

# Woodhead Behavior and the Pricing of Residential Mortgages

YONGHENG DENG and JOHN M. QUIGLEY \*

December 2002

\* Deng is from University of Southern California, 650 Childs Way, RGL 328, Los Angeles, CA 90089-0626, Tel: (213) 821-1030, Fax: (213) 740-6170, Email: [ydeng@usc.edu](mailto:ydeng@usc.edu); Quigley is from University of California, Berkeley, CA 94720-3880, Tel: (510) 643-7411, Fax: (510) 643-7357, Email: [quigley@econ.berkeley.edu](mailto:quigley@econ.berkeley.edu). We are grateful for the comments of Robert Van Order. Deng acknowledges financial support from the Lusk Center for Real Estate at USC; Quigley acknowledges financial support from the Berkeley Program on Housing and Urban Policy.

## **Woodhead Behavior and the Pricing of Residential Mortgages**

### **ABSTRACT**

Mortgage terminations arise because borrowers exercise options. This paper investigates the apparently irrational behavior of those borrowers who do not terminate their mortgages even when the exercise value of the option is deeply in 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 develop an error correction maximum likelihood (ECML) estimator using martingale transforms 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 ECML model yields prices for seasoned mortgage pools that deviate by 0.3 to almost 3.0 percent 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, martingale transform.

JEL Codes: G21, R2, D12

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 in our companion paper (See Deng et al., 2000).<sup>1</sup> 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

---

<sup>1</sup> Archer et al. (1996) and others have analyzed observed heterogeneity among borrowers in reduced form framework.

heterogeneity among agents provides no help at all in the pricing of mortgage pools (although models recognizing completely unobserved heterogeneity do 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 the development of martingale transforms (Barlow and Prentice, 1988, and Therneau et al., 1990), to develop an error correction maximum likelihood approach (ECML) for the proportional hazard model in the presence of heterogeneity among mortgage holders. The model we develop is completely general and need not be limited by specifying an arbitrary constellation of mass-points 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. The model developed here permits spot prices to be updated continuously with the information revealed by the behavior of borrowers holding unmatured 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 I below we sketch out the basic model and the estimation strategy employed. In section II 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 III we consider the pricing implications of these models.

## **I. The Model**

The proportional hazard model introduced by Cox (1972) 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  $\eta_p, \eta_d, r, H, Y$ , and  $X$  can be expressed in the following form:

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

In this formulation  $g_{jk}(r, H, Y)$  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.  $\eta_p$  and  $\eta_d$  are unobserved error terms associated with the hazard functions for prepayment and default respectively.  $\theta$  is a vector of parameters (e.g.,  $\gamma$  and  $\beta$ ) of the hazard function.  $\gamma_{jk}$  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 and SDA experience.”<sup>2</sup>

---

<sup>2</sup> 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

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. The errors,  $(\eta_p, \eta_d)$  in equation (1) may be decomposed into two parts:  $(\mu_p, \mu_d)$  and  $(\xi_{pi}, \xi_{di})$ . The first  $(\mu_p, \mu_d)$  is a pair of fixed-effect error terms representing, for example, a proportionate shift in the non parametric or PSA and SDA experience baselines. The second  $(\xi_{pi}, \xi_{di})$  is a pair of random-effect error terms representing unobservable borrower-specific heterogeneities. The joint survivor function for the  $i$ th borrower can be rewritten in the following form:

$$\begin{aligned}
& S(t_{pi}, t_{di} | r_i, H_i, Y_i, X_i, \mu_p, \mu_d, \xi_{pi}, \xi_{di}, \theta) \\
& = \exp \left\{ -\mu_p \xi_{pi} \sum_{k=1}^{t_{pi}} \exp(\gamma_{pk} + \beta'_{p1} g_{pki}(r_i, H_i, Y_i) + \beta'_{p2} X_i) \right. \\
& \quad \left. - \mu_d \xi_{di} \sum_{k=1}^{t_{di}} \exp(\gamma_{dk} + \beta'_{d1} g_{dki}(r_i, H_i, Y_i) + \beta'_{d2} X_i) \right\}.
\end{aligned} \tag{3}$$

Assume the  $\xi_i$ 's are iid with mean 1 and some unknown variance. It is obvious that, ceteris paribus, if  $\xi_i$  takes a value greater than 1, the  $i$ th individual has a higher risk of termination than the sample average. Conversely, if  $\xi_i$  takes a value less than 1, the  $i$ th individual has a lower risk of termination than the sample average.

---

maturity. Similarly the Bond Market Association has developed a Standard Default Assumption (SDA) that is widely used as a benchmark to measure loan default experience. The SDA series begins at 0.02 percent annual constant rate in the first month and increases by 0.02 percent in each successive month until month 30, when the series levels out at 0.6 percent per year for the next 30 months. Then the series declines by 0.0095 percent each month from month 61 to month 120. At that point, the default rate remains level through maturity. (See Hayre [2001] pp. 168-169 for details.) Prepayments and defaults are often reported as simple linear multiples of PSA and SDA schedules, respectively. Therefore, by adopting the PSA and SDA schedules as the baselines, the factors of proportionality estimated from the hazard model can be expressed simply as a percentage of the “PSA and SDA experience.”

Consider  $(\xi_{pi}, \xi_{di})$  in expression (3) as missing data. Stinebrickner (1999) has noted that consistent evaluation of expression (3) can be obtained by substituting a consistent estimate  $(\hat{\xi}_{pi}, \hat{\xi}_{di})$  for the unobserved  $(\xi_{pi}, \xi_{di})$ .

The development of counting process theory and martingale-based residuals in survival models provides a useful instrument for the unobserved  $(\xi_{pi}, \xi_{di})$ . Following Barlow and Prentice (1988), and Therneau et al. (1990), the martingale residual for the  $i$ th individual is defined as

$$\begin{aligned}\hat{M}_i(t) &= N_i(t) - \hat{E}_i(t) \\ &= N_i(t) - \int_0^t Y_i(s) \exp(X_i(s) \hat{\beta}) d\hat{h}_0(s),\end{aligned}\tag{4}$$

where  $N_i(t)$  takes a value 1 at time  $t$  if individual  $i$  has experienced the event of interest and 0 otherwise;  $Y_i(s)$  is a censor indicator that takes value 1 if individual  $i$  has survived up to time  $s$ , and 0 otherwise;  $h_0(s)$  is the baseline hazard function.<sup>3</sup>

The martingale residuals can be interpreted as the difference over time in the observed number between events and the expected number of events. In other words, the martingale residuals are an estimate of the excess number of events observed in the data but not predicted by the model.

However, the distribution of the martingale residual is highly skewed.<sup>4</sup> But well-defined procedures (e.g., martingale transforms) are available to convert the residuals to approximate the normal distribution.

---

<sup>3</sup> The term ‘‘martingale residuals’’ is motivated by the property that, if the true value of  $\beta$  and  $h_0(s)$  were used in Equation (4), then the function  $M_i$  would be martingale.

<sup>4</sup> From Equation (4), it is clear that martingale residuals are distributed between  $-\infty$  to 1.



More concretely, consider a three-stage approach to estimate a random-effect error correction maximum likelihood survival function of the competing risks hazard model, by replacing  $(\xi_{pi}, \xi_{di})$  in expression (3) with a set of consistent estimates  $(\hat{\xi}_{pi}, \hat{\xi}_{di})$ :

In the first stage, estimate the competing risks hazard model of mortgage prepayment and default, and obtain the martingale residuals,  $M_{pi}$  and  $M_{di}$ , for the prepayment and default functions, respectively. In the second stage, regress the martingale transform residuals,  $M_{pi}^X$  and  $M_{di}^X$ , on a set of observed covariates from the data. (See Appendix A for details of the martingale transform procedures.) In the third stage, correct the random-effect errors from expression (3); simply replace  $(\xi_{pi}, \xi_{di})$  by the exponential of the predicted martingale residuals  $(\exp(\hat{M}_{pi}), \exp(\hat{M}_{di}))$ .

$$\begin{aligned}
& S(t_{pi}, t_{di} | r_i, H_i, Y_i, X_i, \mu_p, \mu_d, \xi_{pi}, \xi_{di}, \theta) \\
& = \exp \left\{ -\mu_p \exp(\hat{M}_{pi}) \sum_{k=1}^{t_{pi}} \exp(\gamma_{pk} + \beta'_{p1} \mathbf{g}_{pki}(r_i, H_i, Y_i) + \beta'_{p2} X_i) \right. \\
& \quad \left. - \mu_d \exp(\hat{M}_{di}) \sum_{k=1}^{t_{di}} \exp(\gamma_{dk} + \beta'_{d1} \mathbf{g}_{dki}(r_i, H_i, Y_i) + \beta'_{d2} X_i) \right\}. \tag{5}
\end{aligned}$$

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.  $t_i = \min(t_{pi}, t_{di})$ . Define  $F_p(t_i | \xi_{pi}, \xi_{di})$  as the probability of mortgage termination by prepayment of the  $i$ th borrower in period  $t$ ,  $F_d(t_i | \xi_{pi}, \xi_{di})$  as the probability of mortgage termination by default of the  $i$ th borrower in period  $t$ , and  $F_c(t_i | \xi_{pi}, \xi_{di})$  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 <sup>5</sup>

$$\begin{aligned}
F_p(t_i | \xi_{pi}, \xi_{di}) &= S(t_i, t_i | \xi_{pi}, \xi_{di}) - S(t_i + 1, t_i | \xi_{pi}, \xi_{di}) \\
&\quad - \frac{1}{2} \left\{ S(t_i, t_i | \xi_{pi}, \xi_{di}) + S(t_i + 1, t_i + 1 | \xi_{pi}, \xi_{di}) \right. \\
&\quad \left. - S(t_i, t_i + 1 | \xi_{pi}, \xi_{di}) - S(t_i + 1, t_i | \xi_{pi}, \xi_{di}) \right\},
\end{aligned} \tag{6}$$

$$\begin{aligned}
F_d(t_i | \xi_{pi}, \xi_{di}) &= S(t_i, t_i | \xi_{pi}, \xi_{di}) - S(t_i, t_i + 1 | \xi_{pi}, \xi_{di}) \\
&\quad - \frac{1}{2} \left\{ S(t_i, t_i | \xi_{pi}, \xi_{di}) + S(t_i + 1, t_i + 1 | \xi_{pi}, \xi_{di}) \right. \\
&\quad \left. - S(t_i, t_i + 1 | \xi_{pi}, \xi_{di}) - S(t_i + 1, t_i | \xi_{pi}, \xi_{di}) \right\},
\end{aligned} \tag{7}$$

and

$$F_c(t_i | \xi_{pi}, \xi_{di}) = S(t_i, t_i | \xi_{pi}, \xi_{di}). \tag{8}$$

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

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

where  $N$  is the sample size and  $\delta_{ji}$ ,  $j = p, d, c$ , 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.

## II. 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).

### A. The Data

The database contains 1,489,372 observations on single family mortgage loans issued

---

<sup>5</sup> The dependence of these functions on  $r, H, Y, X, \mu_p, \mu_d$ , and  $\theta$  has been omitted for notational simplicity.

The term  $\frac{1}{2} \left\{ S(t_i, t_i | \xi_{pi}, \xi_{di}) + S(t_i + 1, t_i + 1 | \xi_{pi}, \xi_{di}) - S(t_i, t_i + 1 | \xi_{pi}, \xi_{di}) - S(t_i + 1, t_i | \xi_{pi}, \xi_{di}) \right\}$ , which appears in both equations (6) and (7) is an adjustment for mortgage duration data measured in discrete rather than continuous time.

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 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 446,098 observations. Loans are observed in each quarter from the quarter of origination through the quarter of termination, maturation, or through 1992:1 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  $g_{pk}(r, H, Y)$  in section I) 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.<sup>6</sup>

---

<sup>6</sup> 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  $\tau_i$ , when the local market interest rate is  $m_{j,\tau_i+k_i}$ , where  $j$  indexes the local region, the “*Call Option*” is defined as:  $Call\_Option_{i,k} = 1 - V_{i,r}^* / V_{i,m}$ ,

where  $V_{i,r}^* = \sum_{S=1}^{TM_i-k_i} 1/(1+r_i)^S$ , and  $V_{i,m} = \sum_{S=1}^{TM_i-k_i} 1/(1+m_{j,\tau_i+k_i})^S$ .

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  $g_{pk}(r, H, Y)$  in section I) measures the probability that homeowner equity is negative, i.e., the probability that the put option is in the money.<sup>7</sup> The details of the calculation of these variables are reported in Deng et al. (2000).

As proxies for other “trigger events,” we include a measure of the quarterly unemployment rate and the annual divorce rate by state (i.e.,  $X$  in section I).

The correlate of unobserved heterogeneity across borrowers is computed in the following way. At each quarter since origination, we calculate whether the intrinsic value of call option is in the money (this merely indicates whether current market interest rates on new first mortgages<sup>8</sup> are lower than the contract interest rate). We then compute a time-varying covariate,  $W$ , for each borrower reflecting the number of quarters since origination that an in-the-money call was not

---

<sup>7</sup> Specifically, the market value  $M_i$  of property  $i$ , purchased at a cost of  $C_i$  at time  $\tau_i$  and evaluated  $k_i$  quarters thereafter is  $H_{i,k} = C_i \left( I_{j,\tau_i+k} / I_{j,\tau_i} \right)$ , 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:  $Put\_Option_{i,k} = \Phi \left( \left( \log V_{i,m} - \log H_{i,k} \right) / \sqrt{\omega^2} \right)$ , where  $\Phi(\cdot)$  is cumulative standard normal distribution function,  $\omega^2$  is an estimated variance, and  $V_{i,m}$  is defined in footnote 5. The term  $\omega^2$  is defined more precisely in Deng et al., (2000).

<sup>8</sup> Interest rates on new mortgage contracts are available by quarter and region at <http://www.freddiemac.com>.

exercised. A borrower who systemically passes up profitable opportunities to prepay the mortgage is more likely to be a “woodhead.” Our measure,  $W$ , treats differences in “astuteness” among borrowers, in their “costs of calculation,” and in their “transactions costs” as observationally equivalent.

We have computed the measure  $W$  at each quarter for each mortgage in the sample. There are a total of 16,454,954 of these event histories in our sample of mortgages. Table I summarizes some of this information. Panel A presents the distribution of  $W$  among mortgages, separately for the full sample and for differently seasoned mortgage pool.<sup>9</sup> 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 85 percent of borrowers missed at least one opportunity. About 15 percent of borrowers in the sample missed more than twelve profitable opportunities, while for ten-year seasoned mortgage pools, about 22 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  $W$  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 mean of values of in-the-money calls (in percent) by different seasoning of mortgage pools. These averages are reported separately for borrowers who never

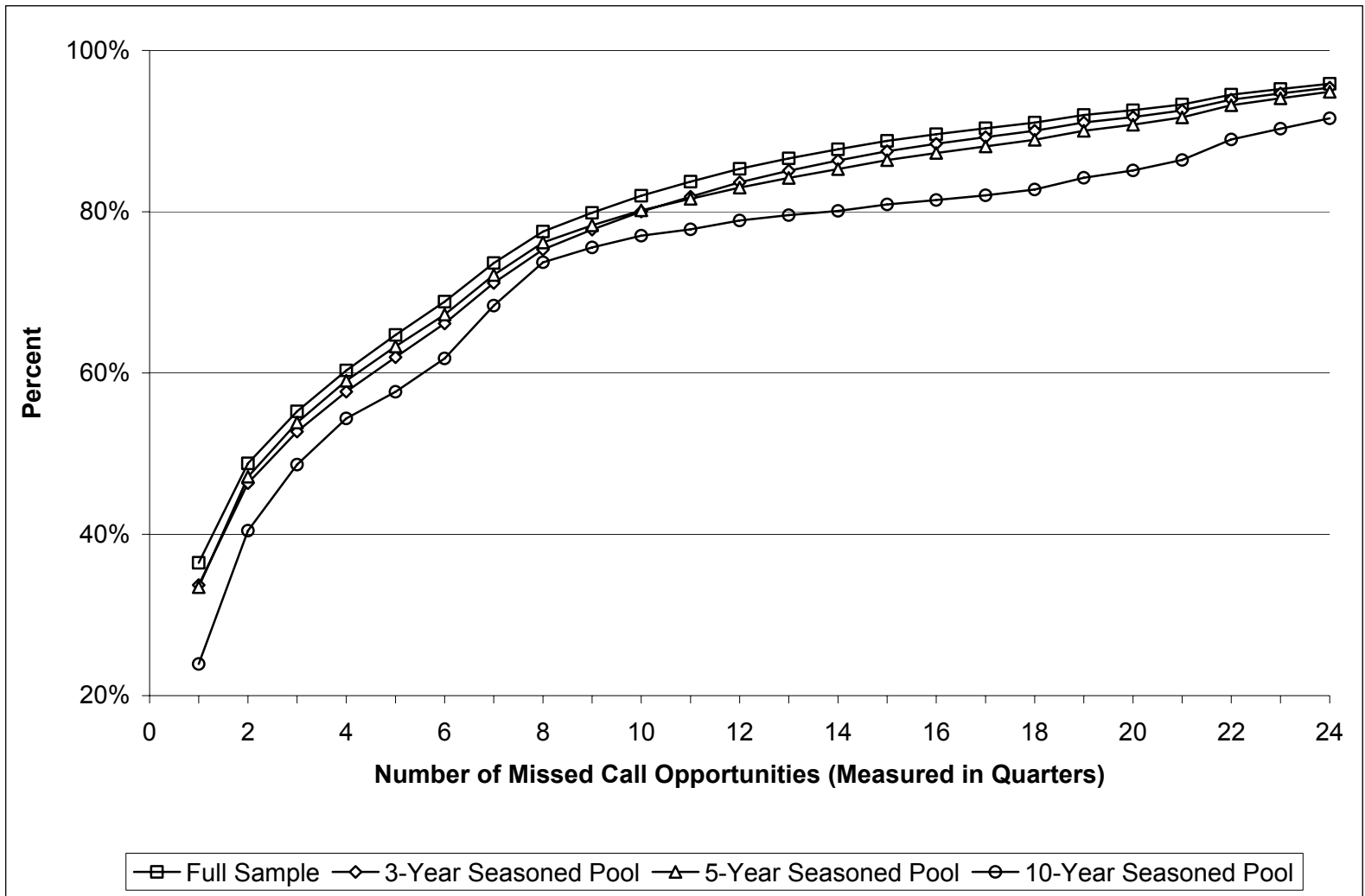
---

<sup>9</sup> The three (five, or ten) year seasoned pool is a sub-sample of mortgage loans which have durations greater than three (five, or ten) years. Our full sample represents a pool containing the newly issued mortgage loans.

**Table I.**  
**Number of Loans and Payable Events by Number of Missed Call Options**

This table summarizes the distribution of  $W$ , a measure that counts the frequency of missed profitable refinancing (call options) opportunities. We have computed the measure  $W$  at each quarter for each mortgage in the sample. There are a total of 446,098 loan records and 16,454,954 of payable event histories in our sample of mortgages. Panel A presents the distribution of  $W$  among mortgages, separately for the full sample and for differently seasoned mortgage pools. Panel B presents the number of payable events, separately for the full sample and for mortgage pools of different seasoning. Column percentages are in parentheses.

	Full Sample	3-Year Seasoned Pool	5-Year Seasoned Pool	10-Year Seasoned Pool
<b>PANEL A – NUMBER OF LOANS</b>				
W = 0	96,977 (21.74)	84,204 (20.80)	75,177 (20.57)	29,428 (15.38)
W = 1-2	102,936 (23.07)	87,818 (21.70)	83,001 (22.72)	41,614 (21.76)
W = 3-4	58,089 (13.02)	51,691 (12.77)	48,681 (13.32)	27,424 (14.34)
W = 5-8	79,623 (17.85)	73,565 (18.17)	63,861 (17.48)	36,791 (19.23)
W = 9-12	40,675 (9.12)	39,707 (9.81)	30,833 (8.44)	14,827 (7.75)
W > 12	67,798 (15.20)	67,798 (16.75)	63,835 (17.47)	41,197 (21.54)
Total	446,098	404,783	365,388	191,281
<b>PANEL B – NUMBER OF PAYABLE EVENTS</b>				
W = 0	3,180,544 (19.33)	3,090,680 (19.14)	2,935,867 (18.95)	1,555,563 (15.30)
W = 1-2	3,725,256 (22.64)	3,622,341 (22.43)	3,542,499 (22.86)	2,248,700 (22.11)
W = 3-4	2,248,305 (13.66)	2,200,448 (13.63)	2,151,872 (13.89)	1,499,145 (14.74)
W = 5-8	3,049,864 (18.53)	2,992,905 (18.54)	2,843,530 (18.35)	2,018,672 (19.85)
W = 9-12	1,475,546 (8.97)	1,464,590 (9.07)	1,315,406 (8.49)	821,245 (8.08)
W > 12	2,775,439 (16.87)	2,775,439 (17.19)	2,704,868 (17.46)	2,025,307 (19.92)
Total	16,454,954	16,146,403	15,494,042	10,168,632



**Figure I. Cumulative Frequency of Missed Call Opportunities.** This figure presents the cumulative frequency of missed call opportunities (measured in quarters), separately for the full sample and for mortgage pools of different seasoning.

**Table II.**  
**Mean Value of In-The-Money Calls (In Percent) at Termination**  
**by Number of Missed Call Opportunities and Seasoning of Mortgage Pools**

Table entries are the mean values of the extent to which the call-options are in the money at the time of termination, separately by mortgage pools with different seasoning. These averages are reported separately for borrowers who never passed a profitable prepayment opportunity ( $W=0$ ) 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.  $W$  measures number of quarters that the call option has been in-the-money but the borrower has not exercised the call option. The sample consists of 446,098 mortgages.

	Full Sample	3-Year Seasoned Pool	5-Year Seasoned Pool	10-Year Seasoned Pool
$W = 0$	-15.16	-14.66	-13.30	-6.45
$W = 1-2$	-4.50	-4.53	-3.92	-1.81
$W = 3-4$	3.28	2.71	2.98	1.02
$W = 5-8$	5.41	5.35	4.85	3.91
$W = 9-12$	11.91	11.69	9.46	5.39
$W > 12$	16.85	16.85	16.10	13.71

passed up a profitable prepayment opportunity ( $W = 0$ ) 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. The table reports two striking regularities.

First, for mortgages of given duration, the averages increase monotonically with  $W$ . Larger values of this variable are associated with much larger potential gains from exercise. The average gain from exercise is about 1 to 3 percent for those who passed up three or four opportunities, 4 to 5 percent for those who passed up five to eight opportunities, up to 12 percent for those who passed up nine to twelve opportunities, and up to 17 percent for those who passed up more than twelve profitable opportunities to refinance. The pattern of average values is similar for mortgage pools of differing seasoning.

Second, the average values of the call option associated with  $W > 2$  declines with mortgage seasoning. Foregoing three or four profitable refinance opportunities is associated with an average intrinsic value of the call option of 3 percent in the full sample, and with an average value of 1 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 more seasoned mortgages. The average value of the call option for those remaining in the pool is higher.

### *B. 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.

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 flexible baseline functions for the prepayment and default equations, i.e., the baseline functions are estimated non-parametrically at the same time while we estimate the parametric function of the proportional factors.<sup>10</sup> Thus, the row labeled “Baseline” in the table reports the average shift in the non-parametric baselines for prepayment and default functions, respectively, estimated according to equation (2).

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

---

<sup>10</sup> We also estimate these models using PSA and SDA baseline functions. These results are reported in Appendix B, Table B2. In general, our results are robust to the functional form imposed on the baseline.

**Table III.**

**Maximum Likelihood Estimates for Competing Risks of Mortgage Prepayment and Default**

Models 1 to 3 are estimated by MLE approach. In model 3, 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 4 is estimated by three-stage error correction maximum likelihood ECML approach. Prepayment and default functions are modeled as correlated competing risks estimated jointly. Flexible non-parametric baselines for prepayment and default functions are estimated simultaneously with the competing risks factors. t-ratios are in parentheses.

	Model 1		Model 2		Model 3		Model 4	
	Prepay	Default	Prepay	Default	Prepay	Default	Prepay	Default
Call Option (fraction of contract value)	4.799 (112.00)	6.801 (16.64)	6.343 (82.01)	5.735 (8.19)	6.523 (76.90)	5.753 (8.09)	7.204 (87.39)	5.715 (8.14)
Put Option (probability of negative equity)	-5.300 (10.74)	8.852 (8.58)	-5.804 (11.75)	8.854 (8.72)	-5.733 (11.41)	9.346 (9.15)	-5.258 (9.48)	8.898 (8.74)
Call Option Squared	1.427 (9.53)	0.608 (0.49)	4.085 (21.49)	-1.656 (1.02)	4.637 (21.33)	-0.350 (0.21)	5.686 (29.82)	-1.653 (1.02)
Put Option Squared	5.710 (9.10)	-9.174 (6.80)	6.313 (10.02)	-9.217 (6.92)	6.267 (9.80)	-9.629 (7.19)	6.065 (8.62)	-9.278 (6.94)
State Unemployment Rate (percent)	-0.039 (7.58)	0.083 (1.67)	-0.042 (8.15)	0.093 (1.84)	-0.043 (8.13)	0.096 (1.88)	-0.072 (13.10)	0.094 (1.85)
State Divorce Rate (percent)	-0.009 (0.81)	0.471 (3.95)	-0.016 (1.43)	0.477 (4.00)	-0.022 (1.84)	0.482 (4.02)	0.006 (0.43)	0.475 (3.97)
0.6<LTV≤0.75	0.065 (2.48)	2.145 (2.65)	0.059 (2.18)	2.154 (2.65)	0.052 (1.77)	2.137 (2.62)	0.064 (2.05)	2.149 (2.65)
0.75<LTV≤0.8	0.044 (1.90)	2.491 (3.12)	0.044 (1.83)	2.495 (3.13)	0.036 (1.38)	2.493 (3.11)	0.057 (2.10)	2.493 (3.13)
0.8<LTV≤0.9	0.094 (3.77)	3.438 (4.37)	0.110 (4.24)	3.439 (4.37)	0.100 (3.54)	3.416 (4.31)	0.140 (4.79)	3.433 (4.36)
LTV>0.9	-0.024 (0.78)	3.878 (4.94)	0.004 (0.12)	3.879 (4.93)	0.010 (0.29)	3.896 (4.92)	-0.009 (0.26)	3.877 (4.93)
W			-0.044 (22.01)	0.034 (1.89)	-0.037 (14.81)	0.053 (2.86)	-0.036 (18.13)	0.037 (2.05)
Baseline Intercept	3.709 (7.58)	0.001 (0.83)	4.070 (7.55)	0.001 (0.82)			3.108 (7.24)	0.001 (0.82)
Baseline Intercept (“ruthless”)					4.407 (7.36)	0.001 (0.81)		
Baseline Intercept (“woodheads”)					0.604 (2.98)	0.000 (0.00)		
Fraction “woodheads”					0.044 (3.34)			
Log Likelihood	-73,974		-73,734		-73,683		-66,507	
Schwarz B.I.C.	74,094		73,864		73,823		66,637	

likely to exercise options.

Model 2 simply adds the number of missed opportunities,  $W$ , a time-varying covariate, to both prepayment and default functions in this model. This specification is analogous to those used by financial analysts in estimating prepayment rates for mortgage pools.<sup>11</sup>

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.

Model 3 extends the model 2 by recognizing of unobserved heterogeneity. In model 3, we allow for the possibility that there are two distinct groups of borrowers; we 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. We estimate the distribution of unobserved heterogeneity and the average behavior of the two groups jointly with the competing risk functions.

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  $W$  remains 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 distinct groups estimated in model 3 in their exercise of the prepayment option. For prepayment, those in the high risk group are about 7 times riskier (e.g., 4.407 versus 0.604) than borrowers in the low risk group. This difference is highly significant. For the default option, there is no significant behavioral difference between the two groups of borrowers. For model 3, over 95 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,”  $W$ , adjusted by the

---

<sup>11</sup> 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).

duration of the loan, and the unobserved heterogeneity among mortgage borrowers.

We begin by estimating the Cox model of competing risks of prepayment and default specified by Model (1). We then collect the martingale residuals of prepayment and default for each individual borrower and compute the martingale transform residuals. (See Appendix A. for details.)

We then regress the martingale transform residuals upon our measure of the “number of missed opportunities” each borrower has had up to the current quarter year and the duration of the loan, such that

$$M_{pi}^X = \alpha_p + \beta_{p1}W(T_i) + \beta_{p2}T_i + \varepsilon_i, \quad (10)$$

$$M_{di}^X = \alpha_d + \beta_{d1}W(T_i) + \beta_{d2}T_i + \varepsilon_i. \quad (11)$$

where  $M_{pi}^X$  and  $M_{di}^X$  are the martingale transforms specified in Appendix A, Equations (A1) and (A2),  $W(T_i)$  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 measured at the termination time  $T_i$ ),  $T_i$  is the age of the mortgage in quarters, and  $\varepsilon_i$  is a random error term which follows a standard normal distribution. The estimated coefficients (reported in Appendix B. Table B1) are highly significant in both prepayment and default functions.

Finally, we estimate the model specified in equation (5) in which the random-effect errors in the joint survival function are corrected by using the expected martingale residuals,  $\hat{M}_{pi}$  and  $\hat{M}_{di}$ , following the martingale transform procedures described in Appendix A.

A comparison with model 3, in which heterogeneity is specified as two distinct groups, indicates that the magnitudes of the coefficients of the option values are 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 model 4 are very substantially higher than those of the other models reported in Table III. The log likelihood yields a value of  $-66,507$  for model 4 as compared to a value of  $-73,683$  for the discrete measure (Model 3). Indeed, after correcting the random-effect using expected martingale residuals, we can no longer to estimate a model that distinguishes two separate groups among

borrowers. Table III also reports the Schwarz B.I.C. index for each model.<sup>12</sup> It shows clearly that model 4 dominates all other models based on the index.

### *C. Summary*

Table III presents a summary and comparison of the ECML model with other less general models of mortgage holder behavior. Model 1 reports the ML estimates of the competing risks model assuming no unobserved heterogeneity across borrowers. Model 2 adds the number of missed opportunities  $W$  to the analysis as an exogenous time-varying covariate. Model 3 assumes in addition a bivariate distribution of unobserved heterogeneous error terms and estimates that distribution simultaneously with the competing risks functions by ML. Model 4 reports the results when it is assumed that the number of missed opportunities  $W$  is correlated with unobserved individual heterogeneity and when the model incorporating this is estimated by ECML methods. Differences in values of log likelihood function are substantial. The ECML model is clearly superior on statistical grounds.

## **III. Pricing Implications**

In this section we evaluate the economic importance of this more general model in the pricing of mortgages, pools of mortgages, or mortgage-backed securities. We adopt a Monte Carlo simulation pricing approach to estimate 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

---

<sup>12</sup> Schwarz's B.I.C. is defined as  $-2L + m \times \ln(N)$ , where  $N$  is the sample size,  $L$  is the maximized log-likelihood of the model and  $m$  is the number of parameters in the model. The index takes into account both the statistical goodness of fit and the number of parameters estimated to achieve this particular degree of fit, by imposing a penalty for increasing the number of parameters. Lower values of the index indicate the preferred model.

functions of risk-neutral drift factors. The basic model we use is the DS generalization of the term structure model<sup>13</sup> of Balduzzi, Das, Foresi and Sundaram (1996). Appendix B. Figure B1 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 III 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 individual interest rate paths.

Table IV summarizes the relative pricing differences of mortgage pools at three assumed contract interest rates. The first two columns present the mean percentage difference in equilibrium prices between model 1 and model 4 and between model 2 and model 4, respectively. Model 2 represents heterogeneity by the inclusion of the variable  $W$  directly in the competing risks model.<sup>14</sup> The last column presents the percentage differences in prices between model 3 and model 4. Model 3 specifies the unobserved heterogeneity among borrowers in two categories and also includes the variable  $W$  directly in the computing risks model.

The simulations are reported separately for mortgage pools with coupon interest rates 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.<sup>15</sup> 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  $W$  observed from the sample, reported in Table I, Panel A, as the basis for the simulation. For example, in the simulations using the 3-year seasoned sample, we assume

---

<sup>13</sup> Our simulation is based upon equation (23) of DS, using the parameters reported by DS in Table II, Column 2.

<sup>14</sup> Model 2 is, perhaps, close to the representation of heterogeneity which may appear in models used by practitioners to price mortgage pools. See Richard and Roll (1989) for a discussion.

<sup>15</sup> 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 IV.**  
**Mean Percentage Differences in Equilibrium Prices Simulated by**  
**Econometric Models at Various Coupon Interest Rates**

The percentage price differences reported in this table are measured by the price differences implied by different econometric models divided by simulated pool price based on model 4. 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. t-ratios are in parentheses.

	Model 1 vs. Model 4	Model 2 vs. Model 4	Model 3 vs. Model 4
<b>A. 8.25 PERCENT</b>			
Full Sample	1.56% (127)	0.52% (114)	0.27% (88)
3-Year Seasoned Pool	1.30 (161)	0.62 (167)	0.45 (151)
5-Year Seasoned Pool	1.50 (256)	0.83 (157)	0.64 (232)
10-Year Seasoned Pool	2.42 (417)	1.72 (447)	1.45 (389)
<b>B. 8.75 PERCENT</b>			
Full Sample	2.11 (153)	0.62 (122)	0.32 (92)
3-Year Seasoned Pool	1.64 (184)	0.67 (169)	0.47 (154)
5-Year Seasoned Pool	1.80 (271)	0.87 (217)	0.66 (197)
10-Year Seasoned Pool	2.99 (425)	1.90 (367)	1.56 (320)
<b>C. 9.25 PERCENT</b>			
Full Sample	2.55 (176)	0.68 (133)	0.35 (101)
3-Year Seasoned Pool	1.92 (202)	0.68 (176)	0.47 (163)
5-Year Seasoned Pool	2.01 (261)	0.87 (188)	0.65 (1.74)
10-Year Seasoned Pool	3.47 (417)	1.98 (306)	1.59 (270)

20.8 percent of mortgages have never missed a single profitable call opportunity, 21.7 percent of mortgages have  $W = 1$ , 12.77 percent have  $W = 3$ , 18.17 percent have  $W = 5$ , 9.81 percent have  $W = 9$ , and 16.75 percent have  $W = 13$ .

$W$ , of course, varies with duration, and a mortgage pool manager does not observe the time-varying path of  $W$  ex ante. She can observe directly, however, the distribution of  $W$  in mortgage pools with different seasoning. 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 IV, we first estimate the prepayment and default risks of each mortgage pool based on the parameters of Models 1 to 4, the distribution of  $W$  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 rate paths over the distribution of  $W$ . The detailed estimates are reported in the appendix.<sup>16</sup> As the comparison in the table shows, the pricing differences estimated from different models are quite large. Model 1 to Model 3 all tend to overprice the mortgages compared to the model estimated using ECML approach. The comparison also indicates that the gain from the ECML estimation technique is larger for more seasoned mortgage pools. For example, for a 5-year seasoned mortgage pool, the pricing differences between ECML model and Model 3 are about 0.65 percent, while the pricing differences between ECML and Model 3 for a 10-year seasoned mortgage pool are between 1.4 to 1.6 percent.

---

<sup>16</sup> Appendix B, Table B3 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 1, Model 2, Model 3, and Model 4. The first column presents the average price differences between Model 1 and Model 4. The second column presents the mean absolute price differences between Model 2 and Model 4. The last column presents the mean absolute price differences between Model 3 and Model 4. The simulation results reported in Appendix B, Table B3 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 nature of the results was not affected by changes in assumed volatility.



The ECML 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. This suggests that the ECML model has practical importance for the pricing and valuation of mortgages and mortgage-backed securities.

#### **IV. Conclusion**

The mortgage market is large and has grown greatly in importance. Until recently, it was estimated that the outstanding volume of residential mortgages would 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 mortgage contracts. 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 an error correction maximum likelihood estimator (ECML) 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 mortgages suggests that this method offers advantages in precision in comparison with conventional methods.

The ECML 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 ECML model are substantially different, at least for this body of data. This suggests that the model may have considerable practical application in the pricing of seasoned pools of mortgages.

## Appendix A. Martingale Transformations

The distribution of the martingale residuals specified in section II equation (4) is highly skewed. That poses some difficulty in applying martingale residuals directly to our analysis. Following martingale transform procedures, we convert the residuals to approximate the normal distribution. For this data set, the martingale transformations that yield approximately normal distribution for prepayment function and default function are power functions (see Therneau, Grambsch, and Fleming, 1990):

$$M_{pi}^X = \left\{ \text{abs} \left[ \log(1 - M_{pi}) \right] \right\}^{7/10}, \quad (\text{A1})$$

$$M_{di}^X = \left\{ \text{abs} \left[ \log(1 - M_{di}) \right] \right\}^{9/100}. \quad (\text{A2})$$

The expected martingale residuals  $\hat{M}_{pi}$  and  $\hat{M}_{di}$  are obtained through the reverse transformation

$$\hat{M}_{pi} = \begin{cases} 1 - \exp \left[ \text{abs} \left( \hat{M}_{pi}^X \right)^{10/7} \right], & \text{if } M_{pi} \leq 0 \\ 1 - \exp \left[ - \left( \text{abs} \left( \hat{M}_{pi}^X \right)^{10/7} \right) \right], & \text{if } M_{pi} > 0 \end{cases} \quad (\text{A3})$$

$$\hat{M}_{di} = \begin{cases} 1 - \exp \left[ \text{abs} \left( \hat{M}_{di}^X \right)^{100/9} \right], & \text{if } M_{di} \leq 0 \\ 1 - \exp \left[ - \left( \text{abs} \left( \hat{M}_{di}^X \right)^{100/9} \right) \right], & \text{if } M_{di} > 0 \end{cases} \quad (\text{A4})$$

## Appendix B. Supplementary Tables and Figures

**Table B1.**  
**Second-Stage Estimates of Martingale Transform Residuals**

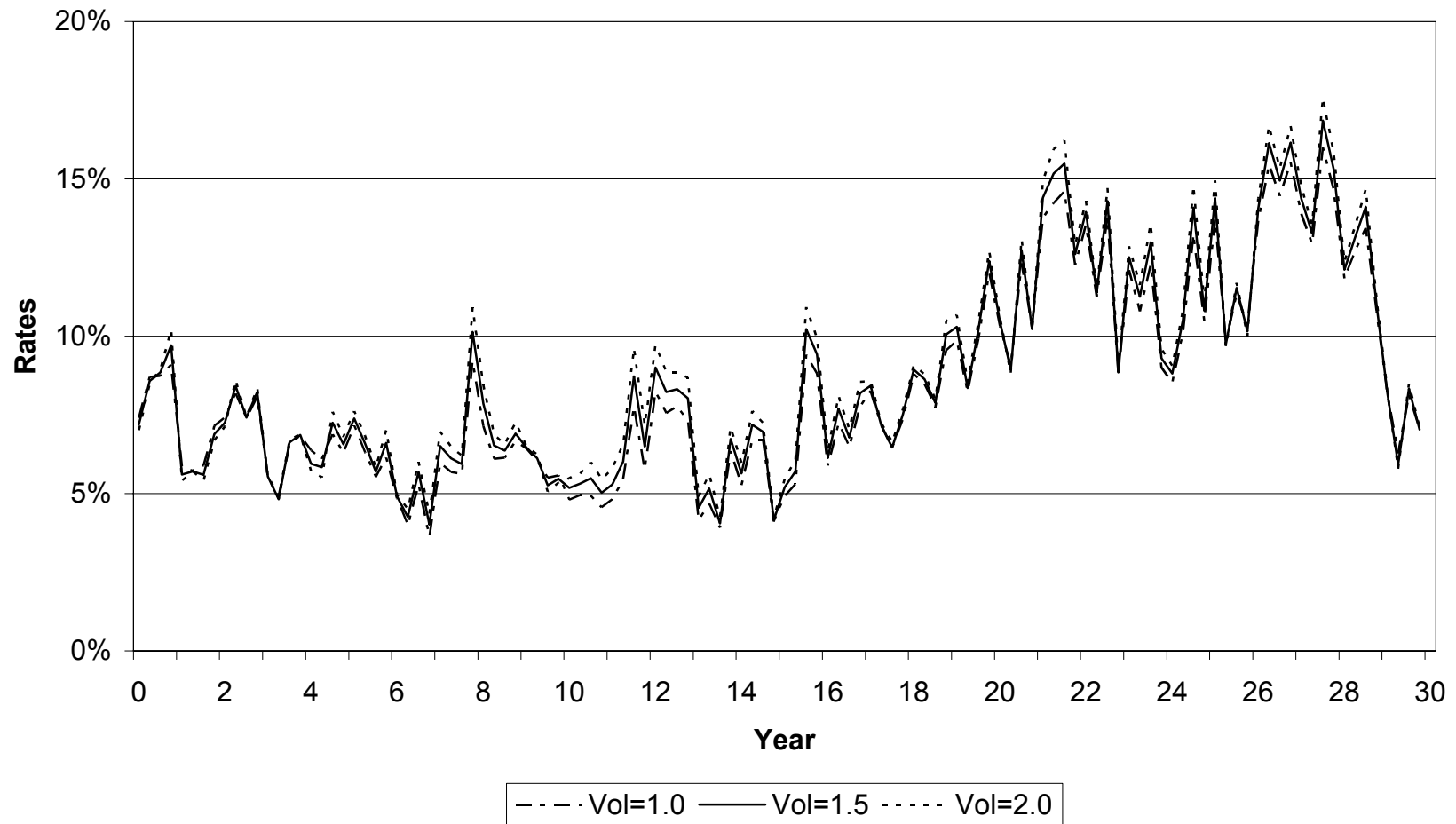
The table reports the second-stage regressions, equations (10) and (11) estimated using martingale transformation residuals from the first-stage Cox model based on a sample of 22,293 mortgages.  $M_{pi}^X = \{abs[\log(1-M_{pi})]\}^{7/10}$  and  $M_{di}^X = \{abs[\log(1-M_{di})]\}^{9/100}$  are used as transformation functions for prepayment and default, respectively. Skewness and Kurtosis for the prepayment martingale transform regression are 0.067 and 0.344, respectively. Skewness and Kurtosis for the default martingale transform regression are 0.004 and 9.149, respectively. t-ratios are in parentheses.

Variables	Prepayment	Default
Constant	1.648 (226.87)	0.510 (264.72)
W (measured at Termination)	-0.014 (33.48)	0.007 (61.62)
Duration	-0.018 (99.09)	0.007 (33.85)
R <sup>2</sup>	0.356	0.205

**Table B2.****Estimates of Competing Risks of Mortgage Prepayment and Default Using Parametric Baselines**

Models 1 to 3 are estimated by MLE approach. In model 3, 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 4 is estimated by three-stage error correction maximum likelihood (ECML) approach. Prepayment and default functions are modeled as correlated competing risks estimated jointly. Baselines for prepayment and default functions are specified by PSA and SDA curves, respectively. t-ratios are in parentheses.

	Model 1		Model 2		Model 3		Model 4	
	Prepay	Default	Prepay	Default	Prepay	Default	Prepay	Default
Call Option (fraction of contract value)	4.948 (120.56)	6.827 (18.71)	6.295 (92.25)	4.315 (7.10)	6.459 (86.19)	4.215 (6.83)	6.668 (87.97)	4.292 (7.06)
Put Option (probability of negative equity)	-6.434 (13.12)	9.074 (9.30)	-7.305 (14.87)	8.993 (9.61)	-7.360 (14.77)	9.562 (10.01)	-8.673 (16.14)	9.048 (9.67)
Call Option Squared	1.384 (10.35)	0.339 (0.30)	3.479 (21.78)	-4.750 (3.37)	3.947 (21.42)	-3.701 (2.58)	3.817 (23.20)	-4.749 (3.36)
Put Option Squared	6.860 (11.10)	-9.283 (7.28)	7.859 (12.76)	-9.471 (7.72)	7.978 (12.88)	-9.888 (8.04)	9.732 (14.62)	-9.542 (7.77)
State Unemployment Rate (percent)	-0.047 (9.56)	0.003 (0.07)	-0.061 (12.22)	0.052 (1.20)	-0.063 (12.50)	0.051 (1.19)	-0.109 (20.81)	0.052 (1.21)
State Divorce Rate (percent)	-0.014 (1.37)	0.356 (3.33)	-0.031 (2.90)	0.382 (3.53)	-0.040 (3.52)	0.392 (3.60)	-0.054 (4.46)	0.380 (3.50)
0.6<LTV≤0.75	0.077 (2.94)	2.184 (2.91)	0.076 (2.77)	2.204 (2.93)	0.074 (2.53)	2.203 (2.92)	0.101 (3.16)	2.199 (2.93)
0.75<LTV≤0.8	0.065 (2.79)	2.539 (3.50)	0.072 (2.95)	2.546 (3.51)	0.069 (2.63)	2.553 (3.50)	0.124 (4.37)	2.544 (3.50)
0.8<LTV≤0.9	0.123 (4.89)	3.462 (4.81)	0.146 (5.57)	3.475 (4.82)	0.141 (5.02)	3.459 (4.78)	0.217 (7.07)	3.469 (4.82)
LTV>0.9	0.007 (0.22)	3.900 (5.40)	0.041 (1.28)	3.915 (5.42)	0.049 (1.42)	3.934 (5.43)	0.070 (1.93)	3.913 (5.42)
W			-0.041 (22.00)	0.087 (5.54)	-0.036 (16.63)	0.107 (6.39)	-0.023 (12.32)	0.090 (5.76)
Baseline Intercept	3.553 (18.12)	0.005 (0.98)	4.964 (17.36)	0.002 (0.95)			6.850 (15.21)	0.002 (0.95)
Baseline Intercept (“ruthless”)					5.357 (15.84)	0.002 (0.94)		
Baseline Intercept (“woodheads”)					0.587 (2.27)	0.000 (0.00)		
Fraction “woodheads”					0.023 (3.03)			
Log Likelihood	-74,321		-74,062		-74,018		-67,394	
Schwarz B.I.C.	74,431		74,182		74,148		67,514	



**Figure B1. Simulated Interest Rates (ATSM).** The figure shows 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.

**Table B3.**  
**Mean Differences in Equilibrium Prices Simulated from One Million Dollar**  
**Mortgage Pools at Different Coupon Interest Rates**

This table reports mean differences in equilibrium prices implied by different econometric models simulated from one million dollar mortgage pools. These price differences are reported separately for the full sample and for sub-samples with 3-year, 5-year and 10-year seasoning. These simulations are based on 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. t-ratios are in parentheses.

	Model 1 vs. Model 4	Model 2 vs. Model 4	Model 3 vs. Model 4
<b>A. 8.25 PERCENT</b>			
Full Sample	\$16,512 (129)	\$5,463 (115)	2,801 (89)
3-Year Seasoned Pool	13,488 (163)	6,492 (168)	4,681 (152)
5-Year Seasoned Pool	15,480 (262)	8,589 (260)	6,644 (233)
10-Year Seasoned Pool	25,036 (435)	17,834 (452)	14,969 (391)
<b>B. 8.75 PERCENT</b>			
Full Sample	22,279 (157)	6,572 (123)	3,352 (93)
3-Year Seasoned Pool	17,103 (188)	6,954 (171)	4,851 (155)
5-Year Seasoned Pool	18,651 (280)	8,985 (219)	6,787 (198)
10-Year Seasoned Pool	30,946 (448)	19,710 (370)	16,112 (321)
<b>C. 9.25 PERCENT</b>			
Full Sample	26,978 (181)	7,236 (135)	3,655 (101)
3-Year Seasoned Pool	19,979 (208)	7,142 (179)	4,885 (165)
5-Year Seasoned Pool	20,847 (271)	8,991 (190)	6,708 (176)
10-Year Seasoned Pool	35,907 (442)	20,532 (310)	16,428 (271)

## REFERENCES

- Archer, Wayne R., David C. Ling, and Gary A. McGill, 1996, The Effect of Income and Collateral Constraints on Residential Mortgage Terminations, *Regional Science and Urban Economics* 26(2-3), 235-261.
- Barlow, W. E., and R. L. Prentice, 1988, Residuals for Relative Risk Regression, *Biometrics* 75, 65-74.
- 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., 1972, Regression Models and Life-Tables, *Journal of the Royal Statistical Society B* 34(2), 187-220.
- 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.
- 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.
- Hayre, Lakhbir, ed., 2001, *Salomon Smith Barney Guide to Mortgage-Backed And Asset-Backed Securities* (John Wiley & Sons, Inc.).

- 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.
- Therneau, T. M., P. M. Grambsch, and T. R. Fleming, 1990, Martingale Based Residuals for Survival Models, *Biometrika* 77, 147-160.