (1.52)	-0.043 (8.62)	0.062 (1.52)
State Divorce Rate (percent)	-0.005 (0.50)	0.395 (3.66)	-0.005 (0.46)	0.395 (3.66)
0.6<LTV≤0.75	0.059 (2.23)	2.144 (2.85)	0.069 (2.63)	2.144 (2.85)
0.75<LTV≤0.8	0.050 (2.12)	2.490 (3.43)	0.058 (2.54)	2.490 (3.43)
0.8<LTV≤0.9	0.120 (4.74)	3.416 (4.74)	0.162 (6.50)	3.416 (4.74)
LTV>0.9	0.022 (0.69)	3.805 (5.27)	0.068 (2.29)	3.805 (5.27)
$\hat{\varepsilon}$	42.770 (28.41)		26.471 (32.26)	
Baseline Intercept	0.209 (16.65)	0.014 (0.97)	0.165 (16.42)	0.014 (0.97)
Log Likelihood	-74,074		-73,987	

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 in Models 5 and 6 were drawn from error specifications 1 and 2, respectively. Prepayment and default functions are modeled as correlated competing risks estimated jointly.

3.3. *Summary*

Table V presents a summary and comparison of the SMLE model with other less general models of mortgage holder behavior. Model 1 reports the MLE estimates of the competing risks model assuming no unobserved heterogeneity across borrowers. Model

TABLE V.

COMPARISON OF MLE WITH SMLE

	MLE (w/o Het.)				MLE (w/ Het.)		SMLE (w/ Het.)	
	Model 1		Model 2		Model 4		Model 6	
	Prepay	Default	Prepay	Default	Prepay	Default	Prepay	Default
Call Option (fraction of contract value)	5.070 (124.62)	6.813 (18.94)	6.198 (91.95)	6.813 (18.94)	6.383 (85.29)	7.233 (19.12)	6.402 (117.60)	6.813 (18.83)
Put Option (probability of negative equity)	-6.011 (12.55)	9.310 (9.42)	-6.646 (13.80)	9.310 (9.41)	-6.716 (13.77)	9.511 (9.53)	-5.761 (12.27)	9.310 (9.38)
Call Option Squared	1.808 (13.75)	1.040 (0.94)	3.566 (22.71)	1.040 (0.94)	4.090 (22.21)	2.561 (2.15)	3.804 (26.33)	1.040 (0.93)
Put Option Squared	6.486 (10.45)	-9.572 (7.41)	7.249 (11.65)	-9.572 (7.38)	7.324 (11.66)	-9.679 (7.34)	6.315 (10.14)	-9.572 (7.35)
State Unemployment Rate (percent)	-0.046 (9.17)	0.062 (1.52)	-0.057 (11.45)	0.062 (1.52)	-0.060 (11.76)	0.056 (1.36)	-0.043 (8.62)	0.062 (1.52)
State Divorce Rate (percent)	-0.057 (5.38)	0.395 (3.66)	-0.073 (6.76)	0.395 (3.66)	-0.083 (7.29)	0.382 (3.53)	-0.005 (0.46)	0.395 (3.66)
0.6<LTV≤0.75	0.080 (3.02)	2.144 (2.85)	0.079 (2.89)	2.144 (2.85)	0.076 (2.57)	2.133 (2.83)	0.069 (2.63)	2.144 (2.85)
0.75<LTV≤0.8	0.074 (3.14)	2.490 (3.43)	0.081 (3.30)	2.490 (3.43)	0.077 (2.91)	2.476 (3.39)	0.058 (2.54)	2.490 (3.43)
0.8<LTV≤0.9	0.132 (5.16)	3.416 (4.74)	0.151 (5.71)	3.416 (4.74)	0.144 (5.06)	3.390 (4.69)	0.162 (6.50)	3.416 (4.74)
LTV>0.9	0.033 (1.05)	3.805 (5.27)	0.062 (1.91)	3.805 (5.27)	0.071 (2.00)	3.798 (5.24)	0.068 (2.29)	3.805 (5.27)
M			-0.034 (18.75)		-0.029 (13.80)			
$\hat{\varepsilon}$							26.471 (32.26)	
Baseline Intercept	0.385 (17.98)	0.014 (0.97)	0.514 (17.27)	0.014 (0.97)			0.165 (16.42)	0.014 (0.97)
Baseline Intercept ("ruthless")					0.558 (15.69)	0.016 (0.97)		
Baseline Intercept ("woodheads")					0.048 (2.20)	0.000 (0.00)		
Fraction "woodheads"					0.020 (3.26)			
Log Likelihood	-74,523		74,352		-74,300		-73,987	

Note: T-ratios are in parentheses. Models 1, 2, and 4 are estimated by MLE approach. In Model 4, a bivariate distribution of unobserved heterogeneous error terms is estimated simultaneously with the competing risks hazard functions. This distribution identifies separately the baselines of the two groups and estimates the fraction of the population in each group. Model 6 is estimated by SMLE with a simulated heterogeneous error distribution following specification 2.

2 adds the number of missed opportunities M to the analysis as an exogenous time-varying covariate. Model 4 assumes in addition a bivariate distribution of unobserved heterogeneous error terms and estimates that distribution simultaneously with the competing risks functions by MLE. Model 6 reports the results when it is assumed that the number of missed opportunities M is correlated with unobserved individual heterogeneity and when the model incorporating this is estimated by simulated MLE methods. Differences in values of log likelihood function are substantial. The SMLE model is clearly superior on statistical grounds.

4. PRICING IMPLICATIONS

In this section we evaluate the economic importance of our more general model in pricing mortgages, pools of mortgages, or mortgage-backed securities. We adopt a Monte Carlo simulation pricing approach to simulate prices for mortgage pools. We implement this simulation using the dynamic term structure model recently proposed by Dai and Singleton (2000). This model, a three-factor affine term structure model (ATSM), attempts explicitly to balance the requirements of precision in econometric representation of the state variables and the computational burdens of pricing and estimation. The Dai-Singleton (DS) model consists of a specific stochastic long run mean and volatility of interest rates, affine functions of risk-neutral drift factors. The basic model we use is the DS generalization of the term structure model[10] of Balduzzi, Das, Foresi and Sundaram (1996). Appendix Figure B reports the average path of simulated interest rates over thirty years using these equations and parameters.

In our application, we simulate 3 million short rates over a thirty-year period at intervals of 10^{-5} year. We then randomly sample 2,000 quarterly interest rate paths over the thirty-year period. These 2,000 randomly sampled interest rate paths are applied to the prepayment and default functions reported in Table V to compute the quarterly prepayment and default risks of the mortgage pools. Finally, the prepayment and default risk-adjusted mortgage amortization cash flows are discounted back using the randomly sampled 2,000 interest rate paths.

[10] Our simulation is based upon equation (23) of DS, using the parameters reported by DS in Table II, Column 2.

Table VI summarizes the relative pricing differences of mortgage pools at three assumed contract interest rates. The first column presents the mean percentage difference in equilibrium prices between model 2 and model 6. Model 2 represents heterogeneity by the inclusion of the variable M directly in the competing risks model.[11] The second column presents the percentage differences in prices between model 4 and model 6. Model 4 specifies the unobserved heterogeneity among borrowers in two categories and also includes the variable M directly in the computing risks model.[12]

The simulations are reported separately for the mortgage pools with a coupon interest rate of 8.25 percent, 8.75 percent, and 9.25 percent. The DS interest rate term structure used in our simulation has a long run mean of 8.27 percent.[13] We assume the average initial loan-to-value ratio is eighty percent, and the unemployment rate and divorce rate are the sample average, seven percent and six percent, respectively. We use the distribution of M observed from the sample, reported in Table I, Panel A, as the basis for the simulation. For example, in the simulations using the full sample, we assume 47.3 percent of mortgages have never missed a single profitable call opportunity, 16.9 percent of mortgages have M = 1.5, 10.7 percent have M = 3.5, 9.4 percent have M = 6.5, 5.7 percent have M = 10.5 and 10.1 percent have M = 14.

M, of course, varies with duration. But a mortgage pool manager does not observe the time-varying path of M ex ante. She can observe directly, however, the distribution of M from various seasoned mortgage pools. As reported in Table I, this distribution is skewed to the right as a mortgage pool seasons, since the remaining sample in a seasoned pool tends to be less risky in exercising the refinance option.

To produce the comparisons reported in Table VI, we first estimate the prepayment and default risks of each mortgage pool based on the parameters of Models 2, 4, and 6, the distribution of M and the stochastic term structure simulated using the three-factor ATSM. We then compute the cash flows for each mortgage pool. Finally, we compute the equilibrium price of each pool using the 2,000 randomly sampled interest

[11] Model 2 is, perhaps, close to the representation of heterogeneity which might appear in models used by practitioners to price mortgage pools. See Richard and Roll (1989) for a discussion.
[12] Model 4 is thus a somewhat more general specification than in Deng et al. (2000).
[13] The long run mean and other parameters of the interest rate term structure used in the simulation are based on estimated parameters reported by Dai and Singleton (2000), in Table II, Column 2.

TABLE VI.

MEAN PERCENTAGE DIFFERENCES IN EQUILIBRIUM PRICES SIMULATED BY ECONOMETRIC MODELS AT VARIOUS COUPON INTEREST RATES

(t-ratios are in parenthesis)

	Model 2 vs. SMLE Model	Model 4 vs. SMLE Model
A. 8.25 PERCENT		
Full Sample	0.40% (685)	0.18% (40)
5-Year Seasoned Pool	0.09 (348)	-0.02 (30)
10-Year Seasoned Pool	-0.15 (597)	-0.19 (722)
B. 8.75 PERCENT		
Full Sample	0.24 (355)	0.01 (3)
5-Year Seasoned Pool	-0.02 (41)	-0.13 (148)
10-Year Seasoned Pool	-0.22 (643)	-0.25 (753)
C. 9.25 PERCENT		
Full Sample	0.08 (96)	-0.14 (37)
5-Year Seasoned Pool	-0.13 (242)	-0.24 (246)
10-Year Seasoned Pool	-0.28 (668)	-0.31 (765)

Note: The percentage price differences reported in Table VI are measured by the price differences implied by different econometric models divided by simulated pool price based on the SMLE model. The Simulated prices use the volatility parameter of 1.5 percent, used by Dai and Singleton and the other parameters reported in Dai and Singleton (2000) Table II, Column 2, page 1964.

rate paths over the distribution of M. The detailed estimates are reported in the appendix.[14] As the comparison in the table shows, the pricing differences estimated from

[14] Appendix Table C reports the means and t-statistics of differences in equilibrium prices of one million dollar mortgage pools based on the estimated prepayment and defaults implied by Model 2, Model 4, and Model 6. The first column presents the average price differences between Model 2 and Model 6. The second column presents the mean absolute price differences between Model 4 and Model 6. The simulation

different models are quite large. For more seasoned pools, both Model 2 and Model 4 tend to under price the mortgages compared to the model estimated using SMLE approach. The comparison also indicates that the gain from the SMLE estimation technique is substantially larger when the mortgage pool is exposed to a higher risk of in-the-money call exercise. For example, for a ten-year seasoned mortgage pool with a weighted average coupon rate of 8.25 percent, the pricing differences between SMLE model and the more primitive models are between fifteen to nineteen basis points. For a mortgage pool with weighted average coupon rate of 9.25 percent, the pricing differences are as much as thirty basis points.

The SMLE estimation technique provides a substantially better fit to the data, as noted above, suggesting that it is a superior technique for analyzing heterogeneity in the behavior of mortgage borrowers. In addition, however, it has substantially different pricing implications, by as much as twenty to forty basis points. This suggests that the SMLE model has practical importance for the pricing and valuation of mortgage and mortgage-backed securities.

5. CONCLUSION

The mortgage market is large and has grown greatly in importance recently. The outstanding volume of residential mortgages will soon exceed the stock of outstanding U.S. government debt.

Contingent claims theory provides a coherent framework for the analysis of the financial behavior of the economic actors who hold these outstanding residential mortgages. As an empirical matter, however, mortgage holders do not behave as ruthlessly as the theory predicts. This has implications for the pricing of mortgage pools and mortgage-backed securities in addition to the well being of borrowers.

This paper develops a simulated maximum likelihood estimator to analyze the importance and extent of non-ruthless behavior in the market. The model uses information on behavioral correlates of heterogeneity among borrowers to extend the competing risks model of mortgage termination. Analysis based upon a large sample of

---

results reported in Appendix Table C use the volatility parameter and the other parameters reported by DS. In addition, we conducted this analysis with several different assumptions about volatility. The qualitative mortgages suggests that this method offers advantages in precision in comparison with conventional methods.

The SMLE approach also supports the real-time pricing of pools of mortgage or mortgage-backed securities. An extensive Monte Carlo simulation indicates that the pricing implications of the SMLE model are substantially different, at least for this body of data. This suggests that the model may have considerable practical application as well.

University of Southern California, Los Angeles, CA 90089-0626, U.S.A; ydeng@usc.edu,
University of California, Berkeley, CA 94720-3880, U.S.A.; quigley@econ.berkeley.edu.

nature of the results was not affected by changes in assumed volatility.

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.

Dai, Qiang and Kenneth J. Singleton (2000): “Specification Analysis of Affine Term Structure Models,” *Journal of Finance*, 55(5) 1943-1978.

Balduzzi, Pierluigi, Sanjiv R. Das, Silverio Foresi, and Rangarajan K. Sundaram (2000): “A Simple Approach to Three Factor Affine Term Structure Models,” *Journal of Fixed Income*, 6, 43-53.

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.

Fabozzi, Frank J., editor (1195) *Handbook of Mortgage Backed Securities*, fourth edition, New York: McGraw-Hill.

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 Estimating Mortgage Prepayment Models,” *Journal of Housing Economics*, 9(4), 215-232.

Heckman, James J. and B. Singer (1984): “Econometric Duration Analysis,” *Journal of Econometrics,* 24, 63-132.

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.

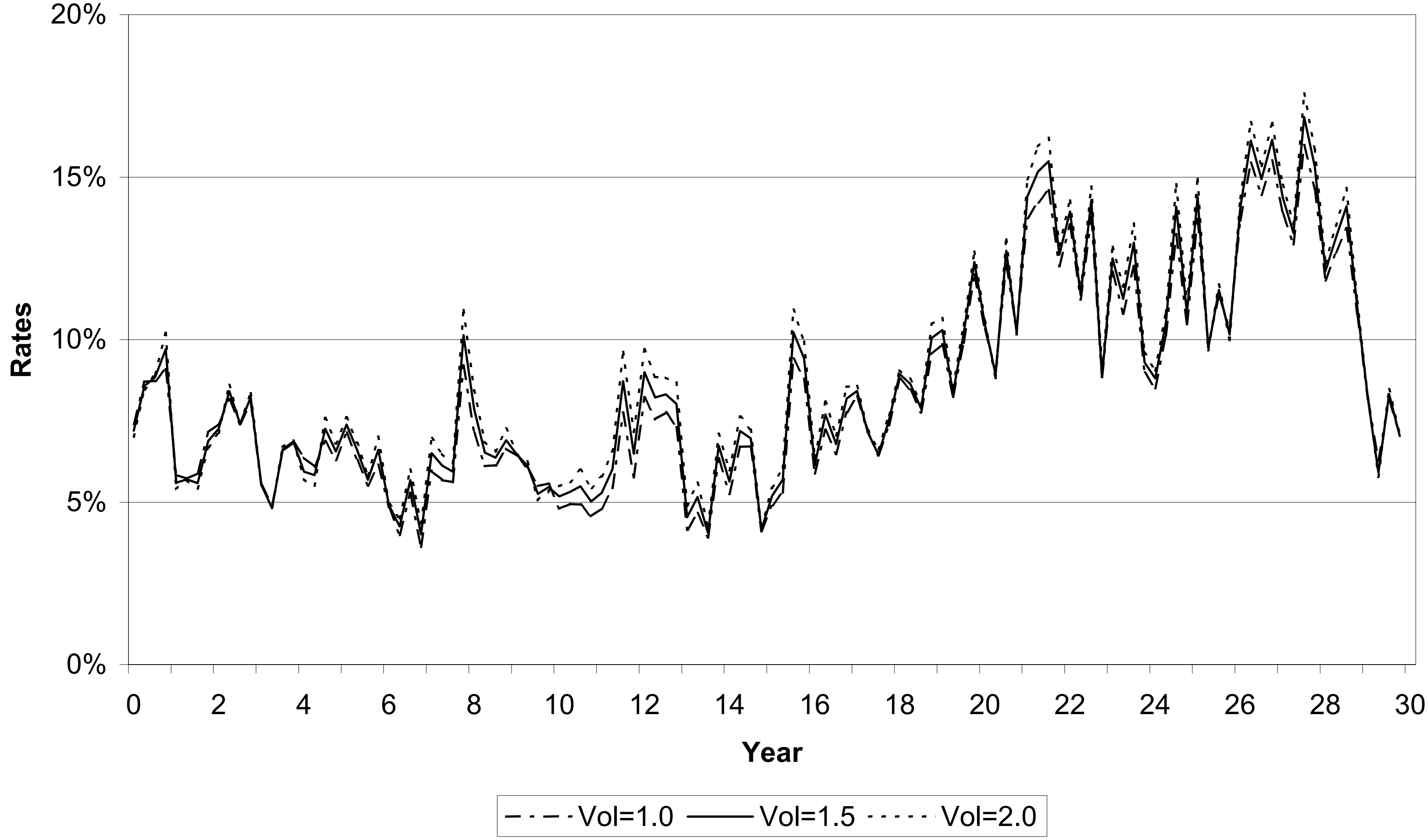
APPENDIX TABLE A.

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)
Call Option Squared	1.589 (26.76)
Put Option Squared	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 from the sample of 22,293 mortgages. The total number of events is 804,582, one for each payment event of each mortgage. The Probit Model is estimated by ML. T-ratios are in parentheses.

APPENDIX FIGURE B.
SIMULATED INTEREST RATES (ATSM)



Note: Average of 2,000 Interest rate paths simulated from Dai and Singleton (2000), equation (23) using parameters reported in Table II, Column 2 of Dai and Singleton. Interest rate paths are simulated for three volatility assumptions. The calculations in Table VI are based on a volatility of 1.5.

APPENDIX TABLE C.

MEAN DIFFERENCES IN EQUILIBRIUM PRICES SIMULATED FROM ONE MILLION DOLLAR MORTGAGE POOLS AT DIFFERENT COUPON INTEREST RATES

(t-ratios are in parentheses)

	Model 2 vs. SMLE Model	Model 4 vs. SMLE Model
A. 8.25 PERCENT		
Full Sample	$4,252 (678)	$1,901 (40)
5-Year Seasoned Pool	1,005 (347)	-263 (30)
10-Year Seasoned Pool	-1,732 (597)	-2,145 (719)
B. 8.75 PERCENT		
Full Sample	2,698 (354)	129 (3)
5-Year Seasoned Pool	-174 (41)	-1,515 (148)
10-Year Seasoned Pool	-2,518 (644)	-2,878 (752)
C. 9.25 PERCENT		
Full Sample	957 (96)	-1,669 (37)
5-Year Seasoned Pool	-1,484 (242)	-2,840 (246)
10-Year Seasoned Pool	-3,373 (671)	-3,659 (767)

Note: Simulated prices reported in Table C use the volatility parameter of 1.5 percent, used by Dai and Singleton and the other parameters reported in Dai and Singleton (2000) Table II, Column 2, page 1964.