2001 V29 3: pp. 411–450
REAL ESTATE
ECONOMICS

Movers and Shuckers: Interdependent Prepayment Decisions

John M. Clapp,* Gerson M. Goldberg,** John P. Harding,***
and Michael LaCour-Little****

Au: Give e+mail address for Clapp, Goldberg and LaCour-Little

We model competing risks of mortgage termination where the borrower faces a repeated choice to continue to pay, refinance the loan, move or default. Most previous empirical work on mortgage prepayment has ignored the distinction between prepayments triggered by refinancing and moving, combining them into a single prepayment rate. We show that financial considerations are the primary drivers of the refinance choice while homeowner characteristics have more influence on the move decision. We demonstrate that these differences are statistically significant and that combining these two distinct choices into a single measure of prepayment shifts coefficients toward zero and produces inaccurate predictions of aggregate termination rates. For example, a combined model underestimates the effect of the market price of the loan on refinancing; it misses entirely the opposite effects of borrower income on moving and refinancing. Our results suggest that existing prepayment models are inconsistent predictors of mobility-driven prepayment and underestimate the effect of market conditions and borrower characteristics on refinancing and housing decisions. Our findings have great significance to mortgage investors because mobility-driven prepayments are likely to be a more significant source of prepayments in the next decade.

Predicting aggregate prepayment is as an important issue in mortgage valuation (Hayre, Chaudhary and Young 2000). It is well established (Patruno 1994) that two distinct borrower motivations drive full prepayments of mortgages: (1) the desire to sell the current house and relocate[1] (the "movers") and (2) the desire to refinance the existing mortgage—usually to obtain a lower rate of interest and/or

*School of Business Administration, University of Connecticut, Storrs, CT.
**Gabelli School of Business, Roger Williams University, Bristol, RI.
***School of Business Administration, University of Connecticut, Storrs, CT 06269-2041 or johnh@sba.uconn.edu.
****CitiMortgage, Inc., St. Louis, MO.

[1] Our discussion in this paper focuses on fixed-rate, 15- and 30-year, conventional residential mortgage loans. Almost all conventional loans are originated with "due-on-sale" clauses that require the borrower to prepay the loan when there is a sale of the property. Government insured or guaranteed loans (FHA and VA loans) are fully assumable. For any assumable loan, a move need not generate a loan payoff.

lower payment (the "shuckers").[2] We use the term "prepayment" to refer to early voluntary terminations by movers and shuckers combined. While lenders, traders and investors may only care about total prepayments, we demonstrate in this paper that the ability to predict prepayments benefits from distinguishing the causal factors pertinent to movers from those for refinancers. We obtain total prepayment by adding the two separate components.

Mortgage investors and servicers currently face a difficult challenge in predicting future prepayments. Approximately 70% of all outstanding mortgages have coupons of 7.50% or less.[3] Unless mortgage rates drop to unprecedented levels, the importance of borrower mobility to total prepayment rates will increase. This shift towards greater weight on movers as opposed to shuckers necessarily occurs whenever a refinancing boom subsides.

By necessity, most prepayment models have been estimated using data that combines terminations triggered by refinancing with terminations triggered by borrower mobility into a single prepayment rate.[4] In addition, the 1990s were characterized by sharply declining mortgage rates and extraordinarily high rates of refinancing. Therefore, models estimated from aggregate prepayments during the 1990s may provide little guidance to future mobility-driven prepayment.

Studies of mobility have generally focused on the factors that affected housing demand and assumed that changes in housing demand caused relocation (Boehm 1981). Empirical studies generally show that changes in age of the household head, permanent income, marital status or household size drive homeowner mobility. Before the mid-1980s, researchers did not consider mobility to be a function of interest rates, except to the extent that interest rates

---

[2] Some mortgagors refinance their loans to obtain cash for home remodeling or other purposes.

[3] Source: *The Market Pulse*, Spring 2000 published by The Mortgage Information Corporation.

[4] Some notable exceptions include Archer, Ling and McGill (1996), who used American Housing Survey data to study borrowers who refinance; LaCour-Little (1999), who used a proprietary sample of loans refinanced with the same lender; and Pavlov (2000b), who estimated equations for refinancers and movers and for all terminations (movers, refinancers and defaults) combined. Clapp, Harding and LaCour-Little (2000) modeled borrower mobility as interdependent with refinancing. While these studies avoided the problems associated with combining movers and refinancers, only Pavlov (2000b) and Clapp, Harding and Lacour-Little (2000) estimated a separate moving equation. None of these studies used a multinomial logit model for competing risks, used borrower characteristics to estimate a move equation, or used as accurate a measure of the market value of the loan as we obtain from the closed, form formula. Also, none of these papers demonstrated that adding up prepayments from the three sources is superior to a combined prepayment model.

affected permanent income. Green and Shoven (1986) and Quigley (1987) estimated the extent to which a borrower's below-market mortgage rate deters mobility.

Since its inception, research on mortgage prepayments has been built on the basic premise that the change in interest rates after origination is a major determinant of the conditional prepayment rate.[5] In recent years, most mortgage prepayment literature has been based on contingent claims analysis. The prepayment and default options can be viewed as options to call the underlying debt contract, and the borrower's decisions are dictated by the objective to maximize the value of these options.

This paper combines lender data on loan terminations with housing transaction data to separate movers from refinancers. We combine theoretical models of mortgage prepayment and mobility into a single framework to predict the different effects that loan, borrower and local housing market conditions will have on that choice. We then empirically test those predictions.

The main purpose of the paper is to test the importance of separating the competing risks. We test the four-choice model (continue to pay, refinance, move or default) against an alternative that is standard in the current mortgage literature: a three-choice model with movers and refinancers combined. We find that adding up estimates from the four-choice model improves predictions of aggregate prepayments.

We make two other innovations. First, we use an estimate of the callable value of the mortgage from the closed-form formula developed by Collin-Dufresne and Harding (1999). Most previous research has used a proxy for the value of the option. (For examples, see Richard and Roll 1989 or Deng, Quigley and Van order 1996.) Second, we improve the standard methodology for estimating current-loan- to-value ratios by modeling housing markets at the local level. We use a database containing all transactions in the local housing market to estimate a current loan-to-value ratio for each loan in each quarter.[6] Most previous research has used state or metropolitan-level indices.[7]

[5] "Conditional" means conditional on survival to a given date. For an early study, see Curley and Guttentag (1974). More recent studies were summarized in the 1994 special issue on mortgage prepayment of the Journal of Fixed Income.

[6] See Clapp, Harding and LaCour-Little (2000) for results based on local house price indices. Clapp (2000) elaborated on the method.

[7] Pavlov (2000b) asserted that he used a semiparametric method based on Pavlov (2000a) to estimate the value of each house through time.

Modeling Refinancing, Moving and Default

This section develops a theory of borrower choice where each month the borrower must decide whether to make the next regularly scheduled payment, refinance the mortgage, move and prepay the mortgage or default. Many factors influence the borrower's decision including personal characteristics (income, age, etc.), loan characteristics (loan amount and rate), financial market conditions (current interest rates) and housing market conditions. In the sections that follow, we combine a theoretical model of each choice with consideration of other factors that might cause a borrower to deviate from the normative theoretical result to determine the variables to include in our models.

Hendershott and Van Order (1987) showed that the right to refinance the mortgage provides the borrower a call option on the mortgage debt with a strike price equal to the unpaid mortgage balance. Furthermore, the right to default provides the borrower a call option on the mortgage debt with strike price equal to the current value of the house.[8] Viewing the problem narrowly as the decision to exercise a call option or not, the relationship between the market value of the loan and the unpaid mortgage balance is the primary determinant in the choice to refinance. Similarly, the level of the house price relative to the market value of the loan is critical in the choice to exercise the default option.

While the standard options framework provides important insights to the borrower's choice problem, it does not fully explain refinance or default decisions. Empirical evidence shows that mortgagors do not exercise their options to prepay and default in the same manner that investors exercise financial options. (See Vandell 1995 for discussion of this evidence.) Increasingly, research has focused on transaction costs and institutional constraints to explain these differences. Further, the contingent claims approach does not address the move decision. The economic theory on mobility points to a strong role for borrower characteristics in the move choice. A choice model with the move alternative needs more than the standard options-related variables.

Letting $H(t)$ and $r(t)$ denote the house price and mortgage rate, respectively, we use the following definitions:

$F(t)$ = the unpaid principal balance at time t.

---

[8] The default option can be equivalently viewed as a put option on the house with strike price equal to the market value of the mortgage.

$N(r,t)$ = the market value at time t of a noncallable debt security with the same payment schedule as the mortgage when the time t mortgage rate $= r$.
$H_d(t)$ = the borrower's demand for housing services at time t.
$H_s^i(t)$ = the quantity of housing services supplied by structure i at time t.
$J(H, H_d, H_s^i, r, t)$ = the market value at time t of the borrower's joint option to terminate the mortgage by prepayment or default.
$C(r,t)$ = the market value at time t of a call option on $N(r,t)$ with strike price $= F(t)$, a pure refinance option.
$D(H,r,t)$ = the market value of a call option on $N(r,t)$ with strike price $H(t)$, a pure default option.
$B(H, H_d, H_s^i, r, t)$ = the market value of the borrower's right to terminate the mortgage by moving and payoff.[9]
$K_k(t)$ = transaction costs; k = R (refinance), D (default) or M (move).

Assuming that H, H_d, H_s^i and r vary stochastically and invoking the standard "perfect market" assumptions, the market value of the mortgage is

$$M\left(H, H_d, H_s^I, r, t\right) = N(r,t) - J\left(H, H_d, H_s^I, r, t\right). \quad (1)$$

When a borrower prepays a mortgage, she extinguishes the default option (and *vice versa*). In addition, under certain circumstances, default and prepayment can substitute for one another. Consequently, $J(H, H_d, H_s^i, r, t) < C(r, t) + D(H, r, t) + B(H, H_d, H_s^i, r, t)$.

The mortgage value must satisfy the standard partial differential equation (pde) for valuing contingent claims as a function of time and stochastic variables (Cox, Ingersoll and Ross 1985). The boundary conditions are the loci of state variables that separate one decision area from another. We deviate from the standard mortgage literature by considering H_d and H_s^i to be stochastic variables as well as H and r. We also relax the perfect market assumption and discuss how institutional factors and personal characteristics influence the choices of borrowers.

---

[9] The market value of the right to terminate by moving can take on both positive and negative values. For example, when market rates are well above the loan rate, expected future mobility-driven prepayment increases the loan value.

The Refinancing Boundary and Mediating Conditions

The traditional definition[10] of the prepayment boundary without transaction costs is

$$\text{Prepay when: } M\left(H, H_d, H_s^i, r, t\right) \geq F(t). \tag{2}$$

With perfect, frictionless markets, this rule provides the correct refinancing decision.[11] When transaction costs are introduced, it is customary to modify the rule as follows:

$$\text{Prepay when: } M\left(H, H_d, H_s^i, r, t\right) \geq F(t) + K_R(t).^{12} \tag{3}$$

Inequality (3) implies that the borrower has perfect information and acts strictly to maximize the value of the option to refinance. Under those assumptions, the borrower should exercise the option to call the debt whenever the market value of the mortgage exceeds the current balance by enough to cover the costs of refinancing. Inequality (3) treats transaction costs as a constant increase in the strike price of the call option.

However, borrowers do not exercise the option to refinance as ruthlessly as do owners of other financial options (see Green and LaCour-Little 1999). This has led some researchers, such as Stanton (1994) and Green and LaCour-Little (1999), to treat transaction costs, $K_R(t)$, in Inequality (3) as varying across borrowers. However, in both studies, even implausibly high levels of transaction costs could not fully explain the observed prepayment behavior.

Since transaction costs alone seem insufficient to explain the underexercise of the prepayment option, a number of researchers have incorporated the effects of institutional constraints on a borrower's ability to refinance. For example, Archer, Ling, and McGill (1996) (ALM) used American Housing Survey

[10] The traditional definition of mortgage value would not include housing demand and supply.

[11] At first glance, it might appear that this rule directs the borrower to exercise the option as soon as it is "in the money" and thereby ignores the time value of the option. This is not the case. The underlying security is the noncallable debt and because $M(H, r, t) = N(r, t) - C(r, t) - D(H, r, t)$, when $F(t)$ equals $M(r, t)$, it must be less than the market value of the underlying noncallable debt. The difference correctly captures the time value of the option.

[12] This rule is technically correct only when the borrower is making a one-time decision to incur the transaction costs associated with prepaying the mortgage. The correct approach would explicitly recognize the intertemporal nature of the borrower's problem and provide for an optimal intertemporal refinancing strategy. See Harding (1994) for an analysis of the optimal intertemporal refinancing strategy.

data from 1985 and 1987 to examine the influence of post-origination income and collateral constraints on prepayment behavior. ALM found higher annual payment-to-income and loan-to-value (LTV) ratios were negatively related to prepayments, after controlling for call option values. Caplin, Freeman, and Tracy (1997) found that regional recessions depressed prepayment rates by as much as 50% in states with declining property markets. Peristiani *et al*. (1997) found strong evidence that poor credit history as well as high current LTV significantly reduced the probability of refinancing. These empirical findings are intuitive, for if collateral value declines below loan balance, the borrower will not be able to refinance without infusing equity from other sources. Similarly, a borrower whose income or financial position deteriorates may be unable to refinance due to payment-to-income or credit quality constraints.

In addition, making the right refinancing decision requires ongoing monitoring of market conditions and ready access to lenders. To the extent that particular demographic groups (*e.g*., minorities) have more limited access to information or lenders, we would expect that group to have higher K_R.

The Move Boundary and Mediating Conditions

Mobility is a mechanism whereby households adjust their housing consumption to changes in circumstances (Rossi 1955). The theory assumes utility-maximizing consumers who choose housing consumption, local public goods and other consumption subject to a budget constraint. A household's decision to move is based on housing "dissatisfaction," household characteristics and exogenous circumstances (*e.g*., job or family composition changes). The dissatisfaction that ultimately results in a move is the direct result of "changes in the needs of a household, changes in the social and physical amenities offered by a particular location, or a change in the standards used to evaluate these factors" (Speare 1974, p. 175).

Hanushek and Quigley (1978) extended this framework by modeling the demand for housing services. Let H_s^i represent the bundle of housing services supplied by a particular structure, *i*. Let H_d be the household's demand for housing services. At the time a household purchases unit *i*, we assume $H_d = H_s^i$. With the passage of time, both H_d and H_s^i change randomly.[13] In the absence of transaction costs, a household would move as soon as $H_d \neq H_s^i$. In the presence of transaction costs, a move will only occur when the present value of the expected utility losses from disequilibrium exceeds the costs associated with moving. Letting $G(|H_d(t) - H_s^i(t)|)$ represent the expected present value

[13] Both of these can be viewed as stochastic processes combining continuous and jump processes.

of utility losses from disequilibrium, we have the following mathematical representation of the move decision:

$$\text{Move if: } G\left(\left|H_d(t) - H_s^i(t)\right|\right) \geq K_M(t). \tag{4}$$

The cost of moving, $K_M(t)$, is a function of demographic variables such as age of the decision-maker. Boehm and Ihlandfeldt (1986) reported that Inequality (4) has long served as the basis for most economic studies of intrametropolitan residential mobility.

Most early economic studies of mobility did not explicitly address the role of the mortgage in the decision to move—implicitly relegating any influence to $K_M(t)$. However, Green and Shoven (1986) and Quigley (1987) documented a significant "lock-in" effect arising from below-market-rate financing. They found that homeowners with low mortgage rates (relative to current market rates) delayed moving. Therefore, we extend Inequality (4) to:

$$\begin{aligned}\text{Move if: } & G\left(\left|H_d(t) - H_s^i(t)\right|\right) - K_M(t)\\ & + (M(H, H_d, H_s^i, r, t) - F(t) + K_R(t)) \geq 0.^{14}\end{aligned} \tag{5}$$

Factors that increase either the market value of the mortgage ($M(H, H_d, H_s^i, r, t)$) or the utility gain from moving ($G(t)$) encourage moving.

Turning to the mediating variables, the age of the head of household has consistently been shown to have a strong, significant negative effect on mobility (*e.g.*, Quigley and Weinberg 1977; Myers, Choi and Lee 1997). A study by South and Crowder (1998) confirmed the importance of age and found that being married, having children and currently having a job significantly deterred mobility and that mobility increased with income. They also found that, controlling for these variables, African-Americans had lower mobility than whites did.[15] Those studies that were able to track changes in family size, composition and income

[14] Using the same $K_R(t)$ as in Inequality (3) implicitly assumes that the move is to another owner-occupied unit financed with a new mortgage. Homeowners who move to a rental unit or do not finance their purchase will experience lower transaction costs.

[15] The finding of lower mobility for minorities has been reported by numerous earlier studies as well. See Quigley and Weinberg (1977). Yinger (1997) estimated that African-Americans and Hispanics paid discrimination "tax" of almost $4,000 every time they searched for a house to buy. Ross (1998) tested whether both race and job access had an independent effect on the probability of a joint residential move and job change. He found no evidence that race directly influenced the joint probability. However, because African-Americans are heavily concentrated in central cities, they had poorer job access and consequently lower job-related mobility.

found that they had a positive effect on mobility.[16] Unfortunately, our data only provides a snapshot description of the borrower at loan origination.

In summary, we expect socioeconomic characteristics to have the effects found in previous literature: negative for age and minority status, positive for income. Given our inability to measure changes in demographic variables influencing demand, we expect time in the house to measure housing dissatisfaction. Variations over time in the underlying continuous and jump stochastic processes will increase housing disequilibrium, $|H_d - H_s^i|$. The dissatisfaction effect can be either offset (in part or in whole) or reinforced by the second term in Inequality (5)—the mortgage effect.[17]

The Default Boundary and Mediating Conditions

The mortgage literature has typically modeled the default boundary in a manner analogous to the refinance boundary:

$$\begin{aligned} &\text{Default when } M\left(H, H_d, H_s^i, r, t\right) \geq H(t) + K_D(t). \\ &\text{Default when } M\left(H, H_d, H_s^i, r, t\right) \geq H(t) + K_R(t). \end{aligned} \tag{6}$$

Inequality (6) describes a policy whereby the borrower defaults if and only if the value of the house plus all costs associated with default is less than the current market value of the loan. In general, the costs associated with default include the costs of moving, the costs of a damaged credit record and the expected cost of deficiency judgements.

The options literature suggests that the default decision should be based strictly on variables that affect $M(H, H_d, H_s^i, r, t)$, $H(t)$ and $K_D(t)$. We interpret K_D broadly to allow borrower characteristics to enter via transaction costs. This suggests that the essential variables to include in a model of default will be measures of the market value of the mortgage, measures of the current house price and variables that could influence the transaction costs of default.

Recent industry research has identified an additional, non-option-based variable that influences default. Primary market lenders, credit rating agencies and the secondary market Government Sponsored Enterprises (GSEs) have all concluded that borrower credit ratings are at least as important as loan-to-value

[16] Elder and Rudolph (2000) found that change in job, divorce or the death of a spouse increased mobility.

[17] In our sample period (1993–1998), the market variations in interest rates were much smaller than those studied by Green and Shoven (1986) and Quigley (1987) and so we do not expect as strong an influence from this term.

ratios in predicting default. People who have historically been able to manage debt well tend to be better able to avoid default (see Avery *et al*. 1996). Consequently, we include the lender's proprietary initial credit score in our models.[18]

Estimation Method and Model Specification

This section discusses two alternative methods for estimating the model of borrower choice. For practical reasons, we limit our discussion to estimation techniques for which software is readily available and which have been widely used in the literature.[19] We compare the Cox proportional hazard model to the multinomial logit model for estimating competing termination risks. First, we present and critique the widely used Cox model. Then we discuss the multinomial logit approach using restructured data. This model avoids the proportionality assumption and provides explicitly for competing risks, but requires a different assumption—the independence of irrelevant alternatives (IIA). We close this section with a discussion of our choice of covariates for each of the competing risks.

Cox Proportional Hazard Model

Time to failure is the underlying random variable used in the Cox proportional hazard model (Kalbfleisch and Prentice 1980). The model begins with a baseline time profile of the probability of termination conditional on the loan having survived to time t, $h_0(t)$. This baseline refinancing hazard can be shifted up or down by a factor that depends on the covariates, Z_{it} for observation i at time t:

$$h(t \mid Z_{it}) = h_0(t)e^{z_{it}\beta}. \tag{7}$$

The vector of coefficients is estimated from a quasi-likelihood function:[20]

$$L_p(\beta) = \prod_{t=1}^{T} \prod_{l(t)}^{L(t)} \frac{e^{Z_{l(t)}\beta}}{\sum_{k \in R(t)} e^{Z_k\beta}}, \tag{8}$$

[18] The initial credit score reflects the borrower's credit at the time of loan application. Credit scores change over time; however, the lender providing our data did not track credit scores after loan origination.

[19] Han and Hausman (1990), Sueyoshi (1992) and McCall (1996) have suggested a maximum likelihood estimator approach that estimates competing risks models simultaneously, accounts for the fact that the risks may be correlated and allows the covariates to be time varying. Deng, Quigley and Van Order (2000) have used this approach to model the competing risks of mortgage prepayment or default. The software to estimate these models is not commercially available.

[20] This function is a quasi-likelihood because the probabilities are each between 0 and 1 but they do not sum to 1. The Cox model assumes that each quasi-probability is independent of every other probability; there is no path dependence in the model.

where $l(t) = 1, 2, \ldots, L(t)$ indexes the $L(t)$ loans that are prepaid in month t. If $L(t) = 0$, then that term of Equation (8) is ignored. $R(t)$ is the set of loans that remain in the risk set at time t: all loans that have not terminated (through refinancing, moving or defaulting) at the *beginning of* time t. Thus, the set $R(t)$ includes the set $L(t)$, future refinances, moves and defaults, as well as all loans that are right censored at the end of the observation period. The notation in Equation (8) allows for the typical situation where loans are originated at different times and fail at different times, with irregular spacing between failures. Note that $h_0(t)$ cancels out of the numerator and the denominator of Equation (8).

Limitations of the Cox Model

The literature documents several weaknesses with the Cox model. The most significant weakness is the limited manner in which it handles competing risks. The Cox model estimates three separate equations. For any one of these risks at time t, the numerator of Equation (8) contains only the loans that terminate because of that risk whereas the denominator contains all the loans that are at risk. As loans terminate for any reason, the number of loans in the denominator declines. Consequently, the three hazard functions that we estimate implicitly assume independence after controlling for the explanatory variables. This contradicts the interdependencies in our model as defined by Inequalities (3), (5) and (6).

A second problem with the Cox model is the proportionality assumption. If the move hazard for 50-year old borrowers is one-half that for 30-year old borrowers, the proportionality assumption requires that this ratio remains constant as the loan seasons. A common method for dealing with violations of the proportionality assumption is to add new covariates to the model by interacting the log of mortgage age with the variables of concern (Quigley 1987). If the number of variables violating the assumption is too large, alternative modeling choices may be more appropriate.[21]

Multinomial Logit

The multinomial logit model treats the dependent variable as a polytomous qualitative choice variable. We begin our discussion of this approach with a review of the close relationship between logit (with a dichotomous choice) and the Cox proportional hazard model.

[21] Other concerns with the Cox model include modeling unobserved heterogeneity (usually related to omitted variables) and dealing with "ties" (two terminations at the same time.)

Consider a single prepayment risk. Previous literature shows that bivariate logit with a restructured data set provides a convenient method for dealing with some of the limitations of the Cox model (Bergström and Edin 1992; Narendranathan and Stuart 1993; Jenkins 1995). The information for each loan is restructured to include one observation for each time period in which that loan is active (*i.e.*, from origination up to and including the period of termination). Once the data are restructured, the likelihood function is identical, in discrete time, to the continuous-time likelihood function for the Cox model.[22] This discrete time model is often used for grouped data where a large number of tied failure dates can cause computational problems for the Cox model (Bergström and Edin 1992).

By using indicator variables for time, we can estimate a flexible baseline hazard:

$$h_{it} = \frac{e^{\theta(t)+\beta' X_{it}}}{1 + e^{\theta(t)+\beta' X_{it}}} \Rightarrow \ln\left(\frac{h_{it}}{1 - h_{it}}\right) = \theta(t) + \beta' X_{it}, \tag{9}$$

where h_{it} is the probability of failure for individual i at time t and $\theta(t)$ is the inner product of estimated coefficients and a vector of time indicator variables with one element for each time period in the data. Note that the covariates may be time varying.

The multinomial logit model provides a logical extension of this reasoning to the competing risks model. The restructured data and the use of θ(t) to model the baseline hazard generalize from the bivariate logit model. Thus, letting Y_{it} represent the ith borrower's decision at time t, the log-likelihood function is

$$\ln L = \sum_{t=1}^{T} \sum_{i=1}^{n_t} \sum_{j=0}^{3} d_{ijt} \ln(\Pr(Y_{it} = j)), \tag{10}$$

where

$$\Pr(Y_{it} = j) = \frac{e^{\theta(t)+\beta_j' X_{it}}}{1 + \sum_{k=1}^{3} e^{\theta(t)+\beta_k' X_{it}}} \quad \text{for } j = 1, 2, 3$$

and

$$\Pr(Y_{it} = 0) = \frac{1}{1 + \sum_{k=1}^{3} e^{\theta(t)+\beta_k' X_{it}}}.$$

[22] See Jenkins' Equation (11) compared to his Equation (10). Cox (1972) was the first to show that logit is the discrete-time analog of the proportional hazard model.

In Equation (10), n_t is the number of observations in the restructured data at time t (t =1, ... T), j indexes the possible choices (continue, default, refinance, move) and d_{ijt} equals one when the alternative j is chosen in the ith observation at time t, otherwise zero.

Competing risks are included in Equation (10) by having probabilities that must sum to one. Thus, an increase in the probability of one risk must necessarily be associated with a decline in the probability of at least one other risk. This competition in probability space is an important advantage over the Cox model. Because of this, the multinomial logit model gives different results than the Cox model.

The multinomial logit model requires independence of irrelevant alternatives (IIA): The odds ratio for any pair of choices is assumed independent of any third alternative. Elimination of one of the choices should not change the ratios of probabilities for the remaining choices. Choices that are close, in the sense that their utilities are highly correlated, violate the IIA assumption. The multinomial logit model also assumes that choices at any point in time are independent of those at any other point in time.[23] Limited path dependence may be introduced into the model by adding explanatory variables (*e.g.*, a burnout variable).[24]

Choice of Explanatory Variables

Table 1 provides a list of explanatory variables included in the choice models and the expected sign of the impact. We discuss the variables by group: Loan Characteristics, Housing Market and Economic Conditions, Borrower Characteristics, and Time and Season Indicators.

Loan characteristics. As described in the model, $M(H, H_d, H_s^i, r, t)$ should have a positive effect on all three termination hazards. Previous empirical research on mortgage terminations has used easily calculated proxies for $M(H, H_d, H_s^i, r, t)$. Richard and Roll (1989) and Pavlov (2000b) used the ratio of the current interest rate to the rate that prevailed at the time of origination. Deng, Quigley and Van Order used the ratio of $N(r, t)$ to $F(t)$ in a sequence of papers (see, for example, Deng, Quigley and Van Order 2000). While both variables are correlated with $M(H, H_d, H_s^i, r, t)$, in this paper we have tested an alternative proxy that is more closely related to $M(H, H_d, H_s^i, r, t)$.

---

[23] The same assumption is implicit in the Cox model.

[24] We experiment with the standard measures of mortgage burnout and do not find them to be significant in our models.

Table 1 ■ Variable list with theoretical effects and actual multinomial logit results.

	Postulated Effect			Actual Effect		
	Refinance	Move	Default	Refinance	Move	Default
Loan characteristics						
Market price of loan	+	+	+	$+^{***}$	$-^{*}$	0
Original loan balance	+	+	+	$+^{***}$	0	0
15 year loan (indic.)	0	−	−	0	$-^{*}$	0
Points	0	−	0	0	$-^{***}$	0
Refinance in 1993 or 1994	0	?	0	$-^{***}$	0	$+^{*}$
Housing mkt. and econ. conditions						
Current loan-to-value ratio	−	−	+	$-^{***}$	0	$+^{**}$
Probability of negative equity	−	−	+	0	$-^{**}$	0
Cum. house price appreciation						
× Indicator (age <40 yrs.)	0	0	0	0	0	0
× Indicator (age >40 yrs.)	0	−	0	$-^{*}$	$-^{***}$	0
County unemployment Rate	−	−	+	$-^{***}$	0	0
Borrower Characteristics						
Borrower age	0	−	0	0	$-^{**}$	0
Minority indicator	−	−	0	$-^{***}$	$-^{***}$	0
Borrower income	−	+	0	$-^{**}$	$+^{***}$	0
Obligation ratio	0	0	+	0	0	0
Low credit score (indic.)	−	−	+	$-^{**}$	0	$+^{***}$
High credit score (indic.)	0	0	−	0	0	0

Notes: The table summarizes the postulated effects on the probability of refinancing, moving and default for all explanatory variables. (See the section on Modeling Refinancing, Moving and Default.) The Estimation and Model Specification section discusses how we measure the variables listed. The market price of the loan is the market value of the borrower's remaining payments. Points measure the fees paid by the borrower at origination to "buy down" the mortgage rate. The refinance variable identifies loans that were originated to refinance an existing loan, not finance a new purchase. Current loan-to-value ratio, probability of negative equity and cumulative house price appreciation measure how the borrower's equity changes over time. The actual effects reported to the right are based on the multinomial logit model presented in Table 4.
* = significant at the 10% level;
** = significant at the 5% level;
*** = significant at the 1% level.

Under perfect markets assumptions, $M(H, H_d, H_s^i, r, t)$ can be calculated precisely by solving the standard pde for asset valuation (Cox, Ingersoll and Ross 1985) subject to appropriate boundary conditions. However, even if we restrict the state variables to $H(t)$ and $r(t)$ (ignoring the stochastic processes underlying $G(t)$, there are no general closed-form solutions for the resulting pde. While recent computational advances (Hilliard, Kau and Slawson 1998) facilitate solving the two state variable pde for mortgage values numerically, these numerical methods are still very time intensive.[25]

We use the closed-form formula for mortgage valuation developed by Collin-Dufresne and Harding (1999) (CDH). The closed form formula is based on the standard result (Cox, Ingersoll and Ross 1985) that the equilibrium value of any contingent claim is equal to the expected discounted value of future cash flows when the expectation is taken with respect to the equivalent martingale probability measure. Letting $CF(H, H_d, H_s^i, r_\tau, \tau, \phi_\tau)$ represent future cash flows and E^Q the expectation under the equivalent probability measure, we have

$$M\big(H, H_d, H_s^i, r, t\big) = E_t^Q\left[\int_t^T e^{-\int_t^\tau r_s ds} CF\big(H, H_d, H_s^i, r_\tau, \tau, \phi_\tau\big)\, ds\right].^{26} \quad (11)$$

To derive the closed form result, one must limit the model to a single state variable (the interest rate) and use a loglinear model of future cash flows with no path dependent explanatory variables. The closed form formula provides an approximation to $M(H, H_d, H_s^i, r, t)$. It does not explicitly incorporate variations in $H(t)$, $H_d(t)$ or $H_s^i(t)$. However, there is an important advantage to using Equation (11) relative to using $N(r, t)$. Because it incorporates both the possibility of future prepayment and the fact that not all mortgagors exercise their options ruthlessly, the values generated by the formula exhibit negative convexity and are not capped at par plus transaction costs. CDH showed that the closed-form mortgage values explained 85% of the variation between actual mortgage prices and $N(r, t)$. We test all three candidate measures in the choice model and present the tables based on the closed-form valuation. The differences in results from using the other proxies for $M(H, H_d, H_s^i, r, t)$ are discussed in the Results section.

The original loan balance is included in the model to capture the scale effect of loan size. A large mortgage with a premium price provides a larger dollar incentive to refinance, move or default than does a small mortgage at the same price (positive signs in Table 1). The effect should be most significant in the

[25] Based on correspondence with the authors, we understand the calculations reported in Hilliard, Kau, and Slawson (1998) required several hours for each mortgage valuation. Advanced numerical techniques such as "pruning" the lattice are estimated to lower this time to a few minutes or less. Nevertheless, even if the values could be generated in a minute, calculating more than 140,000 different values would require approximately 100 days of calculations.

[26] We normalize the CDH values by the outstanding balance so it represents a standard price per $100 of principal.

equation for refinancing since there are no offsets to the mortgage effect in Inequality (3), whereas the mortgage effect can be offset by changes in $G(t)$ or $H(t)$ for the other two choices.

We include two variables related to the borrower's choice of loan product as explanatory variables: the choice of maturity (15 years *vs*. 30 years) and the points[27] paid by the borrower. Several authors have suggested that borrowers signal their mobility via their choice of products (Dunn and Spatt 1988; Brueckner 1992; Sa-Aadu and Sirmans 1995). Because of the inherent rapid amortization, borrowers who choose 15-year loans are believed to signal low mobility and low default probability. Recent literature on mortgage points (Brueckner 1994 and Stanton and Wallace 1998) has argued that borrowers who expect to move in the near future will generally find low point/high coupon rate products provide the lowest effective cost of borrowing while borrowers with long expected tenure should choose the opposite combination. This reasoning suggests a negative effect of points on the move probability.

We define an indicator of loan purpose at origination with a value of one indicating refinancing an existing loan. This identifies loans where the borrower has lived in the home for a period longer than the loan age, thereby increasing the likely level of housing dissatisfaction. However, refinancing could also signal private information about expected future mobility. A borrower who plans to move in the near future needs a much larger refinancing incentive than one who plans to stay in the home until the current loan matures. Consequently, the sign on the refinance indicator is ambiguous in the move equation.

Housing market and economic conditions. We include three variables that measure the effect of $H(t)$: the current loan-to-value ratio, the probability of negative equity and cumulative appreciation (or depreciation) in house value since origination.[28] Most past researchers have used the indices of aggregate house price changes published by Office of Federal Housing Enterprise Oversight or one of the secondary mortgage market GSEs. These indices are based on the sample of conforming loans purchased by the GSEs and are aggregated

Au: Palov 20001 or 2000b or both ?

[27] The loan data does not contain accurate information on points paid by the borrower for all loans. We estimate the points for each loan following a methodology suggested by Pavlov (2000). We develop a model of the coupon rate based on current treasury rates and loan and borrower characteristics. The residuals from this equation provide a measure of points since a borrower paying a rate substantially below the predicted rate must have "bought down" the rate by paying above average points.

[28] Curtailments and second mortgages also affect the homeowner's equity. We cannot observe curtailments or secondary financing; consequently, our discussion of current loan-to-value ratios is based on the ratio of amortized original loan balance to estimated current house value.

over an entire metropolitan region. In this paper, we model the local housing market using a transaction database containing every house transaction in the local market including those financed with jumbo loans.

We model local housing markets with locally weighted regressions, one for each point on a grid placed over the locations of the properties that secure the loans, determined from addresses with GIS software. We then weight each of the sold properties (from the transaction's database described below) inversely with a function of distance from each grid point. Characteristics of the sold properties are controlled with ordinary least squares (see Appendix and Clapp 2000).

We expect the borrowers with high current loan to value ratios and high probability of negative equity to face additional constraints in their moving and refinancing decisions. High values of these variables should be positively related to the default probability (Inequality (6)).

Cumulative house price appreciation can influence the borrower's choice through the investment aspect of a home. To the extent that the current house represents a good financial investment, borrowers may be reluctant to sell and move. We interact the cumulative house price appreciation with an indicator of the borrower's age (above and below 40 years of age) because Kiel (1994) found that for older homeowners, significant appreciation deterred mobility.[29] High local unemployment is expected to result in more borrowers being unable to qualify for a new loan (negative effect on move and refinance) and more borrowers who have difficulty servicing the existing debt.

Borrower characteristics. We include age, race, income, obligation ratio and credit score in the model of borrower choice. All borrower characteristics are measured as of the time of origination. Based on the empirical mobility literature, we expect borrower age to have a negative effect on the move probability but little or no effect on the other probabilities. To the extent that minorities have more limited opportunities to move and refinance (Yinger 1997), we expect negative effects of minority status on both move and refinance probabilities. We use a single indicator variable to identify borrowers whose race is recorded as African-American, Hispanic or Asian.[30]

---

[29] One explanation for the different effect of appreciation is that younger homeowners (whose demand for housing services is increasing because of life cycle effects) view accumulated appreciation as a necessary condition for mobility because it provides the necessary downpayment on a larger, more expensive home. Older borrowers (who have more stable or declining housing demand) view their home more as an investment. From an investment perspective, rapid appreciation may deter the homeowner from selling.

[30] We also used three separate indicator variables and found the coefficients on all three had the same sign and were similar in magnitude.

As noted earlier, we expect high-income borrowers to have higher opportunity costs for refinancing and therefore a negative effect on the probability of refinancing. Although economic theory suggests mobility should be driven by changes in income, not the level, previous studies have found a positive relationship. However, there is little reason to expect a strong effect on default. We include the credit score of the borrower as a pair of indicator variables. The low score indicator flags borrowers with scores below 800—approximately the bottom 10% of the sample. The high score indicator flags borrowers with credit scores greater than 1000—a group that includes approximately 50% of the borrowers. Borrowers with weak credit are more likely to be constrained in their ability to refinance and move. To the extent that lenders apply more stringent underwriting to refinance loans than purchase money loans, the effect of bad credit will be stronger on the refinance probability. Based on previous research on default, we expect low credit scores at origination to be related to higher default probabilities.

Time and season indicators. We include a measure of loan age and season of the year in the multinomial choice model. These indicators allow a flexible estimation of the underlying base hazard rate for each probability. The Cox models do not need these indicators.[31]

Data

Loan Data

A major loan originator/servicer that prefers anonymity provided loan level information on 2,057 fixed-rate residential mortgage loans originated in 1993–1994 and tracked through 1998. The lender was among the top 20 loan servicing firms nationwide during this originating time period, utilizing multiple origination channels and a proprietary credit score. We selected loans from three representative California counties.[32] We did not prescreen the loans in any fashion. Rather, we used the entire population of fixed rate mortgages originated/purchased by the lender during 1993 and 1994 in the three counties selected. We eliminated a small portion (3.5%) of the loans from the sample if the information was incomplete, if they had maturities less than 15 years or if

---

[31] The baseline hazard is estimated in the Cox model as the probability of a particular risk when all the explanatory variables take on the value of zero. In order to facilitate the estimation of the baseline hazard, the Cox model is estimated after transforming all continuous variables to have mean zero and standard deviation of one.

[32] The three counties are Contra Costa (a suburban county in the San Francisco Bay area), Los Angeles and Orange Counties (two contiguous counties in Southern California).

they failed to meet certain edit checks.[33] The remaining data includes information on 1,985 fixed-rate mortgages with both 30-year and 15-year maturities. Approximately 79% of the loans were originated to refinance an existing mortgage loan on the same property while 21% were loans for home purchases. The majority of the loans (64%) were originated by correspondents or brokers and purchased by the lender providing the data, the lender originated the remainder. Table 2 provides summary statistics describing the sample.

Loans were underwritten according to standard policies in effect during 1993 and 1994, including scoring loans using an internally developed mortgage credit scoring model that adds certain borrower and loan characteristics, including LTV, to traditional credit bureau measures in order to estimate borrow creditworthiness.

Because of high housing costs in California, the loans had an average original loan balance of $167,600. Approximately 73% had original loan amounts below the GSE limits for 1993 and 1994 making them eligible for purchase by Fannie Mae and Freddie Mac. The average interest rate on the loans at origination was 7.48%.

Figure 1 describes the interest rate environment that prevailed during our observation period. Interest rates were at historically low levels in late 1993 and early 1994. By the end of 1994, rates increased more than 2% over that low. By the end of the observation period, mortgage rates had returned to the levels of 1993.

Table 3 contains values for the estimated market price of the loans and other time-varying covariates. All data are quarterly, the smallest time interval common to all variables.[34] The table shows how these covariates change, on average, over time compared to the values at origination.

As of December 31, 1998, 27 loans (1.4%) had terminated by default and 573 loans (28.9%) had terminated by prepayment. Although the raw loan data provided by the lender does not distinguish between refinancers and movers, we use house transaction data (described below) to estimate that moves triggered 252 of the prepayments and refinancing resulted in the remaining 321 prepayments.

[33] For a small number of the loans, we estimate the borrower's age and credit score based on other recorded values.

[34] Unemployment and house price indices are estimated at the quarterly level to avoid excessive noise. Monthly data could be smoothed, but this would introduce time dependence.

Table 2 ■ Descriptive statistics on loan and borrower characteristics by loan status.

Variable	Total	Active	Refinanced	Moved	Defaulted
Home value ($,000)	296.0	286.1	343.0	304.5	167.2
	(209.4)	(204.2)	(234.8)	(201.9)	(70.7)
Loan balance ($,000)	167.6	158.5	205.3	172.0	146.5
	(121.8)	(115.9)	(151.6)	(106.1)	(58.9)
Original loan-to-Value	60.5%	59.9%	61.7%	59.7%	89.1%
(LTV)	(22.7%)	(22.8%)	(22.8%)	(20.4%)	(12.7%)
LTV > 90%	8.3%	9.1%	4.7%	3.2%	59.3%
LTV < 60%	45.4%	46.6%	41.4%	48.4%	3.7%
Interest rate	7.48%	7.41%	7.79%	7.43%	8.16%
	(0.65%)	(0.62%)	(0.70%)	(0.56%)	(0.65%)
Estimated points	2.00%	2.10%	1.64%	1.87%	2.11%
	(1.43%)	(1.42%)	(1.36%)	(1.45%)	(1.47%)
Mortgage payment	1279.13	1206.57	1593.82	1295.80	1104.53
	(948.55)	(873.75)	(1273.04)	(808.38)	(477.41)
Original refinance	79.1%	80.6%	70.4%	85.7%	40.7%
15-year indicator	31.0%	33.1%	27.1%	27.4%	3.7%
Borrower age	46.7	47.1	46.4	46.1	38.2
	(11.2)	(11.4)	(11.1)	(9.3)	(11.2)
Minority indicator	23.2%	27.4%	14.3%	9.9%	40.7%
Borrower monthly	8079	7674	8987	9484	4961
income	(9033)	(8424)	(8051)	(12854)	(2730)
Obligation ratio	30.1%	29.9%	31.3%	29.4%	36.6%
	(9.6%)	(9.7%)	(9.2%)	(9.6%)	(5.7%)
Mortgage credit score	1002.7	1002.6	1001.8	1036.7	703.4
(CRDSCOR)	(170.5)	(171.8)	(161.0)	(141.2)	(174.4)
CRDSCOR > 1000	63.9%	62.7%	66.0%	72.6%	18.5%
CRDSCOR < 800	10.3%	10.4%	8.4%	4.8%	77.8%
Number of loans	1985	1385	321	252	27
% of total	100.0%	69.8%	16.2%	12.7%	1.4%

Notes: Means are above; standard deviations are below in parentheses. All loans were originated in 1993 and 1994. Values are reported as of the time of loan origination. An active loan is one that had not terminated by 12/31/98. The loans labeled Refinanced and Moved are classified using house sale transaction data to identify movers (see Data section). For loans originated to purchase a home, the reported home value is the minimum of the purchase price and the appraised value. For loans originated as refinance loans, the home value is the appraised value.

The Transactions Data

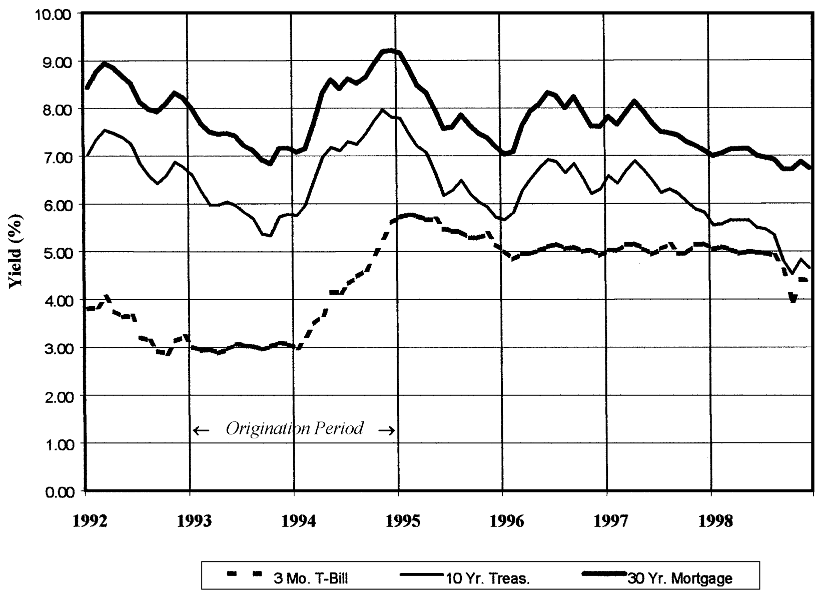
We purchased six years of transactions data from the California Market Data Cooperative, Inc. (CMDC). CMDC collects, verifies and, if necessary, corrects all property transactions from the county records. The sales for Contra Costa, Los

Table 3 ■ Descriptive statistics on time varying covariates.

Variable	At Orig.	12/31/1993	12/31/1994	12/31/1995	12/31/1996	12/31/1997	12/31/1998
Market price of mortgage	100.0	101.3	93.8	96.6	99.1	99.9	102.6
($ per $100 of principal)	(1.8)	(1.9)	(2.5)	(2.3)	(2.2)	(2.1)	(2.0)
Noncallable price	99.7	101.2	92.2	95.6	99.0	100.3	104.5
($ per $100 of principal)	(3.9)	(4.1)	(4.8)	(4.9)	(4.9)	(4.9)	(4.8)
Ratio of mortgage rates	100.0%	103.0%	82.4%	102.4%	97.6%	104.5%	110.4%
(current to origination)	(0.0%)	(4.3%)	(6.9%)	(8.6%)	(8.2%)	(8.8%)	(9.4%)
Current LTV	60.5%	56.8%	60.2%	59.7%	58.1%	55.1%	49.4%
	(22.7%)	(21.5%)	(23.0%)	(23.7%)	(24.9%)	(26.2%)	(25.0%)
Current LTV >90%	9.7%	4.4%	11.5%	11.5%	7.9%	5.2%	4.1%
Current LTV <60%	46.1%	51.5%	46.2%	47.8%	51.0%	57.7%	68.2%
Probability of	5.3%	3.6%	4.7%	4.6%	4.4%	4.2%	3.8%
negative equity	(10.8%)	(8.8%)	(10.5%)	(12.1%)	(13.6%)	(14.7%)	(13.2%)
Probability of negative	15.4%	9.9%	13.5%	11.4%	9.3%	7.8%	7.7%
equity >90 percentile	(36.1%)	(29.8%)	(34.2%)	(31.8%)	(29.1%)	(26.9%)	(26.7%)
Cumulative house price	0.0	−3.2	−5.8	−6.0	1.0	15.5	41.8
appreciation ($000's)	(0.0)	(8.5)	(12.8)	(16.9)	(24.3)	(37.4)	(59.2)
Unemployment rate	8.3%	8.4%	7.1%	6.9%	6.3%	5.3%	5.3%
	(1.7%)	(1.7%)	(1.4%)	(1.4%)	(1.7%)	(1.4%)	(1.6%)
Age of loans	0.0	2.3	5.1	9.1	13.1	17.1	21.1
(in quarter years)	(0.0)	(1.0)	(1.9)	(1.9)	(1.9)	(1.9)	(1.9)
Number of active loans	1985	1309	1943	1879	1798	1686	1474

Notes: Means are above; standard deviations are below in parentheses. This covers the six possible years of early termination, compared to the values at origination. The market price of the mortgage, noncallable price and ratio of rates vary primarily with interest rates. The market price is calculated using the CDH closed-form formula (normalized to equal par at origination). The noncallable value is the present value of the remaining mortgage payments when discounted at the current mortgage rate. The ratio of rates is the interest rate on the loan divided by the current mortgage yield (×100). Current LTV is the estimated current loan-to-value ratio calculated using the amortized loan balance divided by the estimated house value (×100). The probability of negative equity is the probability that the true house value is less than the loan amount, based on the standard error of the estimate of house price and the normal distribution.

Figure 1 ■ Interest rates January 1992 through December 1998. The figure shows the 3-month Treasury Bill rate (discount basis), the 10-year Treasury rate and the Freddie Mac primary market mortgage rate (30-year fixed-rate) over the observation period. The mortgages studied were originated during 1993 and 1994 and were monitored through 1998.



Angeles and Orange counties are from the period from January 1993 through December 1998. CMDC data contain a full street address for each property that sold as well as the date of sale, sales price, appraised value and recorded first mortgage loan. They also contain considerable detail on the property, including square footage, bathrooms, bedrooms and year built.

Identifying Movers and Refinancers

We match the full street address of the collateral underlying the loan, the origination date, loan amount and appraised value to the housing transactions data to identify movers. When we find a house sale in the transaction data with the same address and a sale date close to the date of loan termination, we identify the prepayment as being the result of a move. When we find no match, we conclude that the prepayment was caused by a refinance.[35] Because we only

[35] Since refinances are identified by the *failure* to find a match, it is possible that some loans are incorrectly classified as refinances because of a defect in the matching program.

have transaction data recorded by CMDC through December 1998, to the extent that there is a lag in reporting transactions in the CMDC data, this procedure will tend to underestimate moves (and overestimate refinancing) in the last two quarters of 1998.

Results

Refinance, Move and Default Models

Table 4 presents the multinomial logit and Cox proportional hazard estimates of coefficients for the refinance, move and default models, respectively.[36] Comparing signs and significance of coefficients[37] across the two models, it becomes apparent that there are few differences and no statistically significant coefficients have opposite signs: The coefficient estimates are robust to the method used to generate the estimates. The discussion that follows will focus on the multinomial logit estimates because they allow for competing risks.

The effects in Table 4 are consistent with the postulated effects in Table 1.[38] We find the variables measuring the market value of the loan (market price and original loan balance) to be very influential in the refinance model and insignificant in the move and default models. Consistent with the literature that predicts borrowers will self select their rate/point combination based on expected future mobility, the coefficient on estimated points is negative and significant in the move model but not significant in the refinancing or default models. We also find that loans that were originated as refinance loans in 1993 or 1994 are somewhat slower to refinance again. This could reflect learning on the part of the borrower about the real cost of refinancing.

To address this problem, we augmented the automated search algorithm with a manual effort to find a reasonable match to the property. In addition, we applied this combined automated and manual match procedure to try to identify a record in the CMDC data for every origination identified as a purchase transaction in 1994. The procedure generated matches for 86% of the purchase money mortgages in Contra Costa, 80% in Los Angeles County and 70% in Orange County. The complements of these percentages are upper bounds on the misclassification of prepayments.

[36] The coefficients for year and quarter indicators in the multinomial logit model are not reported in Table 4. We use those coefficients to generate a "baseline" hazard for the multinomial logit model comparable to the Cox baseline. These baseline hazards are presented in Figures 2 and 3.

[37] The magnitudes of the reported coefficients vary across the models because we rescale the data in the Cox model (to mean zero and standard deviation one) to facilitate estimation and calculation of the baseline hazard.

[38] The negative coefficient on market price in the move model is the only sign reversal from Table 1 that is marginally significant. Closer inspection of the data shows that this result is influenced by a group of 1993 borrowers who moved very quickly in 1995 when market prices of their loans were low.

Table 4 ■ Results from alternative competing risk models.

	Refinance Models		Move Models		Default Models	
	Multinomial Logit	Cox Prop. Haz.	Multinomial Logit	Cox Prop. Haz.	Multinomial Logit	Cox Prop. Haz.
Loan characteristics						
Market Price of Loan	27.194***	0.882***	−6.686*	−0.387***	10.114	0.282
	(2.738)	(0.113)	(3.536)	(0.143)	(8.869)	(0.373)
Original loan balance ($00,000)	0.293***	0.334***	−0.095	−0.112	−0.081	−0.048
	(0.056)	(0.060)	(0.073)	(0.082)	(0.282)	(0.303)
15-year loan indicator	−0.009	−0.005	−0.289*	−0.299**	−0.981	−0.812
	(0.160)	(0.154)	(0.175)	(0.171)	(1.147)	(1.143)
Points (estimated)	−0.010	0.020	−0.160***	−0.246***	0.079	0.109
	(0.044)	(0.062)	(0.062)	(0.089)	(0.133)	(0.187)
Original refinance indicator	−0.547***	−0.461***	0.207	0.228	1.130*	1.238**
	(0.168)	(0.164)	(0.203)	(0.202)	(0.664)	(0.622)
Housing mkt. and econ. conditions						
Current loan-to-value	−0.014***	−0.301***	0.005	0.148	0.034**	0.835**
	(0.004)	(0.101)	(0.005)	(0.115)	(0.017)	(0.395)
Prob. neg. eq. >90% pctile ind.	0.209	0.128	−0.855**	−0.970***	−1.462	−1.407
	(0.227)	(0.226)	(0.347)	(0.360)	(1.309)	(1.282)
House price appr. ind. (Age <40, $,000)	0.001	−0.001	0.003	0.001	0.006	0.003
	(0.002)	(0.002)	(0.003)	(0.003)	(0.010)	(0.009)
House price appr. ind. (Age >40, $,000)	−0.003*	−0.003**	−0.006***	−0.007***	−0.006	−0.007
	(0.001)	(0.001)	(0.002)	(0.002)	(0.007)	(0.006)
Unemployment rate	−0.100***	−0.168**	−0.053	−0.090	0.275	0.741
	(0.036)	(0.067)	(0.041)	(0.077)	(0.182)	(0.480)

Borrower characteristics						
Borrower age	−0.003	−0.046	−0.014**	−0.163**	−0.019	−0.207
	(0.006)	(0.066)	(0.006)	(0.068)	(0.026)	(0.291)
Minority indicator	−0.700***	−0.717***	−1.074***	−1.076***	−0.309	−0.323
	(0.169)	(0.167)	(0.210)	(0.209)	(0.490)	(0.479)
Borrower income	−0.017**	−0.150**	0.019***	0.182***	−0.036	−0.290
	(0.008)	(0.067)	(0.005)	(0.047)	(0.071)	(0.570)
Obligation ratio	0.001	0.000	0.007	0.068	0.045	0.356
	(0.006)	(0.061)	(0.007)	(0.072)	(0.029)	(0.264)
High credit score indicator	−0.153	0.011	0.247	0.363**	0.961	1.150
	(0.137)	(0.134)	(0.161)	(0.162)	(1.122)	(1.124)
Low credit score indicator	−0.541**	−0.698***	−0.143	−0.244	3.658***	3.503***
	(0.269)	(0.258)	(0.342)	(0.342)	(1.142)	(1.172)
Constant	−31.229***		0.361		−44.286	
	(2.897)		(3.489)		—	
Number of observations	38301	38301		38301		38298
Log likelihood	−3181	−2284		−1833		−156
χ^2 (d.f.)	45417 (72)	204 (16)		93 (16)		158 (16)
Prob. > χ^2	0.0000	0.0000		0.0000		0.0000
Pseudo R^2	0.1144					

Notes: The choice is continue to pay, default, refinance or move. The market price of the loan is based on the CDH closed-formula. Current loan-to-value, probability of negative equity and cumulative house price appreciation are based on the estimated current house price. Borrower characteristics are measured at origination.

* = significant at the 10% level;

** = significant at the 5% level;

*** = significant at the 1% level.

Among housing market variables, we find the two measures of low borrower equity (high current loan-to-value or high probability of negative equity) to be significantly negatively correlated with both refinancing and moving, but strongly positively correlated with default. These results confirm the previous findings reported by Archer, Ling and McGill (1996) and Caplin, Freeman and Tracy (1997) regarding institutional constraints on refinancing, and extend those findings to include an effect on mobility. The positive correlation of high LTV with default is well established (see Vandell 1993, 1995 for surveys).

We also find support for the finding of Kiel (1994) that strong house price appreciation deters older homeowners from selling and moving. High local unemployment rates are also negatively related to refinancing, consistent with results on unemployment in Green and LaCour-Little (1999). Our data shows no unemployment effect on moving or default, contrary to the view that higher unemployment rates are positively related to default (Deng, Quigley and Van Order 2000). We also find a strong credit score effect in the default equation. Specifically, the low credit score indicator is strongly positively related to default, consistent with the conventional wisdom on credit scoring. However, we have so few defaults in our data that neither of these default results is particularly compelling.

Our results on the effect of borrower characteristics on prepayments add to a small, but growing, body of empirical literature. As expected from both theory and previous empirical work, we find strong effects of borrower age (negative), race (negative for minorities) and income (positive) on the probability of moving. These results are broadly consistent with Archer, Ling and McGill (1997). Borrower characteristics have less influence in the refinance model, consistent with the findings of LaCour-Little (1999). As expected, borrower income enters the refinance model with a negative sign. Minority status is also negatively correlated with the decision to refinance, possibly reflecting more limited access to lenders by minorities or higher search costs, consistent with Kelly (1995).[39] Finally, we find that borrowers with poorer credit histories are less likely to refinance, consistent with Perisitiani *et al*. (1997).

Value of Separating Movers from Shuckers

We test the statistical significance of separating movers and shuckers by testing whether the estimated vectors of coefficients from the move and refinance models are identical using both a Wald and likelihood ratio test.[40] Both tests strongly

---

[39] The same results hold when we separate the three minority groups; this is also true for the move and default equations.

[40] See Long (1997, p. 162) for details.

reject the null hypothesis that these two outcomes are indistinguishable. To test the economic significance of separating movers and refinancers in mortgage prepayment analysis, we combine them into a single prepayment variable and rerun the multinomial logit model with only three choices: continue to pay, prepay, or default. Table 5 compares the results from this combined termination model with the full competing risks model: The move, refinance and default equations from Table 4 are repeated in Table 5 to facilitate comparison.

Table 4 shows that two variables (market price and income) have significantly different signs in the move and refinance model. We expect that combining the two hazards would make it difficult to estimate the effect of these variables on prepayment. Six other variables[41] are significant in either the move or refinance equation but not the other. We expect the absolute value of the coefficients on these variables to be shifted toward zero when movers and refinancers are combined. Such a shift in coefficients would distort predictions of aggregate terminations when, for example, a change in market conditions favors moving but not refinancing or when the cost of refinancing declines without significantly altering the cost of moving.

Table 5 confirms these expectations. The coefficient on borrower income in the combined model for prepayment is not significantly different from zero. Consequently, a one standard deviation change in borrower income has almost no effect on the predicted probability of prepayment, but it does reduce the probability of default. When these two effects are added together, the aggregate termination rate (holding all other variables at their means) declines.[42] In the full MNL competing risks model, an increase in income has a significant positive effect on the move probability, offsetting the negative effect on refinancing and an insignificant negative effect on default: The overall termination rate increases slightly with income.[43]

The coefficient on the market price of the loan is positive and significant in both the refinance model and the combined prepayment model—but the magnitude of the coefficient in the combined model is reduced by a factor of two

[41] Those variables are: estimated points, original refinance indicator, house price appreciation (age > 40), borrower age and both credit score measures.

[42] Estimated effects on probabilities of termination for a given reason are total effects, after allowing for competition from the other risks. We use Long's (1997) equation (6.13, p. 165). The MNL model, unlike the Cox model, allows for these competing risks during estimation by requiring that termination probabilities sum to one.

[43] The effect of a one standard deviation increase in income on the overall termination rate is different at different values of the other explanatory variables. For example, if the market price of the loan is significantly above par, the negative effect on refinancing predicted in the full model would be greater.

Table 5 ■ MNL competing risk models.

	Full Competing Risks Model			Combined Model	
	Refinance	Move	Default	All Prepays	Default
Loan characteristics					
Market price of loan	27.194	−6.686	10.114	13.275	10.131
	(2.738)	(3.536)	(8.869)	(2.091)	(8.868)
Original loan balance	0.293	−0.095	−0.081	0.122	−0.081
($00,000)	(0.056)	(0.073)	(0.282)	(0.045)	(0.282)
15-year loan indicator	−0.009	−0.289	−0.981	−0.145	−0.981
	(0.160)	(0.175)	(1.147)	(0.119)	(1.147)
Points (estimated)	−0.010	−0.160	0.079	−0.075	0.079
	(0.044)	(0.062)	(0.133)	(0.035)	(0.133)
Original refinance	−0.547	0.207	1.130	−0.242	1.129
indicator	(0.168)	(0.203)	(0.664)	(0.126)	(0.664)
Housing market and econ. conditions					
Current loan-to-value	−0.014	0.005	0.034	−0.006	0.034
	(0.004)	(0.005)	(0.017)	(0.003)	(0.017)
Prob. neg. eq. >90%	0.209	−0.855	−1.462	−0.185	−1.464
pctile indicator	(0.227)	(0.347)	(1.309)	(0.185)	(1.308)
House price appr. ind.	0.001	0.003	0.006	0.002	0.006
(Age <40, $,000)	(0.002)	(0.003)	(0.010)	(0.002)	(0.009)
House price appr. ind.	−0.003	−0.006	−0.006	−0.003	−0.006
(Age >40, $,000)	(0.001)	(0.002)	(0.007)	(0.001)	(0.007)
Unemployment rate	−0.100	−0.053	0.275	−0.068	0.275
	(0.036)	(0.041)	(0.182)	(0.027)	(0.181)
Borrower characteristics					
Borrower age	−0.003	−0.014	−0.019	−0.008	−0.019
	(0.006)	(0.006)	(0.026)	(0.004)	(0.026)
Minority indicator	−0.700	−1.074	−0.309	−0.847	−0.310
	(0.169)	(0.210)	(0.490)	(0.127)	(0.490)
Borrower income	−0.017	0.019	−0.036	0.001	−0.036
	(0.008)	(0.005)	(0.071)	(0.005)	(0.071)
Obligation ratio	0.001	0.007	0.045	0.004	0.045
	(0.006)	(0.007)	(0.029)	(0.005)	(0.029)
High credit score	−0.153	0.247	0.961	0.026	0.960
indicator	(0.137)	(0.161)	(1.122)	(0.104)	(1.122)
Low credit score	−0.541	−0.143	3.658	−0.412	3.658
indicator	(0.269)	(0.342)	(1.142)	(0.208)	(1.142)
Constant	−31.229	0.361	−44.286	−17.554	−44.304
	(2.897)	(3.489)	—	(2.159)	—
Number of observations	38301			38301	
Log likelihood	−3181			−2886	
χ^2 (*d.f.*)	45417 (72)			45276 (48)	
Prob > χ^2	0.0000			0.0000	
Pseudo R^2	0.1144			0.0976	

Note: The table compares the results of the full multinomial logit competing risk model (from Table 4) with the multinomial logit results when terminations from refinance and move are combined (Prepays). (Coefficient above; standard errors in parentheses).

(Table 5). The practical significance of the change in coefficient size is important. For example, a decline in interest rates of 200 basis points would increase the quarterly probability of refinancing (for the average loan with a current market price slightly below par) by a factor of nine from 0.4% to 3.6%.[44] In the combined prepayment model, the same change in rates would result in an increase in the probability of prepaying from 1.3% to 3.5%. The predicted aggregate annualized termination rate for the full competing risks model is 15.2% compared with 13.3% for the combined model. This 1.9% difference in prepayment speed is significant for pricing mortgage-backed securities and servicing rights.

The value of the competing risks model is also shown by the effect of credit scores on aggregate termination rates. In the full competing risks model, shifting the low credit score indicator from zero to one (while holding all other variables at their means) is predicted to increase the annual termination rate by 0.4%: The probability of default increases more than the probability of refinancing declines and the effect of the shift on the probability of moving is small. This is reasonable for many borrowers facing financial distress: Moving is a viable alternative to default. In the combined model, the effect of shifting to low credit scores is a 1.6% *reduction* in the annual termination rate: The negative prepayment effect overwhelms the positive default effect. This appears to be an artifact associated with combining the move probability with the refinance probability.

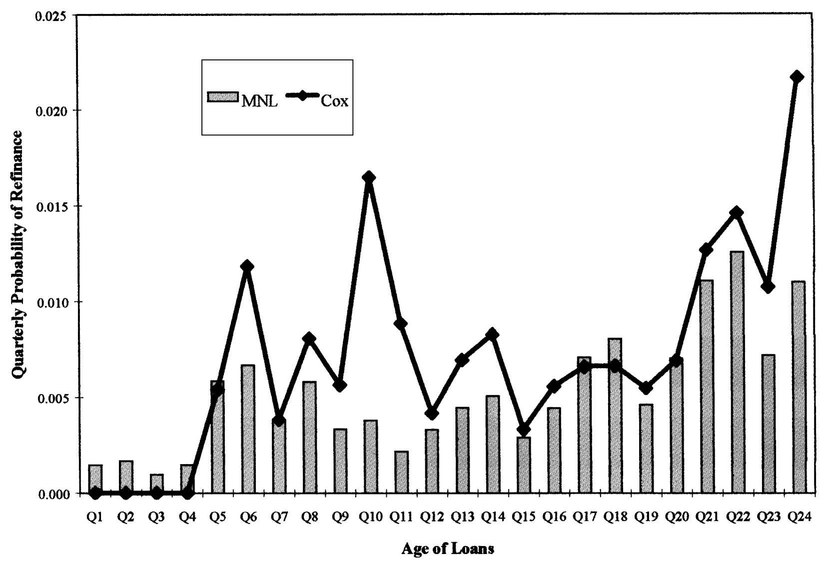
The negative coefficient for points in the move model is still captured as a negative effect on prepayment in the combined model (See Table 5), but the magnitude and precision of the estimate are reduced and the effect is not statistically significant. The estimated effect of current loan-to-value in limiting refinancing is also sharply reduced in the combined model. Similar changes can be seen in the coefficients of unemployment and the indicator of an original refinance loan. The combined model shows borrower income to have no significant effect on the prepayment rate. This provides evidence that models estimated using combined data are likely to be poor predictors of future mobility-driven prepayment.

Baseline Hazards for Refinancing and Moving

Figures 2 and 3 compare the baseline hazards for refinancing and moving resulting from the two different models: multinomial logit and Cox proportional

[44] There would be a corresponding changes in the probability of continuing to pay, moving and defaulting from 98.85%% to 95.91%%, 0.67% to 0.39% and 0.01% to 0.03%, respectively.

Figure 2 ■ Comparison of baseline refinance hazards. The figure compares the estimated quarterly probability of refinancing over time from the multinomial logit model (MNL) and the Cox proportional hazard model (Cox). The baseline shows how the probability of refinancing changes as a typical loan ages using the mean values, over the entire sample, for all explanatory variables.

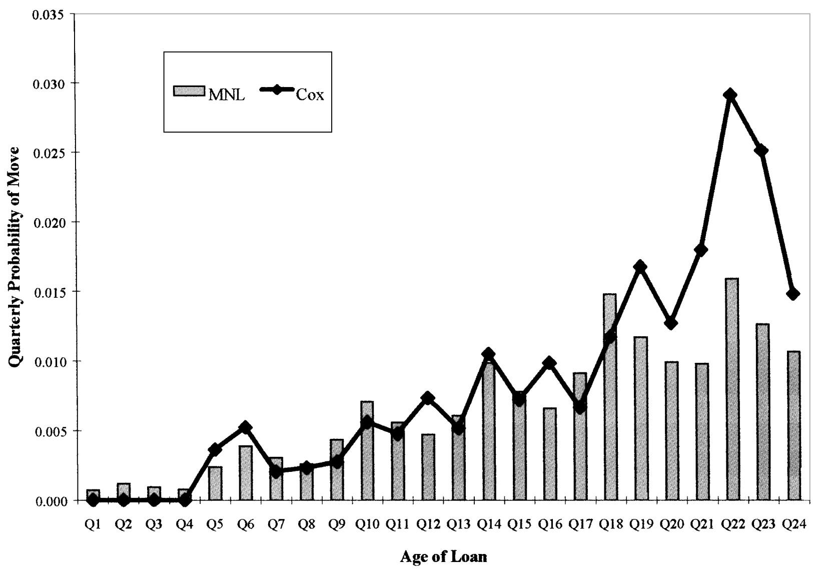


hazard. In each figure, the baseline from the multinomial logit model is shown as a series of bars while the Cox baseline is shown as a line.[45]

Both figures show general agreement between the two baselines. The steady upward trend in Figure 3 is consistent with the hypothesis that housing dissatisfaction increases over time leading to a higher probability of moving. The sharp increase in the Cox refinancing hazard (and corresponding decline in the Cox move hazard) for quarters 23 and 24 is attributable to the underestimation of the moves in those quarters due to the lagged reporting of house transactions. The upward spike in the Cox refinancing hazard in quarter 11 is attributable to an anomalous increase in refinancings that quarter. The multinomial logit specification is less sensitive to large quarterly fluctuations because it is estimated with year and season indicators, not quarterly indicators.

[45] The multinomial logit baseline is calculated by setting the value of all variables except the year and season indicators at their sample means. Given the previously mentioned transformation of the data for the Cox model, this is equivalent to the standard Cox baseline that shows the hazard over time when all explanatory variables are set to zero.

Figure 3 ■ Comparison of baseline move hazards. The figure compares the estimated quarterly probability of moving over time from the multinomial logit model (MNL) and the Cox proportional hazard model (Cox). The baseline shows how the probability of moving changes as a typical loan ages using the mean values, over the entire sample, for all explanatory variables.



Tests of Proportionality and IIA Assumptions

The Cox proportional hazard model assumes that the effects of explanatory variables are the same at all times in the sample. We test the proportionality assumption for each risk using a test proposed by Grambsch and Therneau (1994). The test rejects the assumption of proportionality of all variables considered jointly in the move and refinance equations (using a 10% significance level), but does not reject the assumption for the default risk. We further examine the violations of nonproportionality by interacting the log of mortgage age with the market price of the loan, the current loan-to-value ratio and the minority status indicator.[46] We conclude that the violations of proportionality are not material to the primary results of the paper and report the more parsimonious models in Table 4.[47]

46 The null hypothesis of proportionality is rejected for these three variables in both the move and refinance models without interactions.

47 Only the coefficient on the interaction of minority status and mortgage age is significantly different from zero.

Table 6 ■ Results from multinomial refinance models using alternative mortgage price proxies.

	CDH Closed Form	Noncallable Price	Ratio of Rates
Loan characteristics			
Market price of loan	27.194***	11.544***	5.601***
	(2.738)	(1.230)	(0.663)
Original loan balance ($00,000)	0.293***	0.288***	0.292***
	(0.056)	(0.056)	(0.057)
15-year loan indicator	−0.009	0.096	−0.093
	(0.160)	(0.164)	(0.157)
Points (estimated)	−0.010	−0.025	−0.217***
	(0.044)	(0.045)	(0.036)
Original refinance indicator	−0.547***	−0.577***	−0.542***
	(0.168)	(0.169)	(0.169)
Housing market and econ. conditions			
Currentloan-to-value	−0.014***	−0.015***	−0.013***
	(0.004)	(0.004)	(0.004)
Prob. neg. eq. >90% pctile indicator	0.209	0.259	0.191
	(0.227)	(0.228)	(0.227)
House price appr. ind. (Age <40) ($,000)	0.001	0.001	0.001
	(0.002)	(0.002)	(0.002)
House price appr. ind. (Age >40) ($,000)	−0.003*	−0.002	−0.002
	(0.001)	(0.001)	(0.001)
Unemployment rate	−0.100***	−0.091**	−0.091**
	(0.036)	(0.037)	(0.037)

Tests of the IIA assumption for the multinomial logit model are inconclusive. The Hausman and McFadden (1984) test fails to reject the IIA assumption while the Small and Hsiao (1985) test rejects the assumption.[48]

These test results suggest that the data do not fully conform to the underlying assumptions of either model. Nevertheless, we believe that the robustness of the results across both models supports our primary conclusions (see Table 3).

Comparing the CDH Estimate of $M(H, H_d, H_s^i, r, t)$ with Alternatives

We re-estimate all the models reported in Table 4 using the two alternative estimates of $M(H, H_d, H_s^i, r, t)$ discussed previously. Table 6 reports the results for the multinomial refinance model. (Full details are available from the authors.)

[48] Future research will explore the potential nesting of the borrower's decisions.

Table 6 ■ continued

	CDH Closed Form	Non-callable Price	Ratio of Rates
Borrower characteristics			
Borrower age	−0.003	−0.003	−0.003
	(0.006)	(0.006)	(0.006)
Minority indicator	−0.700***	−0.692***	−0.736***
	(0.169)	(0.168)	(0.171)
Borrower income	−0.017**	−0.016**	−0.014*
	(0.008)	(0.008)	(0.008)
Obligation ratio	0.001	0.001	0.005
	(0.006)	(0.006)	(0.006)
High credit score indicator	−0.153	−0.175	−0.045
	(0.137)	(0.138)	(0.139)
Low credit score indicator	−0.541**	−0.597**	−0.509*
	(0.269)	(0.277)	(0.269)
Constant	−31.229***	−15.774***	−9.484***
	(2.897)	(1.489)	(0.945)
Number of observations	38301	38301	38301
Log likelihood	−3181	−3185	−3200
χ^2 (*d.f.*)	45417 (72)	33432 (72)	28575 (72)
Prob. > χ^2	0.0000	0.0000	0.0000
Pseudo R^2	0.1144	0.1133	0.1092

Coefficient above; standard errors in parentheses.
* = significant at the 10% level;
** = significant at the 5% level;
*** = significant at the 1% level.

The models using the CDH estimates have better goodness of fit measures (log-likelihood, χ^2 test statistic and Pseudo R^2).

The coefficient on the CDH market price is more than twice as large as that for the noncallable price because the duration of the CDH price is approximately four years while that of a noncallable mortgage is approximately nine years. For loans with market prices near par, a 1% change in interest rates leads to twice as large a change in the noncallable price as it does the CDH market price because the latter includes the call option.

Overall, the predictions of termination probabilities are similar across all three models. The advantage of using CDH is in the premium price range. The duration of the noncallable premium mortgage remains close to nine years even when the call option is deep in the money. Consequently, a model based on the noncallable value will predict approximately the same proportional change in probability of refinancing for a 100 basis point change in rates regardless of whether the mortgage price is near par or at a significant premium. The CDH duration declines for premium loans and thus a decline in rates when the mortgage is already trading well above par is predicted to have a smaller effect on refinancing probability than it would if the mortgage were trading near par.

Conclusions

Most previous research on residential mortgage prepayment has failed to distinguish between mobility and refinancing risks. Since different (but overlapping) factors drive these two decisions, models that combine them produce inaccurate predictions unless all variables are at their means. We model the move decision as competing with refinance and default decisions.

The values of the refinancing and default options together with a measure of housing dissatisfaction provide a theory of the levels of state variables that separate these competing risks. While many variables influence the move and refinance decisions in the same direction, we identify certain variables that we expect to have different effects on different risks. All three decisions are mediated by transaction costs (a function of borrower characteristics) and institutional constraints (a function of housing market conditions and borrower characteristics). Table 1 compares the theoretical predictions with our empirical results and shows substantial agreement between the two.[49]

The most important drivers of the three decisions are:

- Refinance: the market price of the loan, the original loan balance and the current loan-to-value ratio.
- Move: borrower characteristics (age, minority status and income) and the probability of negative equity.
- Default: credit score and current loan-to-value ratio.

In addition, we find that minority status and previous refinancing strongly reduce refinancing rates. Paying more points signals that borrowers are less likely to move.

We demonstrate the importance of separating movers from refinancers by comparing a prepayment model with movers and refinancers combined to the full competing risks model; Wald and likelihood ratio tests show that the two

[49] The results indicate a weak negative effect of the market value of the loan on moving, contrary to theory. Data exploration shows that this effect is due to a group of borrowers who moved very quickly in 1995 when interest rates increased.

models are significantly different. The significance of variables that primarily influence prepayment through one choice (moving or refinancing) is lost in the combined model because either the absolute value of the coefficient is shifted toward zero or the precision of the estimate is greatly reduced. Furthermore, the estimate of borrower sensitivity to interest rate movements is substantially diminished.

The authors thank David Ling and Richard Green for helpful comments on an earlier version of this paper. The authors also thank Matt Spiegel for sharing his GAUSS code (used as a general guide for the local housing market analysis) and the Research Foundation of the University of Connecticut for grants used to purchase the transactions data. The authors thank three anonymous referees for their helpful comments. John Clapp gratefully acknowledges financial assistance from the Center for Real Estate, University of Connecticut.

References

Archer, W., D. Ling and G. McGill. 1996. The Effect of Income and Collateral Constraints on Residential Mortgage Terminations. *Regional Science and Urban Economics* 26: 235–261.

Archer, W., D. Ling and G. McGill. 1997. Demographic Versus Option-Driven Mortgage Terminations. *Journal of Housing Economics* 6(2): 137–163.

Avery, R., R. Bostic, P. Calem and G. Canner. 1996. Credit Risk, Credit Scoring, and the Performance of Home Mortgages. *Federal Reserve Bulletin* 82(July): 621–664.

Bergström, R. and P.A. Edin. 1992. Time Aggregation And The Distributional Shape Of Unemployment Duration. *Journal of Applied Econometrics* 7: 5–30.

Boehm, T.P. 1981. Tenure Choice and Expected Mobility: A Synthesis. *Journal of Urban Economics* 10: 375–390.

Boehm, T.P. and K.R. Ihlanfeldt. 1986. Residential Mobility and Neighborhood Quality. *Journal of Regional Science* 26(2): 411–423.

Brueckner, J.K. 1992. Borrower Mobility, Self-Selection, and the Relative Prices of Fixed-and Adjustable-Rate Mortgages. *Journal of Financial Intermediation* 2: 401–421.

Brueckner, J.K. 1994. Borrower Mobility, Adverse Selection, and Mortgage Points. *Journal of Financial Intermediation* 3(4): 416–441.

Caplin, A., C. Freeman and J. Tracy. 1997. Collateral Damage: Refinancing Constraints and Regional Recessions. *Journal of Money, Credit and Banking* 29(4): 497–516.

Clapp, J.M. 2000. Method for Estimating Land Value Surfaces Over Time. Draft paper, University of Connecticut.

Clapp, J.M., J. Harding and M. LaCour-Little. 2000. Expected Mobility: Part of the Prepayment Puzzle. *The Journal of Fixed Income* 10(1): 68–78.

Collin-Dufresne, P. and J. Harding. 1999. A Closed-Form Formula for Valuing Mortgages. *Journal of Real Estate Finance and Economics* 19(2): 133–146.

Cox, D.R. 1972. Regression Models and Life Tables (with discussion). *Journal of the Royal Statistical Society* 34: 187–220.

Cox, J.C., J.E. Ingersoll, and S.A. Ross. 1985. An Intertemporal General Equilibrium Model of Asset Prices. *Econometrica* 53(2): 363–384.

Curley, A.J. and J.M. Guttentag. 1974. The Yield on Insured Residential Mortgages. *Explorations in Economic Research* 1: 114–161.

Au: give location of authors

Deng, Y., J.M. Quigley and R. Van Order. 1996. 'Wood-heads' and the Exercise of Mortgage Options. Draft manuscript.

Deng, Y., J.M. Quigley and R. Van Order. 2000. Mortgage Terminations, Heterogeneity and the Exercise of Mortgage Options. *Econometrica* 68(2): 275–307.

Dunn, K.B. and C.S. Spatt. 1988. Private Information and Incentives: Implications for Mortgage Contract Terms and Pricing. *Journal of Real Estate Finance and Economics* 1(1): 47–60.

Elder, H. and P.M. Rudolph. 2000. Mobility and Housing Tenure Transitions of Older Americans. Draft Paper, University of Alabama.

Grambsch, P.M. and T.M. Therneau. 1994. Proportional Hazards Tests and Diagnostics Based on Weighted Residuals. *Biometrika* 81: 515–526.

Green, J. and J.B. Shoven. 1986. The Effects of Interest Rates on Mortgage Prepayments. *Journal of Money, Credit, and Banking* 18(1): 41–50.

Green, R.K. and M. LaCour-Little. 1999. Some Truths About Ostriches: Who Never Refinances Their Mortgage and Why They Don't. *Journal of Housing Economics* 8: 233–248.

Han, A. and J.A. Hausman. 1990. Flexible Parametric Estimation of Duration and Competing Risk Models. *Journal of Applied Econometrics* 5: 1–28.

Hanushek, E.A. and J.M. Quigley. 1978. An Explicit Model of Intra-Metropolitan Mobility. *Land Economics* 54: 411–429.

Harding, J. 1994. Rational Mortgage Valuation Using Optimal Intertemporal Refinancing Strategies and Homogeneous Borrowers. Ph.D. dissertation. University of California: Berkeley, CA.

Hausman, J. and D. McFadden. 1984. Specification Tests for the Multinomial Logit Model. *Econometrica* 52(7): 1219–1240.

Hayre, L.S., S. Chaudhary and R.A. Young. 2000. Anatomy of Prepayments.*The Journal of Fixed Income* 10(1): 19–49.

Hendershott, P. and R. Van Order. 1987. Pricing Mortgages: An Interpretation of Models and Results. *Journal of Financial Services Research* 1: 77–111.

Hilliard, J.E., J.B. Kau and V. C. Slawson, Jr. 1998. Valuing Prepayment and Default in a Fixed Rate Mortgage. *Real Estate Economics* 26(2): 431–468.

Au: give page range

Jenkins, S.P. 1995. Practitioner's Corner: Easy Estimation Methods for Discrete-Time Duration Models. *Oxford Bulletin of Economics and Statistics* 57(1).

Kalbfleisch, J.D. and R.L. Prentice. 1980. *The Statistical Analysis of Failure Time Data*. John Wiley & Sons: New York, NY.

Kelly, A. 1995. Racial and Ethnic Disparities in Mortgage Prepayment. *Journal of Housing Economics* 4: 350–372.

Kiel, K.A. 1994. The Impact of House Price Appreciation on Household Mobility. *Journal of Housing Economics* 3: 92–108.

LaCour-Little, M. 1999. Another Look at the Role of Borrower Characteristics in Predicting Mortgage Prepayments. *Journal of Housing Research* 10(1): 45–60.

Long, J.S. 1997. *Regression Models for Categorical and Limited Dependent Variables*. Sage Publications: Thousand Oaks, CA.

McCall, B.P. 1996. Unemployment Insurance Rules, Joblessness, and Part-time Work. *Econometrica* 64(3): 647–682.

Myers, D., S.S. Choi and S.W. Lee. 1997. Constraints of Housing Age and Migration on Residential Mobility. *Professional Geographer* 49(1): 14–28.

Narendranathan, W. and M.B. Stewart. 1993. Modelling the Probability of Leaving Unemployment: Competing Risks Models with Flexible Base-line Hazards. *Applied Statistics* 42(1): 63–83.

Patruno. G.N. 1994. Mortgage Prepayments: A New Model for a New Era. *The Journal of Fixed Income* 4(3): 42–56.

Pavlov, A.D. 2000a. Space-Varying Regression Coefficients: A Semi-parametric Approach Applied to Real Estate Markets. *Real Estate Economics* 28(2): 249–283.

Pavlov, A.D. 2000b. Competing Risks of Mortgage Prepayments: Who Refinances Who Moves, and Who Defaults? *Journal of Real Estate Finance and Economics* (forthcoming). **Au: any update ?**

Peristiani, S., P. Bennett, G. Monsen, R. Peach and J. Raiff. 1997. Credit, Equity, and Mortgage Refinancings. *Federal Reserve Bank of New York Policy Review* (July): 83–99.

Quigley, J.M. 1987. Interest Rate Variations, Mortgage Prepayments, and Household Mobility. *The Review of Economics and Statistics* 69(4): 636–643.

Quigley, J.M. and D.H. Weinberg. 1977. Intra-Urban Residential Mobility: A Review and Synthesis. *International Regional Science Review* 2(1): 42–66.

Richard, S.F. and R. Roll. 1989. Modeling Prepayments on Fixed Rate Mortgage-Backed Securities. *Journal of Portfolio Management* 15(3): 73–82.

Ross, S.L. 1998. Racial Differences in Residential and Job Mobility: Evidence Concerning the Spatial Mismatch Hypothesis. *Journal of Urban Economics* 43: 112–135.

Rossi, P.H. 1955. *Why Families Move*. The Free Press. **Au: give publisher location**

Sa-Aadu, J. and C.F. Sirmans. 1995. Differentiated Contracts, Heterogeneous Borrowers, and the Mortgage Choice Decision. *Journal of Money, Credit & Banking* 27(2): 498–510.

Small, K.A. and C. Hsiao. 1985. Multinomial Logit Specification Tests. *International Economic Review* 26(3): 619–627.

South, S.J. and K.D. Crowder. 1998. Leaving the'Hood: Residential Mobility Between Black, White, and Integrated Neighborhoods. *American Sociological Review* 63(February): 17–26.

Speare, A. Jr. 1974. Residential Satisfaction as an Intervening Variable in Residential Mobility. *Demography* 11. **Au: give page no.**

Stanton, R. 1994. Rational Prepayment and Valuation of Mortgage-Backed Securities. *Review of Financial Studies* 8(3): 677–708.

Stanton, R. and N. Wallace. 1998. Mortgage Choice: What's the Point? *Real Estate Economics* 26(2): 173–206.

Sueyoshi, G.T. 1992. Semiparametric Proportional Hazards Estimation of Competing Risks Models with Time-varying Covariates. *Journal of Econometrics* 51: 25–58.

Vandell, K. 1993. Handing Over the Keys: A Perspective on Mortgage Default Research. *AREUEA Journal* 21(3): 211–246.

Vandell, K. 1995. How Ruthless is Mortgage Default? A Review and Synthesis of the Evidence. *Journal of Housing Research* 6(2): 245–264.

Wand, M.P., and M.C. Jones. 1995. *Kernel Smoothing*. Chapman and Hall: New York, NY.

Yinger, J. 1997. Cash in Your Face: The Cost of Racial and Ethnic Discrimination in Housing. *Journal of Urban Economics* 42: 339–365.

P1: GTT
REEC.cls BL013-04 June 19, 2001 11:7

Appendix: Methods Used for the Locally Weighted Regressions

The local regression model (LRM) has two parts: OLS for the linear relationship between log of sales price and property characteristics (*e.g.*, square footage and building age) and local polynomial regressions (LPR) for the nonlinear effects of space and time. The OLS part of the LRM model is from standard hedonic regression analysis. This Appendix presents technical aspects of the locally-weighted part of the LRM. More detail on both parts of the LRM is contained in Clapp (2000).

Our local regression model (LRM) can be outlined as follows:

1. Use a standard hedonic regression to control for square footage, bathrooms, age and other characteristics of properties in the transactions database. The residuals from this regression contain information about variation in house value over space and time.
2. Place a three-dimensional grid over these residuals as a function of latitude, longitude and time. The values of the explanatory variables for each observation are subtracted from each target point.
3. For a given grid point, weight the residuals and the transformed (in step 2) explanatory variables inversely with distance in each of the three dimensions. The three weights are multiplied together.
4. Run a local polynomial (smoothing) regression to find the value of a standard house at each point in space and time (each grid point).
5. Iterate the OLS and LPR parts of the model to obtain independence.
6. Each loan is at a given point in space: Use the results from step 5 to interpolate the price index that applies to the location (latitude and longitude) for the loan.
7. Multiply cumulative percent changes in the price index by the collateral value of the house in order to estimate current LTV and growth/decline in housing wealth.[50]
8. Obtain the probability of negative equity over time for each loan from $\hat{P}_t$ (the estimated house value), the standard deviation of $\hat{P}_t$, the remaining balance on the loan and the normal pdf.

[50] House value at origination is the minimum of appraised value or sales price; for a refinance, it is appraised value.

Technical Details

The "kernel" is the weighting function for the LPR. It is any smooth, unimodal, symmetric function. We have chosen a Gaussian kernel:[51]

$$f(X;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)}. \tag{A1}$$

We experimented with alternative choices given by the following general form:

$$K(x;h) = (h)^{-1}\sum_{i=1}^{n} K\,\{(X_i - x)/h\} = K_h(X_i - x), \tag{A2}$$

where n is sample size, H is the bandwidth, and

$$\lim_{n\to\infty} h = 0 \qquad \text{and} \qquad \lim_{n\to\infty} nh = \infty.$$

Here, h is strictly analogous to the standard deviation in Equation (A1). The larger the bandwidth, the more observations are given high weight in the regression. The results are not sensitive to choices of a functional form for the kernel.

The local polynomial regression equation is given by:

$$Y_i = \beta_0 + \beta_1(X_i - x) + \beta_2(X_i - x)^2 + \cdots + \beta_p(X_i - x)^p. \tag{A3}$$

Note that the regression is a polynomial in the explanatory variables; generally $p = 1$, but 2 or 3 are possible. The polynomial allows for curvature in the functional form. With LPR, interest centers on the constant term: This is the $\hat{Y}$ estimate when covariates are at the target points.

The kernel is applied to each observation:

$$\mathrm{Min}(\hat{\beta}) \sum_{i=1}^{n} \{Y_i - \beta_0 - \cdots - \beta_p(X_i - x)^p\}^2 K_h(X_i - x). \tag{A4}$$

The $\hat{\beta}$s are given by equation (A5) with the terms defined by Equations (A6) and (A7). Note the GLS form of Equation (A5).

$$\hat{\beta} = \left(X_x^T W_x X_x\right)^{-1} X_x^T W_x Y, \tag{A5}$$

[51] For ease of exposition, all equations are given for a single covariate. The product kernel described above allows obvious generalization to three dimensions.

where

$$X_x = \begin{bmatrix} 1 & X_1 - x & \cdots & (X_1 - x)^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_n - x & \cdots & (X_n - x)^p \end{bmatrix} \tag{A6}$$

is an $n \times (p + 1)$ design matrix and the weights are given by an $n \times n$ matrix:

$$W_x = diag\{K_h(X_1 - x), \ldots, K_h(X_n - x)\}. \tag{A7}$$

Bandwidth Selection

Empirical results are sensitive to the choice of bandwidth. A very small bandwidth would mean few observations in each local regression: The results would be highly variable, increasing the mean squared error. Large (above optimal) bandwidths will cause the estimated value surface to be biased. That is, the surface will not be sensitive to underline variations in value over space and time. The choice of optimal bandwidth in three dimensions is complicated because all three dimensions must be considered simultaneously.

The literature suggests several methods for optimal bandwidth selection (Wand and Jones 1995). This paper chooses bandwidths (one for each of the three dimensions) to minimize out-of-sample mean squared error. In addition, the bandwidths used here are locally adaptive. If a given local regression has fewer than m observations, then the bandwidth is increased until at least m observations are included. The results reported here use $m = 30$.

An interesting feature of LPR is that OLS is a special case. When the bandwidth becomes very large, each local regression gives a high weight to every observation in the sample. The weights used in Equation (A4) become essentially equal for all observations, and (A4) reduces to the ordinary least squares estimator.