Omitted Mobility Characteristics and Property Market Dynamics:

Application to Mortgage Termination

XUDONG AN, JOHN M. CLAPP, and YONGHENG DENG*

First draft: July 1, 2004

Second draft: August 25, 2005

Current draft: May 23, 2006

* The authors are grateful for helpful comments on an earlier draft by Professor Cheng Hsiao at the University of Southern California. Yongheng Deng and Xudong An thank the financial support from the Lusk Center for Real Estate at the University of Southern California.

XUDONG AN *is a Ph.D. candidate in the School of Policy, Planning and Development at the University of Southern California (E-mail:*xudongan@usc.edu*).* JOHN M. CLAPP *is a Professor in the School of Business Administration at the University of Connecticut (E-mail:* john.clapp@business.uconn.edu*).* YONGHENG DENG *is an Associate Professor in the School of Policy, Planning and Development at the University of Southern California (E-mail:* ydeng@usc.edu*).*

Omitted Mobility Characteristics and Property Market Dynamics: Application to Mortgage Termination

XUDONG AN, JOHN M. CLAPP and YONGHENG DENG

ABSTRACT

Property market dynamics depend on changes in long run equilibrium and on impediments to adjustment towards equilibrium. One of the most important aspects of property market dynamics is often attributed to the activities in the mortgage market. However, many impediment factors, such as changes in family status, education, neighborhood effects and job relocation, are often unobservable from micro household-level data. Since these omitted variables contribute to moving decisions and therefore to sale and default decisions, utility functions for sale and default are correlated through these unobservable variables; thus, the IIA assumption of the widely used Multinomial Logit Model (MNL) is violated. Under such circumstances, econometric theory suggests that the Nested Logit Model (NMNL) is a better choice, which obviates the limitation of MNL by allowing correlation in unobserved factors across alternatives.

This paper empirically investigates the omitted household mobility characteristics problem in mortgage termination, and tests NMNL against MNL. Using loan level micro data, we find significant correlation between sale and default due to omitted borrower mobility characteristics. Our simulations find that NMNL out performs MNL in out-of-sample prediction.

JEL codes: G21; C25; C41; C52; D12

Keywords: Mortgage termination; Multinomial Logit Model (MNL); Nested Logit Model (NMNL); Mobility; Refinance; Default.

1. INTRODUCTION

Economists have reached a general consensus that property markets adjust slowly to supply and demand shocks, so that any study of dynamics must deal not only with long run equilibrium but also with impediments to adjustment towards equilibrium.[1] Property market dynamics must be understood through theories of local market equilibrium as mediated by adjustment mechanisms such as the mobility decision.[2] The adjustment path for any given local market is influenced, at a minimum, by transactions costs, information costs and inertial forces such as attachment to a neighborhood (Whitehead and Odling-Schmee, 1975, p. 316).

The complexity of these adjustment mechanisms poses a challenge to empirical research on property markets, especially at the microeconomic level where household and firm heterogeneity implies different responses to shocks. For example, information sets will vary across economic agents, implying variation in rates of adjustment to spatial disequilibrium. Data allowing adequate discrimination across groups of agents is difficult or impossible to obtain. These data limitations have caused researchers to use econometric techniques such as random coefficients or models of unobserved heterogeneity.

An important perspective on property market dynamics may be attributed directly to mortgage market activities and to the financial constraints imposed by mortgage markets. Decisions of firms and households to terminate mortgage contracts due to financial or relocation motivations are not the only adjustment mechanism, but they constitute a category of particular relevance to spatial disequilibrium. This paper focuses on the property market dynamics from the

[1] Whitehead and Odling-Smee (1975) emphasize the need to evaluate adjustment dynamics. Ball and Kirwan (1977) point out that "the persistence of sub-markets will reflect a continuing mismatch between the evolving pattern of demand and the slowly-changing supply (p. 15)." These points have been confirmed by many empirical studies: See Straszheim (1975), Schnare and Struyk (1977), Watkins (2001) and Clapp and Wang (2006).
[2] Maclennan Munro and Wood (1987, pp. 29-32) discuss Alfred Marshall's emphasis on long run equilibrium *versus* Adam Smith's emphasis on spatial rigidities.

mortgage market perspective. It models the decisions of households to move as reflected in mortgage market data, where a move as an adjustment mechanism may be associated with utility-maximizing decisions to either prepay or default on the mortgage. Often, the optimal choice between these two termination events depends on unobserved demographic changes associated with income, job location, or family size; substantial inertial forces include search costs, neighborhood change and attachment to an area. Likewise, the shocks that produce a new optimal bundle of housing attributes cause dissatisfaction with the current home; the new equilibrium is a function of many of the same variables. Thus, our focus on mobility and mortgage termination is a step towards a general understanding of the dynamic paths of property markets.

There have been several major developments in models of optimal mortgage termination. The option-theoretic model predicts that the borrower is exercising a call (prepayment) or put (default) option[3]. Empirical research has shown that option-theoretic models are able to explain a substantial proportion of prepayment and default risks[4]. However, due to inertial forces[5], researchers modify the standard frictionless-model. While many studies turn to transactions costs to explain the prepayment behavior[6], Archer, Ling and McGill (1996, 1997) establish the conceptual framework to distinguish refinance from move in a borrower's prepayment behavior. They argue that borrowers have three choices to terminate mortgages by either move, or

---

[3] See, Findley and Capozza (1977), Dunn and McConnell (1981), Buser and Hendershott (1984), Cunningham and Hendershott (1984), Brennan and Schwartz (1985), Epperson, Kau, Keenan and Muller III (1985), Kau, Keenan, Muller III and Epperson (1992) and Kau and Keenan (1995).

[4] See, Foster and Van Order (1984), Green and Shoven (1986), Schwartz and Torous (1989), Quigley and Van Order (1990, 1995) and Deng, Quigley and Van Order (2000).

[5] E.g. many borrowers prepay when call options are "out-of-money", while others do not default even when the put option is deeply "in-the-money".

[6] For example, Follain et al (1992) and Giliberto and Ling (1992) include borrower transaction costs in optimal call models, while Archer and Ling (1993) and Stanton (1995) further assume heterogeneous transaction costs for different borrowers.

refinance or default, and list several mobility variables affecting the move decision[7]. However, their estimates combine the three choices together.

Taking one step further, Clapp et al (2001) is the first to model the choices of refinance, move and default as three competing risks of mortgage termination. They provide evidence showing that determinants of moving are very different from those of refinance. More specifically, financial considerations (such as value of the call option) are primary drivers of the optimal refinance choice, while proxies for mobility characteristics of homeowners (e.g., age, income and minority indicators) are more important for the move decision. Taking the same approach, Pavlov (2001), Goldberg and Harding (2003) and Deng, Pavlov and Yang (2005) model mortgage termination as three different choices of refinance, move (sale) and default, and show that factors related to household mobility are crucial to the sale decision[8].

Turning to default, more and more researchers have found that option value is far from enough to explain borrower choices to default – many borrowers do not default although their houses have substantial negative equity[9]. Ambrose, Buttimer and Capone (1997) show the importance of transaction costs. Deng, Quigley and Van Order (1996, 2000) show that "trigger events" (i.e., shocks to an equilibrium) such as unemployment and divorce are important to the borrower's default decision. Vandell (1995) and Archer, Ling and McGill (1997) argue for similar trigger events or shocks, which are crucial to default decision. Pavlov (2001) and Deng, Pavlov and Yang (2005) argue that default is primarily driven by the optimality of a move in the presence of negative equity. Moderating variables such as years in the current home or proxies

[7] The list of variables includes lifecycle, education, employment opportunity, household restructuring, and deteriorating neighborhood, etc.
[8] Thereafter, we use "sale" to denote selling the property, paying off the mortgage and moving.
[9] For example, Cauley (1996) reports that there was little increase of default rate even though up to 44 percent of homes purchased between 1989 and 1991 in Los Angeles County had negative equity in 1995.

for transactions costs are considered by these studies. Therefore, there is increasing consensus that household mobility factors are also crucial for default.

Parallel to the conceptual understanding of mortgage termination choices, there have been important developments in empirical econometric models of mortgage termination. One strand of the literature applies the Cox Proportional Hazard (CPH) model;[10] at the same time, applications of logit models to mortgage termination are well established[11]. Logit models are straightforward and appear to do a good job of mimicking the mortgage borrowers' choice making. Recent applications of Multinomial Logit Models (MNL) to mortgage terminations structure the data according to event-history.[12] Clapp et al (2001) and Clapp, Deng and An (2004) argue that MNL with event history data is an attractive alternative to the PHM due to its inherent competing risks nature[13] and its ease of estimation. Many mortgage companies today are using MNLfor mortgage termination modeling.

Despite its wide application, there are certain limitations of MNL. A very important implicit assumption for MNL is the Independence of Irrelevant Alternatives (IIA) property: the predicted odds ratio of two choices will be constant if we eliminate a third choice from the

[10] The CPH model has been used in mortgage termination studies by Green and Shoven (1986), Schwatz and Torous (1989), Vandell, et al. (1993), Deng and Quigley and Von Order (1996), Deng (1997), Ambrose and Capone (2000), Pavlov (2001), Bennett Peach and Peristiani (2001) and Lambrecht, Perraudin and Satchell (2003). More recent studies include Deng and Quigley (2002), Huang and Ondrich (2002), Ciochetti et al (2002, 2003), Deng and Gabriel (2006), Deng, Pavlov and Yang (2004) and Clapp, Deng and An (2004).

[11] See Cunningham and Capone (1990), Philips et al. (1995), Quigley and Van Order (1995), Archer, et al. (1996), Ambrose and Capone (1998), Berkovec et al. (1998), Mattey and Wallace (2001), Clapp et al (2001), Calhoun and Deng (2002), Goldberg and Harding (2003) and Ambrose and Sanders (2003).

[12] Mortgage data allow tracing each loan from origination through termination or end of observation. The data are structured as a panel data, with one observation for each quarter for each loan during the observation period. The time series aspect is preserved when the records are stacked together, assuming the i.i.d. property.

[13] The MNL allows direct competition among the choices: the probabilities of termination risks, and the probability if continuing to pay, must sum to one. An increase in one termination probability must be offset by a decline in probability for one or more of the alternatives, thus risks are "competing".

model[14]. This is problematic when we apply MNL to the modeling of borrower refinance, sale and default choices: as we have discussed earlier, there is growing evidence showing that mobility plays important roles in borrowers' decisions on both sale and default. Borrowers rarely default just because the put option is in the money. It is highly likely that the borrower will choose default if there is negative equity while choose sale if there is positive equity, given that the borrower desires/needs to move. Now suppose that we eliminate the default choice. The ratio between predicted probabilities (expressed as odds) of refinance and sale will change after we eliminate the third choice, default. Econometric theory tells us that this happens because MNL assumes that the unobserved error terms of the three utility functions (refinance, sale and default) follow an i.i.d. extreme value distribution.[15] However, in reality, there are many unobservable household mobility characteristics related to both sale and default that cause correlation of error terms across alternative choices.

Given this problem, the appropriateness of using MNL to model mortgage termination by competing risks of refinance, sale and default is open to question. The Nested Logit (NMNL) model obviates the limitation of MNL by allowing correlation in unobserved factors across alternatives of sale and default. Therefore, NMNL is potentially a better choice when we model mortgage termination choices when researchers do not fully observe household mobility characteristics. The purpose of this paper is to follow the above rationale to investigate the omitted household mobility characteristics problem for mortgage termination, and empirically test whether there are substantial improvements when using NMNL rather than MNL.

[14] This is illuminated by the classical red bus-blue bus problem: the ratio between number of people drive to work and taking blue bus should keep unchanged if we eliminate the choice to take the red bus; but this is not the case in reality.
[15] See Train (2003) for a complete discussion of MNL and i.i.d. assumption.

The rest of the paper is organized as follows: the next section reviews the mortgage termination literature and further explains the omitted mobility characteristics problem in MNL modeling. Section three describes the difference between MNL and NMNL by intuition, and formally presents the two models. Section four and five present our data and empirical results regarding comparison between MNL and NMNL. Section six develops a simulation study in the spirit of MNL and NMNL comparison. Section seven draws conclusions.

2. MORTGAGE TERMINATIONS AND THE OMITTED MOBILITY CHARACTERISTICS PROBLEM

2.1 Mortgage Termination by Refinance, Sale and Default

As indicated in the previous section, we view a mortgage borrower as having four choices every month – to refinance the existing mortgage (pay off the current loan and replace it with another, called “refinance” hereafter); to sell the property, pay off the existing mortgage and move to somewhere else (called “sale”); to stop making payments and turn over possession of the property to the lender(called “default”) [16] and to make the scheduled payment (called “continue”) (see Archer, Ling and McGill 1996, 1997, Clapp et al 2001, Pavlov 2001, Goldberg and Harding 2003, and Deng, Pavlov and Yang 2004 for detailed discussions about each choice). It is important to separate prepayment into refinance and sale because the variables driving optimal choices are significantly different: refinance happens solely because the borrower wants to take advantages of the “in-the-money” call option (to obtain a lower interest rate and/or lower monthly payment); while sale is mainly motivated by the homeowner’s needs/desire to change the bundle of housing characteristics. Clapp et al (2001) identifies two different sets of

[16] Default is typically measured by three months of nonpayment. Note that some defaults are cured: they do not always results in foreclosure.

explanatory variables: refinance is mainly explained by financial considerations like market value of the loan, while homeowner characteristics (e.g., age, income and minority status) are more important for sale. Further, they show that for some explanatory variables common to refinance and sale, the coefficients are significantly different and some coefficients have opposite signs. They conclude that "combining these two distinct choices into a single prepayment shifts coefficients towards zero and produces inaccurate predictions of aggregate termination rates." (Clapp et al 2001, pp.411) Other empirical studies including Pavlov (2001), Goldberg and Harding (2003) and Deng, Pavlov and Yang (2005) find similar results. These studies account for inertial forces associated with the loan contract, the neighborhood and/or the borrower.

2.2 Variables Related to Refinance, Sale and Default

The option-theoretic approach implies that optimal refinance happens when the call option is "in-the-money". Empirical studies have developed several proxies for the call option, for example, computing the difference between the par value of the mortgage and the present value of the remaining payments evaluated using the current market mortgage rate (Deng, Quigley and Van Order 2000), the ratio of current interest rate to the rate at origination (Richard and Roll 1989, Pavlov 2001, and Calhoun and Deng 2002), and the estimated market value of the mortgage derived from a close form formula (Collin-Dufresne and Harding 1999, and Clapp et al 2001). They have substantial explanatory power for the refinance decision.

However, mortgage borrowers do not exercise the option to refinance as ruthlessly as do owners of other financial options[17] (See Green and LaCour-Little 1999, and Deng, Quigley and Van Order 2000). Stanton (1995) and Green and LaCour-Little (1999) developed models of

[17] Although, on the other hand, Hurst and Stafford (2004) show that some borrowers refinanced to convert equity into current consumption during 1991-1994, which caused some "unexpected" refinancing activities, the call option to refinance is generally under-exercised.

mortgage terminations that account for transaction costs. For example, the larger the loan balance, the greater the dollar amount of benefits from refinancing, which increases the probability of refinancing. Others incorporate various constraint effects on a borrower's ability to refinance. For example, Archer, Ling, and McGill (1996) find higher annual payment-to-income and loan-to-value ratios were negatively related to refinance. Bennett, et al (2001) found strong evidence that poor credit history as well as high current loan-to-value ratio (CLTV) significantly reduced the probability of refinance. Deng and Gabriel (2006) find that being a minority adds constraints to the borrower's refinance choice.

Sale shares some common factors with refinance: e.g., both value of the call option and transaction costs to refinance may influence mortgage borrower's decision to sell the house. However, mobility as an adjustment to a new equilibrium bundle of housing is found to be more fundamental in the sale decision[18]. Thus, factors that trigger a move, as well as those that hinder the move, should be considered in the model.

Archer, Ling, and McGill (1996) summarize mobility driven factors related to mortgage termination into two broad categories: the location decision factors and the response to housing disequilibrium factors. Employment opportunity is the most important location-driven mobility factor. People usually move because of job relocation. Pavlov (2001) finds that the local unemployment rate is positively related to move because there might be more attractive opportunities outside the local area. Besides employment, other factors like climate and health are also important location-driven mobility factors.

[18] Mobility is a mechanism whereby households adjust their optimal housing consumption to changes in circumstances (Rossi, 1955). See Maclennan, Munro and Wood (1987) for further elaboration on this point.

Housing disequilibrium is a more complicated issue[19]. The first important type of housing disequilibrium factor is household restructuring. Change of marital status and change of household size both have been found to increase the propensity to move (Krumm 1984, Boehm and Ihlandfeldt 1986 and Quigley 1987). The second important type is neighborhood effects: e.g., lower neighborhood quality may cause renovations to be less valuable, increasing the likelihood to move (Shear 1983), while time in the neighborhood may increase the psychic cost of relocating (Boehm 1981, and Ihlandfelt and Silberman 1985). Recent work by Deng, Pavlov and Yang (2004) also find important neighborhood effects on the sale decision.

The sale decision is affected by many variables related to the optimal consumption of housing and to inertial forces. The first category includes the transaction costs of moving, e.g., search time, relocation expenses, costs of sale, and acquisitions costs (Archer, Ling, and McGill 1996). The second category is the incentives for housing investment: e.g., Green and Shoven (1986) and Quigley (1987) document a significant "lock-in" effect arising from below market rate financing – homeowners with low mortgage rates (relative to current market rates) delayed moving. On the other hand, we can imagine that if the borrower has a high mortgage rate relative to current market rates, he has an added incentive for sale. The third category is the household life cycle: e.g., younger households have longer periods to amortize the cost of moving, so they should be more sensitive to any resulting benefit of relocating (Shear 1983, and Ihlandfelt and Silberman 1985). The last category is education and involvement in workforce. Education tends to increase the likelihood to move, while presence of a working spouse has the opposite effect (Krumm 1984). In addition, income and race are also important to the move decision (see, e.g., Quigley and Weinberg 1977).

[19] The disequilibrium 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, pp. 175).

Turning to default decisions, option theory predicts that negative equity is the most important variable determining the optimality of default. If there is negative equity in the house, the homeowner can exercise the put option by default to maximize his wealth. However, as discussed earlier, empirical evidence suggests that negative equity itself is far from enough to cause a default (Vandell 1995, Cauley 1996, Archer, Ling, and McGill 1996, Clapp et al 2001, Pavlov 2001, and Deng, Pavlov and Yang 2004). While many researchers see "trigger events" or "shocks" like unemployment, divorce and death as important factors to default (Vandell 1995, Archer, Ling, and McGill 1996 and Deng, Quigley and Van Order 2000), other studies have found that default is primarily driven by the needs/desire to move when there is negative equity (Pavlov 2001, Deng, Pavlov and Yang 2004). Therefore, the mobility factors examined above for the sale decision are also important factors for default.

2.3 The Omitted Mobility Characteristics Problem

While an increasing consensus has developed on the importance of mobility factors on both mortgage sale and default decisions, it is challenging to empirically model these mobility factors in mortgage termination. Empirical models need to deal with omitted variables that drive any change in equilibrium as well as those that accelerate or delay adjustments toward equilibrium.

As discussed in previous section, mobility is a very complicated issue that involves many dimensions and is often heterogeneous. Limited availability and accuracy of empirical data greatly restrain our ability to incorporating it in our modeling. Researchers have tried to use various variables to capture mobility effects: e.g., regional unemployment and divorce rate are used as sale and default explanatory variables and sometimes found significant (Deng, Quigley and Van Order 2000, Clapp et al 2001, Pavlov 2001, Deng, Quigley 2002); loan terms and points are used as indicators of move propensity (Clapp et al 2001); age, race (minority indicator), marital status and income are also used as move propensity indicators (Archer, Ling, and McGill 1996, Clapp et al 2001, Goldberg and Harding 2003, and Clapp, Deng and An 2004).

However, two problems remain: first, many of the variables used are just “proxies” that may contain substantial measurement error: e.g., age, income and marital status at origination are usually used in post-origination periods although they definitely can change over time. Second, many of the mobility characteristics are totally omitted from empirical data. For example, profession is an important factor predicting job change; however, it is often omitted. Neighborhood effects are mostly unobservable in loan level micro data. Education, family structure, income shocks, and transaction costs of the move are often omitted from the mortgage data. Empirical researchers studying mortgage termination by household mobility typically have a long list of important variables they wish to be observable and accountable in their models.

To summarize, there is increasing consensus that mobility characteristics are crucial factors for both mortgage sale and default decisions; however, most of these important mobility characteristics are often unobservable from the micro mortgage data, or poorly proxied. In order to accurately model mortgage termination decisions, econometric technique can be used to handle these deficiencies in the data..

3. MULTINOMIAL LOGIT MODEL (MNL) VERSUS NESTED LOGIT MODEL (NMNL)

3.1 IIA property of MNL and the advantage of NMNL

The Multinomial Logit Model (MNL) is one of the most widely used econometric tools in the housing and mortgage literatures[20].

McFadden (1974) proves that the logit model of choice probabilities implies that unobserved utility is distributed as an i.i.d. extreme value function. One of the key assumptions implied by the MNL is that for any two alternatives i and k, the odds ratio of the logit probabilities does not depend on any alternatives other than i and k. This is commonly known as the property of *independence from irrelevant alternatives*, or IIA, initially derived by Luce (1959). The IIA property allows researchers to estimate the parameters of the MNL model consistently using a subset of alternatives, because elimination of irrelevant alternatives does not affect the odds ratio of probability for the remaining choices. This nice feature of MNL allows researchers to reduce computing time significantly.

However, the IIA property may not always hold for choice probabilities faced by mortgage borrowers or mortgage pools. For example, suppose we have a total of 5 terminations with 3 refinances, 1 sale and 1 default in our sample. If we eliminate the default choice, the predicted number of refinances and sales will be 3.75 and 1.25 because, due to MNL's IIA property, the predicted ratio between refinance and sale should be unchanged at 3:1. In reality, in the absence of the default choice, borrowers previously choosing default due to their mobility needs have to choose sale. Therefore, the correct prediction should be 3 refinances and 2 sales rather than 3.75 refinances and 1.25 sales.

This problem can be explained more formally by the underlying assumptions of MNL. MNL is derived based on random utility theory. Suppose in each month (t), each individual (i) faces four choices: refinance, sale, default and continue, denoted by r, s ,d and c, and each choice is associated with a utility function, which can be decompose into deterministic parts and random

[20] In his Nobel Lecture delivered in Stockholm, Sweden on December 8, 2000, McFadden provides a historical review of the development of the micro-econometric analysis of behavior of consumers who face discrete choices (McFadden 2001).

error terms:

$$U_j = V_j + \varepsilon_j \tag{1}$$

where j=c, r, s, and d, and U_j is the utility associated with continue, refinance, sale or default (We suppress the subscript i and t for notation simplicity). V_j is the deterministic part, while ε_j is the error term.

The MNL assumes that ε_j is independently distributed not only across individuals and time, but also across choices, that is, for each individual i at time t, ε_r, ε_s and ε_d are not correlated. However, this is not the case for mortgage termination by refinance, sale and default. As discussed earlier, due to the data limitation, researchers often have to face the problem of omitted variables associated with sale and default. Therefore, V_j can be seen as a function of observed (X) variables, e.g., call option value, put option value, race, credit score, and omitted (Z) variables, e.g. changes of family size or status, profession, education, neighborhood effects, job relocation, income shocks, etc. This can be expressed by the following equation:

$$V_j = X\beta_j + Z\gamma_j\,, \tag{2}$$

Here, β_j and γ_j are parameters for observed and omitted variables that determine utility. In empirical estimation, the unobserved term, $Z\gamma_j$, becomes part of the error term ε_j. As we have discussed, many of the mobility characteristics in the vector Z are common factors for both sale and default. Therefore, $(\varepsilon_s, \varepsilon_d)$ becomes $(\varepsilon_s + Z\gamma_s, \varepsilon_d + Z\gamma_d)$ during the empirical estimation, and these error terms are correlated through Z. Given this correlation, the IIA property of MNL no longer holds (Amemiya 1985).[21]

[21] See McFadden (1974) for a complete discussion.

Under such circumstances, econometric theory suggests that the nested logit model (NMNL) is a better choice, which obviates the limitation of MNL by allowing correlation in unobserved factors across alternatives. Further, the correlation between ε_s and ε_d in NMNL can be estimated simultaneously with β_j . Therefore, even with the omitted mobility characteristics, coefficients (β_j) for observable variables X can be estimated consistently, and the model can provide accurate predictions of sale and default.

3.2 Functional forms of MNL and NMNL

For MNL, the choice set is $C = \{c, r, s, d\}$, and it is assumed that ε_j , *j=c, r, s,* and *d*, are independent and follow a Type I extreme value distribution. The cumulative distribution function of ε_j is:[22]

$$F(\varepsilon_j) = \exp[-\exp(-\varepsilon_j)] \tag{3}$$

The probability for the borrower to choose alternative j is:

$$\Pr ob(Y = j) = \frac{\exp(V_j)}{\sum_l \exp(V_l)}; l, j = c, r, s, d \tag{4}$$

For the NMNL, on the other hand, the choice set is $C = \{c, r, (s, d)\}$. Here sale and default are in one nested group. The disturbances for sale and default are assumed to be correlated and follow a Type II extreme value distribution:

$$F(\varepsilon_s, \varepsilon_d) = \exp\left\{-\left[\exp\left(-\frac{\varepsilon_s}{\rho}\right) + \exp\left(-\frac{\varepsilon_d}{\rho}\right)\right]^\rho\right\}, 0 < \rho \le 1 \tag{5}$$

Under this assumption, the correlation coefficient for ε_s and ε_d is $1-\rho^2$ (Amemiya, 1985).[23] The probabilities of choosing alternative groups are:

[22] The decision is made each time period. We suppress the time subscript throughout.

$$\Pr ob\left(Y=k\right)=\frac{\exp\left(V_k\right)}{\exp\left(V_c\right)+\exp\left(V_r\right)+\left[\exp\left(\frac{V_s}{\rho}\right)+\exp\left(\frac{V_d}{\rho}\right)\right]^{\rho}},k=c,r \qquad (6)$$

$$\Pr ob(Y=s\ or\ d)=\frac{\left[\exp\left(\frac{V_s}{\rho}\right)+\exp\left(\frac{V_d}{\rho}\right)\right]^{\rho}}{\exp\left(V_c\right)+\exp\left(V_r\right)+\left[\exp\left(\frac{V_s}{\rho}\right)+\exp\left(\frac{V_d}{\rho}\right)\right]^{\rho}} \qquad (7)$$

Conditional on not choosing continue or refinance, the probabilities of choosing sale or default are:

$$\Pr ob\left(Y=j\,|\,Y=s\quad or\quad d\right)=\frac{\exp\left(\frac{V_j}{\rho}\right)}{\exp\left(\frac{V_s}{\rho}\right)+\exp\left(\frac{V_d}{\rho}\right)},\quad j=s\quad or\quad d \qquad (8)$$

Therefore, the unconditional probability of choosing sale or default is:

$$\Pr ob\left(Y=j\right)=\Pr ob\left(Y=j\,|\,Y=s\ or\ d\right)\Pr ob\left(Y=s\ or\ d\right),\ j=s,\,d \qquad (9)$$

It is noteworthy that NMNL is a general form of the MNL. In NMNL, if $\rho=1$ the NMNL collapses into MNL. In such a case, ε_s and ε_d are uncorrelated (the correlation coefficient "$1-\rho^2$" equals 0), and the IIA assumption holds. Later on, we will perform Hausman-McFadden test and Small-Hsiao test to test the IIA assumption of MNL.

For either the MNL or NMNL, the log likelihood function for the competing risks model is:

$$Log\ L=\sum_j d_{ji}\log\left(\Pr ob\left(Y=j\right)\right) \qquad (10)$$

[23] The original model of this type is discussed in McFadden (1977).

where the d_{ji} are indicator variables which take the value of 1 if the i^{th} individual chooses continue, refinance, sale, or default, respectively, in any given time period, and 0 otherwise.

4. DATA

4.1 Mortgage Data and Identification of Refinance, Sale and Default

Table 1 describes data from a large loan servicer and originator includes information on 1,985 fixed-rate mortgages with both 30-year and 15-year maturities. All loans are originated during 1993 and 1994 and the properties are in three California counties: Contra Costa, Los Angeles and Orange. Approximately 79% of the loans were originated to refinance an existing mortgage loan on the same property while 21% were loans for home purchases.

Because of high housing costs in California, the loans had an average original loan balance of $167,600. The borrower income is also relatively high, with an average of over $8,000 per month; about 20% of the borrowers have monthly income greater than $10,000. About one quarter of borrowers is classified as minorities.

In order to identify sale, we purchased six years of transactions data from the California Market Data Cooperative, Inc (CMDC). The sales for the three 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.

We match the full street address of the collateral underlying the loan, the origination date, loan amount and appraisal 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. As of December 31, 1998, 27 loans (1.4%) had terminated by default and 573 loans (28.9%) had terminated by prepayment. We estimate that moves triggered 252 of the prepayments and refinancing resulted in the remaining 321.[24]

4.2 Event-history Data Organization

Since the 1,985 loans were traced from origination through termination, or data collecting point, December, 31, 1998, whichever was earlier, we follow the standard technique in the literature to construct an event-history dataset: it has one observation for each quarter for each loan during the observation period. Thus it records a full history of each loan except those censored on the data collecting point. We stack all observations together, and the dataset expands to 38,301 observations.

Tables 2 and 3 give descriptive statistics on these 38,301 observations. Besides the variables we observe at loan origination, we estimate several time varying variables, like market price of the loan, current loan-to-value ratio and probability of negative equity greater than or equal to 90 percent[25]. We also match county level unemployment rate into our event-history data. We will use these 38,301 observations in our model estimation.

The organization of the event-history data is designed to model the borrower choice at every historical point rather than only on termination or censoring points. This is reasonable because at each quarter before termination, the mortgage borrower made the choice to continue to make scheduled payment rather than refinance, sale and default, and this choice making is part

[24] The data set used here is basically the same as the one in Clapp et al (2000, 2001). For more details on data source and manipulation, please refer to those two papers.
[25] Details on how these variables are constructed are provided in tables 2 and 3.

of the borrowers' mortgage termination choice. The problem with the event-history data structure is that we impose another assumption on the data, that is, observations across time for each individual are i.i.d.

4.3 Preliminary Bivariate Data Analysis

Before running any model, we want to take a preliminary look at the data. Table 4 gives means and standard deviations of time-varying variables at loan origination and termination by the choice of refinance, sale or default.

Those refinance loans have higher than normal market price at loan termination, which is consistent with the call option being in-the-money.. Defaulted loans have much higher values for the two variables related to the put option: probability of negative equity greater or equal to 90 percent and current loan to value ratio. At the termination point, the average unemployment rate is higher for defaulted loans, implying a positive relationship between default and unemployment rate. Those sale choice records have the lowest average value of probability of negative equity greater or equal to 90 percent, which is also consistent with the "positive equity encourages sale" story. Thus, adjustment towards optimal mortgage amounts is consistent with simple averages of the data.

5. ESTIMATION AND PREDICTION

5.1 Explanatory Variables in the Models

We choose only six variables in our main model specification. Later we will present an expanded specification with sixteen explanatory variables as used in Clapp et al (2001). We will show that the ten variables added to the main specification are rarely significant and the overall data fitting does not improve with the expanded specification. We believe the main specification includes most of the important measurable variables in our models.

The market price of the loan derived from a close form formula used by Collin-Dufresne and Harding (1999) and Clapp et al (2001) is the proxy for call option. It is expected to have positive impact on refinance. For sale, one may argue that the increase of market price of the loan indicates the benefit of prepaying the loan and getting a new one. However, others may say that the increase of market price may increase the probability of negative equity, thus put constraints on move and sale. Therefore, its impact on sale is ambiguous.

We include two variables related to put option: the current loan-to-value ratio (CLTV) and the probability of negative equity. Both variables should be positively related to default. Since the default option competes with the refinancing option, we expect negative signs in the refinance choice. For example, the borrowers with high current LTV and/or negative equity will face additional constraints from lenders concerned about default. Inertial forces include high probability of negative equity which will lower the borrowers' likelihood of sale because of lack of equity needed to purchase another house. High CLTV will not have an independent constraint on sale because the borrower can obtain a high ratio loan on the new house.

County level unemployment rate is included in our models. On one hand, acting as an proxy for the economics environment, high unemployment in the area where the borrower lives is expected to result in more borrowers being unable to qualify for a new loan (thus a negative effect on sale and refinance) and more borrowers who have difficulty servicing the existing debt (positive effect on default). On the other hand, as discussed previously, high local unemployment may drive the borrowers to move out to find better employment opportunities, implying a positive impact on sale and default.

The literature indicates that minority status adds certain constraints on the borrower, so we include a minority indicator in the model and expect it to have negative signs for refinance and sale choices.

The last variable in the main specification is the indicator of poor credit history (low credit score). Borrowers with poor credit history are likely to default. For refinance and sale, poor credit history constrains the borrower's ability to get new loans; thus it should have a negative effect.

As mentioned, we also add 10 more variables related to loan and borrower characteristics to the main specification in the expanded models: the choice of loan term (15 years vs. 30 years) and the points paid by the borrower are expected to have negative impacts on sale; the indicator of loan purpose at origination (refinance=1) is expected to have a positive impact on refinance because people who have previous experience with refinance are more likely to refinance again to maximize net wealth; age and high income indicator at the time of loan origination are also included in the expanded specification and are expected to have a negative effect to all three choices – older people tend to be more conservative and less mobile and higher income borrowers have higher costs of default in terms of credit reputation; we also include house price appreciation times borrower age greater than or equal to 40 and house price appreciation times borrower age less than 40 to try to capture the attitude of different age group towards investment opportunities.

5.2 Tests on the IIA Property Violation

During model estimation, we perform two diagnostic tests of whether the IIA property of MNL is violated in our mortgage data. The first one is the Hausman-McFadden test, and the second one is the Small-Hsiao test. Details of these two tests are given in Appendix A.

For the main specification, our Hausman-McFadden test gives us a value equal to 30.12, which is over the critical value of 6.57 (χ^2 with 14 degrees of freedom at 95% significance level). The Small-Hsiao test has a value of 186.02, which is also much higher than the critical value.

For the expanded specification with 16 explanatory variables, the Hausman-McFadden test value equals to 52.36, which is over the critical value of about 20 (χ^2 with 34 degrees of freedom at 95% significance level), and the Small-Hsiao test value is 385.41, which is also much higher than the critical value.

Therefore, both specification tests reject our null hypothesis, the IIA assumption of MNL, and suggest that NMNL is more appropriate for the three competing risks of mortgage termination.

5.3 Estimation Results

Table 5 provides the maximum likelihood estimation results for the main specifications of MNL and NMNL. First we can see that MNL produces coefficients consistent with our theoretical expectations. The market value of loan, which is an important indicator for value of the call option, is highly positively related to refinance, while higher unemployment rate and the minority indicator have negative impacts on refinance.

The MNL results do show that the sale choice is very different from refinance. Unlike refinance, sale is not related to the financial variable, market price of loan. Moreover, the coefficients on CLTV and probability of negative equity greater than 90 percent have the expected effects, and have opposite signs from those for refinance. Unemployment and the minority indicator are negatively related to the borrowers' sale decision, with coefficient values similar to those for refinance. For default, CLTV, the proxy for the value of the put option, and the low credit score indicator have positive signs. The only surprise comes from the probability of negative equity greater than 90 percent variable for default choice. It is expected to have a positive effect, while it has a significant negative sign here. It is possible that this variable is collinear with CLTV and picks up the second order effect of CLTV.

We are more interested in the NMNL. We can see that the coefficients are almost the same for the 6 variables except that the significance levels for probability of negative equity greater than 90 percent and the low credit score indicator change. We notice that the sale and default correlation parameter ρ is significant and has a value of 0.82. This is consistent with our previous IIA property test, which tells that the IIA property is violated and there is correlation going on between sale and default. However, comparing the estimation results from MNL and NMNL in terms of data fitting, we find no significant difference.

Turning to the estimation results from the expanded specification (Table 6), we find that the 6 variables used in the main specification still have significant effects as in the main model specification and as theory predicts. We can also see that the 10 added variables are rarely significant and that the overall data fitting does not improve (it even deteriorate based on the B.I.C. measure). Again, NMNL does not give much difference in parameter coefficients although the sale and default correlation coefficient is significant and has a value of 0.84.

Comparing the main specification (Table 5) and the expanded specification (Table 6), a very interesting finding is that, for NMNL, the correlation parameter ρ only increase from 0.82 to 0.84 when we add 10 more explanatory variables to the main specification. This means that the correlation coefficient (1- ρ^2) between the error disturbances of sale and default only decreases from 0.3276 (= 1-0.82^2) to 0.2944 (=1-0.84^2). This is interesting because, as we discussed earlier, we expect correlation between error disturbances of sale and default due to omitted mobility variables. The main model specification with only 6 variables reveals this possibility. However, one might think that when we add a lot more variables like age, income, points, and loan terms, the correlation would disappear. What we observe is that adding 10 variables only decrease the correlation a little. This is consistent with what we discussed earlier: there are important mobility characteristics missing from the seemingly complete model specification (the expanded specification); these missing variables contribute both to sale and default, leading to a violation of IIA property of MNL, even with 16 explanatory variables.

5.4 Out-of-sample Prediction

In order to further test whether NMNL provides significant improvement to MNL, we compare the out-of-sample prediction performance. Basically, we randomly split our sample into two sub-samples, and then use one sub-sample to estimate the two models. We use the estimated coefficients to predict 200 bootstrapped samples from the other sub-sample, and compare the predictive power between MNL and NMNL. Appendix B describes the details of the procedure and the predictive performance criteria.

Table 7 gives the performance comparisons for the 200 bootstrapped samples. Panel 1 is the pool choice predicting error, which measures how accurate the model predicts the refinance, sale and default rates for the aggregated pool of out-of-sample mortgages. We can see that NMNL provides no difference with MNL in terms of refinance and sale prediction, while its prediction error for default is significantly lower than MNL. This supports the use of NMNL where pool default rates are at issue. Panel 2 shows the R-square when we regress the real choice of refinance, sale and default on the predicted probability of each of these choices. It shows no difference between NMNL and MNL.

From the above analysis, it seems that NMNL improves somewhat over MNL in model estimation and in prediction. However, the advantages of NMNL may be much greater than would appear from these tests. When we look back into our data and results, we find two important facts: first, we have only 27 defaults in our data; second, the correlation coefficient between the error disturbance of sale and that of default is pretty low ($1- \rho^2 = 0.3276$). The sparse defaults might be the reason that not many variables are significant for default choice; both this and the low correlation coefficient may cause small or insignificant differences between MNL and NMNL[26]. We need further exploration.

6. SIMULATION

Since we have too few defaults in our data, we simulate a worse economic environment for the mortgage market by shocking values of the explanatory variables. We shock the 6 variables by two standard deviations: this simulates rising interest rates, a deteriorating housing market (decreasing house prices), and an overall economy worse than in the 1990s. Second, since we have a low correlation coefficient, we shock ρ by one standard deviation to increase the correlation coefficient. Based on these two shocks and the estimated coefficients from NMNL, we simulate a set of mortgage termination data. We add noise into the data based on the error disturbance assumptions of NMNL. Then, we re-estimate MNL and NMNL based on the simulated data, and use the new estimates to do predictions. Finally we compare the estimation and prediction from MNL and NMNL.

The rationale of this procedure is as follows: assuming that mortgage termination choices are correlated as argued above, we want to know whether NMNL gives better results than MNL in terms of estimation and prediction. If this is the case, we know that under certain circumstances, NMNL improves over MNL.

[26] The sparse default might also be the cause of the low correlation coefficient.

Table 8 gives termination choices of the simulated datasets. There are 2.37% refinance, 6.08% sale, and 1.74% of default; the largest increases are in sale and default relative to the real data.[27] This conforms to the market environment we are simulating: the interest rate is going up, house prices are going down and the overall economy is worse than that in 1990s.

Table 9 gives estimation results based on the simulated data[28]. Now we can see that there is significant difference between NMNL and MNL: e.g., the coefficient of the low credit score indicator for default in NMNL turns strongly significant with the expected sign, while the coefficient for sale choice becomes insignificant. The minority indicator parameter for default in NMNL becomes significant with a negative sign; moreover, the magnitudes of the NMNL coefficients on all three termination choices are roughly the same, supporting a constrained choice set. Also, the magnitude of other variables like CLTV and probability of negative equity differs between MNL and NMNL. For overall data fitting, MNL is much worse than NMNL.

Turning to prediction, we can see in Table 10 that NMNL performs significantly better than MNL. In panel 1, we see that the pool choice predicting errors are smaller for all three choices in NMNL than those in MNL. This is especially true for the default choice which is more important in this simulation. Similarly, in panel 2, we can see that the R-squares for NMNL prediction models of both sale and default are significantly higher than those for the corresponding MNL prediction models.

7. CONCLUSION

[27] The absolute increase in refinances relative to the real data (2.37% simulated, .84% real) is due to the noise in this particular sample, limitation to the six significant variables (e.g., the current loan to value is omitted), the higher correlation coefficient and other changes such as reduction in the percent minority applicants. The important point is the decline in refinances relative to sale and default.

[28] In order to check potential sampling problems in the simulated disturbances, we simulate different sets of data and the estimation results are identical.

Property markets respond to shocks in market equilibrium mediated by important observed and unobserved characteristics of firms and households. Mortgage market activities are often seen as one of the most important aspects of property market dynamics. As suggested by Whitehead and Odling-Smee (1975), empirical models are required to use implications of theories of optimal decisions while also allowing for heterogeneous characteristics of decision making units. The nested multinomial logit model (NMNL) provides a way to do this for problems that can be cast in the nested framework.

We apply NMNL and the more standard multinomial logit model (MNL) to the decision to terminate a mortgage by prepayment, sale or default. Omitted household mobility characteristics are the mediating variables that challenge MNL models of these three choices. Important household mobility factors include changes in family status, education, neighborhood effects, job relocation opportunities and income shocks, which are usually unobservable in loan level micro data. Since these household mobility characteristics contribute at the same time to both sale and default decisions, these two choices are correlated through unobserved terms. This is not conforming to the underlying assumptions of MNL, and leads to the violation of the IIA property. Therefore, if the omitted household mobility characteristics problem is serious, MNL will give biased estimation and lead to inaccurate predictions of termination risks. But, NMNL can prevent this problem – it allows sale and default to be correlated through unobserved variables, and it can model this correlation and incorporate it into estimation and prediction. Under such circumstances, NMNL is expected to be an improved modeling tool over MNL.

This paper use data on 1,985 mortgage loans originated during 1993 and 1994 in three California counties to investigate the omitted mobility characteristics problem in mortgage termination, and tests NMNL against MNL. Our results on the event-history of these loans support the existence of the omitted mobility characteristics problem: the IIA property of MNL is violated and the correlation between sale and default is significant. However, possibly due to sparse default observations, estimation and prediction from NMNL differ from MNL only with respect to defaults.

In order to overcome the data limitation, we simulate a new set of data, reflecting a different market environment than we really have in the data – interest rates are increasing, house prices are falling and the overall economy is worse than that in the real data. Based on the simulated data, we again investigate the omitted mobility characteristics problem. We find that MNL generates different estimates of termination choice parameters, and has lower performance in prediction than NMNL does.

The success of our model may generalize to any problem that can be structured as a discrete choice with three or more alternatives to "do nothing." For example, consider the impact of relocation decisions on sales prices of houses. If the effect of mobility depends on the pattern of migration among small, medium and large metropolitan areas, the mobility decision can be recast as a decision among these three types of areas as alternatives to not moving. The choice between large and medium sized areas may be correlated through omitted variables such as profession or change in employment status, variables that mediate between the optimal location decision and the actual decision. In this example, the nested logit model provides a useful econometric method for evaluating and predicting mobility among metropolitan areas sorted by size.

APPENDIX A: TESTS ON THE IIA PROPERTY VIOLATION

The Hausman-McFadden test is proposed in Hausman and McFadden (1984) to empirically test whether IIA assumption is violated and a NMNL specification rather than MNL is needed. The null hypothesis is

$$H_0 : F(\varepsilon_s, \varepsilon_d) = \prod_j F(\varepsilon_j), j = s, d ;$$

$$H_1 : H_0 \text{ is not true} \tag{1}$$

Following Hausman and McFadden (1984), we first run an MNL for all four choices with all observations and get parameter estimates $\hat{\beta}$ (the number of variables is *k,* and the number of estimated parameters is 3*k*). Then we run another MNL for only three choices: continue, refinance, and default with observations not terminated by sale, and get new estimates $\hat{\beta}^*$ (the number of variables is *k,* and the number of estimated parameters is 2*k*).

Denote $\hat{q} = \hat{\beta} - \hat{\beta}^*$. Then the test statistics is,

$$\hat{q}' \operatorname{cov}\left(\hat{q}\right) \hat{q} \sim \chi^2(2k) \tag{2}$$

The second specification test is the Small-Hsiao test, which is a modified version of the McFadden-Train-Tye likelihood ratio test. Following Small and Hsiao (1985), we can divide the sample randomly into two parts A and B of (asymptotically equal) sizes of N^A and N^B. Estimate two MNL for four choices separately, and get estimates obtained by maximizing the respective likelihood functions, L^A and L^B. Let $\hat{\theta}_0^A$ and $\hat{\theta}_0^B$ be the two sets of parameter estimates, and let

$$\hat{\theta}_0^{AB} = \left(1/\sqrt{2}\right)\hat{\theta}_0^A + \left[1 - \left(1/\sqrt{2}\right)\right]\hat{\theta}_0^B \tag{3}$$

Then focus on the sub-sample B, and use the subset, which has only continue, refinance and default observations, and estimate a three choice MNL. Let $\hat{\theta}_1^B$ be the estimate obtained by maximizing the conditional likelihood L_1^B (with the subset consisting of the three choice set). The test statistics is:

$$\Delta = -2\left[L_1^B\left(\hat{\theta}_0^{AB}\right) - L_1^B\left(\hat{\theta}_1^B\right)\right] \sim \chi^2(2k) \quad (4)$$

APPENDIX B: COMPASISON OF OUT-OF-SAMPLE PREDICTION

We follow the following procedure to do the out-of-sample prediction: 1)Randomly split the sample into two sub-samples of roughly equal size; 2) Use one sub-sample (called sub-sample 1) to estimate MNL and NMNL models; 3) Bootstrap the other sub-sample (sub-sample 2) to form 200 test samples; 4) Use estimates from above to predict the 200 test samples.

In order to compare the predicting power of MNL and NMNL, we calculate the following two measures: 1) The pool choice predicting error: we aggregate the real choices and the predicted probability of choices of sub-sample 2, and calculate the relative error. The real termination probability for the aggregate of out-of-sample loans was divided by the predicted probability of termination; the absolute value of the percentage error is reported. This is a meaningful measure because mortgage lenders and MBS investors are most interested in the termination risks of a pool of mortgage loans. Since we are assuming our observations to be i.i.d., we can aggregate them. 2) The R-square for the prediction sample measures prediction error at the individual mortgage level. We regress the real choices on the predicted probability of choices of sub-sample 2; the R-square from this regression is our measure. This measure is valid because a better model is supposed to give a better mapping from the predicted choices to real choices.

We do the out-of-sample prediction with 200 bootstrapped test samples as described above, and then calculate means and standard deviations of the two measures of predicting power. Finally we analyze whether the means of the measures from MNL are significantly different than those from NMNL.

References

Ambrose, Brent W, Richard J. Buttimer, Jr. and Charles A. Capone. (1997). "Pricing Mortgage Default and Foreclosure Delay." *Journal of Money, Credit and Banking* 29(3), 314-25.

Ambrose, Brent W. and Charles A. Capone, Jr. (2000). "The Hazard Rates of First and Second Defaults." *Journal of Real Estate Finance and Economics* 20(3), 275-293.

Ambrose, Brent W. and Anthony B. Sanders. (2003). "Commercial Mortgage Backed Securities: Prepayment and Default." *Journal of Real Estate Finance and Economics* 26(2/3), 179-196.

Amemiya, T. (1985). *Advanced Econometrics*. Cambridge, Massachusetts: Harvard University Press.

Archer, W. R., and D. C. Ling. (1993). "Pricing Mortgage-backed Securities: Integrating Optimal Call and Empirical Models of Prepayment." *Journal of American Real Estate and Urban Economics Association* 21(4), 373–404.

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(3/4), 235-261.

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

Ball, M. and R. Kirwan, "Accessibility and supply constraints in the urban housing market." *Urban Studies*, 14 (1977) 11-32.

Bennett, P., R. Peach and S. Peristiani. (2001). "Structural Change in the Mortgage Market and the Propensity to Refinance." *Journal of Money, Credit and Banking* 33(4), 955-975.

Berkovec, J. A., G. B. Canner, S. A. Gabriel, and T. Hannan. (1998). "Discrimination, Competition, and Loan Performance in FHA Mortgage Lending." *Review of Economics and Statistics* 80(2), 241-250.

Boehm, T. (1981). "Tenure Choice and Expected Mobility: A Synthesis." *Journal of Urban Economics* 10(3), 375–389.

Boehm, T. and K. R. Ihlandfeldt. (1986). "Residential Mobility and Neighborhood Change." *Journal of Regional Science* 26, 411-424.

Brennan, M. J., and E. S. Schwartz. (1985). "Determinants of GNMA Mortgage Prices." *Journal of the American Real Estate and Urban Economics Association* 13(3), 209-228.

Buser, S. A., and P. H. Hendershott. (1984). "Pricing Default Free Mortgage." *Housing Finance Review* 3(3), 405-429.

Calhoun, C. A. and Y. Deng. (2002). "A Dynamic Analysis of Fixed- and Adjustable-Rate Mortgage Terminations." *Journal of Real Estate Finance and Economics* 24(1-2), 9-33.

Cauley, S.. (1996). California Real Estate Market. Quarterly UCLA-Anderson Business Forecast.

Ciochetti, Brian A., Yongheng Deng, Bin Gao, and Rui Yao. (2002). "The Termination of Commercial Mortgage Contracts through Prepayment and Default: A Proportional Hazards Approach with Competing Risks." *Real Estate Economics* 30(4), 595-633.

Ciochetti, Brian A., Yongheng Deng, Gail Lee, James Shilling and Rui Yao. (2003). "A Proportional Hazards Model of Commercial Mortgage Default with Originator Bias." *Journal of Real Estate Finance and Economics* 27(1), 5-23.

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

Clapp, J, G. Goldberg, J. Harding and M. LaCour-Little. (2001). "Movers and Shuckers: Interdependent Prepayment Decisions." *Real Estate Economics* 29(3), 411-450.

Clapp, J., Y. Deng, and X. An. (2004). "Alternative Models for Competing Risks of Mortgage Termination." University of Southern California, Lusk Center for Real Estate Working Paper, No. 2004-1002.

Clapp, J. M. and Y. Wang. (2006). “Defining Neighborhood Boundaries: Are Census Tracts Obsolete?” *Journal of Urban Economics* 59(2), 259-284.

Collin-Dufresne, P. and J. P. Harding. (1999). “A Closed Form Formula for Valuing Mortgages.” *Journal of Real Estate Finance and Economics* 19(21), 133-146.

Cunningham, D. F., and C. A. Capone. (1990). “The Relative Termination Experience of Adjustable to Fixed-Rate Mortgages.” *Journal of Finance* 45(5), 1687–1703.

Cunningham, D. F. and P. H. Hendershott. (1984). “Pricing FHA Default Insurance.” *Housing Finance Review* 3(3), 373–92

Deng, Y., J. M. Quigley, and R. Van Order. (1996). “Mortgage Default and Low Down-payment Loans: The Cost of Public Subsidy.” *Regional Science and Urban Economics* 26(3/4), 263-285.

Deng, Y. (1997). “Mortgage Termination: An Empirical Hazard Model with Stochastic Term Structure.” *Journal of Real Estate Finance and Economics* 14(3), 309-331.

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

Deng, Y, and J. M. Quigley. (2002). “Woodhead Behavior and the Pricing of Residential Mortgages.” University of Southern California, Lusk Center for Real Estate Working Paper, No. 2003-1005.

Deng, Y. and S. A. Gabriel. (2006). “Risk-Based Pricing and the Enhancement of Mortgage Credit Availability among Underserved and Higher Credit-Risk Populations.” Forthcoming in *Journal of Money, Credit, and Banking*.

Deng, Y., A. D. Pavlov, and L. Yang. (2005). “Spatial Heterogeneity in Mortgage Terminations by Refinance, Sale and Default.” Forthcoming in *Real Estate Economics* 33(4).

Dunn, K. B., and J. J. McConnell. (1981). “Valuation of Mortgage-Backed Securities.” *Journal of Finance* 36(3), 599-617.

Epperson, J. F., J. B. Kau, D. C. Keenan, and W J. Muller, III. (1985). “Pricing Default Risk in Mortgages.” *Journal of American Real Estate and Urban Economics Association* 13(3), 261-72.

Findley, M. C., and D. R. Capozza. (1977). “The Variable Rate Mortgage: An Option Theory Perspective.” *Journal of Money, Credit and Banking* 9(2), 356-364.

Follain, J. R., O. Scott, and T. L. T. Yang. (1992). “Microfoundations of a Mortgage Prepayment Function.” *Journal of Real Estate Finance and Economics* 5(2), 197–217.

Foster, C., and R. Van Order. (1984). “An Option-Based Model of Mortgage Default.” *Housing Finance Review* 3(4), 351-372.

Giliberto, S. M., and D. C. Ling. (1992). “An Empirical Investigation of the Contingent-Claims Approach to Pricing Residential Mortgage Debt.” *Journal of American Real Estate and Urban Economics Association* 20(3), 393-426.

Goldberg G. M. and J. P. Harding. (2003). “Investment Chracteristics of Low- and Moderate-income Mortgage Loans.” *Journal of Housing Economics* 12(3), 151-180.

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(3), 233-248.

Green, J., and J.B. Shoven. (1986). “The Effect of Interest Rates on Mortgage Prepayment.” *Journal of Money, Credit and Banking* 18(1), 41-50.

Hausman, J., and D. McFadden. (1984). “Specification Testes for the Multinomial Logit Model.” *Econometrica* 52(5), 1219-1240.

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

Huang, Wenyi and Jan Ondrich. (2002). “Stay, Pay or Walk Away: A Hazard Rate Analysis of FHA-Insured Multifamily Mortgage Terminations.” *Journal of Housing Research* 13 (1), 85-117.

Hurst, E. and F. Stafford. (2004). “Home Is Where the Equity Is: Mortgage Refinancing and Household Consumption”. *Journal of Money, Credit and Banking* 36(6), 985-1014.

Ihlandfeldt, K., and J. Silberman. (1985). “Differential Response to Change: The Case of Home Purchase.” *Journal of Urban Economics* 17(2), 127–144.

Kau, J.B. and D.C. Keenan. (1995). “An Overview of the Option-Theoretic Pricing of Mortgages.” *Journal of Housing Research* 6(2), 217-244.

Kau, J.B., D.C. Keenan, W. Muller and J. Epperson. (1992). “A Generalized Valuation Model for Fixed-Rate Residential Mortgages.” *Journal of Money, Credit and Banking* 24(3), 279-299.

Krumm, R. J. (1984). “Household Tenure Choice and Migration.” *Journal of Urban Economics* 16(3), 259–271.

Lambrecht, B. M., W. R. M. Perraudin and S. Satchell. (2003). “Mortgage Default and Possession Under Recourse: A Competing Hazards Approach”. *Journal of Money, Credit and Banking* 35(3), 425-442.

Luce, D. (1959). *Individual Choice Behavior*. New York: John Wiley and Sons.

Maclennan, Duncan, M. Munro and G. Wood. (1987) “Housing Choices and the Structure of Housing Markets.” in *Between State and Market: Housing in the Post-Industrial Era*, edited by Bengt Turner, Jim Kemeny and Lennart J. Lundqvist, Almqvist & Wiksell International 26-52.

Mattey J., N. Wallace. (2001). “Housing-price cycles and prepayment rates of US mortgage pools.” *Journal of Real Estate Finance and Economics* 23(2), 161-184.

McFadden, D. (1974). “Conditional Logit Analysis of Qualitative Choice Behavior.” In *Frontiers in Econometrics,* edited by P. Zarembka, pp. 105-142. New York: Academic Press.

McFadden, D. (1977). “Econometric Models of Probabilistic Choice.” In *Structural Analysis of Discrete Data with Econometric Applications*, edited by C. F. Manski and D. McFadden, pp. 198-272. Cambridge, Massachusetts: MIT Press.

McFadden, D. (2001). “Economic Choices.” *American Economic Review* 91(3), 351-378.

Pavlov, A. D. (2001). “Competing Risks of Mortgage Termination: Who Refinances, Who Moves, and Who Defaults?” *Journal of Real Estate Finance and Economics* 23(2), 185-211.

Philips, R. A., E. Rosenblatt and J. H. VanderHoff. (1995). “The Probability of Fixed and Adjustable Rate Mortgage Termination.” *Journal of Real Estate Finance and Economics* 13(2), 95-104.

Quigley, J. M. (1987). “Interest Rate Variations, Mortgage Prepayments, and Household Mobility.” *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.

Quigley, J. M. and R. Van Order. (1990). “Efficiency in the Mortgage Market: The Borrower’s Perspective.” *Journal of American Real Estate and Urban Economics Association* 18(3), 237-252.

Quigley, J. M. and R. Van Order. (1995). “Explicit Tests of Contingent Claims Models of Mortgage Default.” *Journal of Real Estate Finance and Economics* 11(2), 99-117.

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

Rossi, P. H. (1955). *Why Families Move*. Glencoe: The Free Press.

Shear, W. B. (1983). “Urban Housing Rehabilitation and Move Decisions.” *Southern Economics Journal* 49(4), 1030–1052.

Schnare, A. B. and R. J. Struyk. (1976) “Segmentation in Urban Housing Markets.” *Journal of Urban Economics*, 8, 146-166.

Schwartz, E. S., and W. N. Torous. (1989). “Prepayment and the Valuation of Mortgage-Backed Securities.” *Journal of Finance* 44(2), 375-392.

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

Speare, A. Jr. (1974). "Residential Satisfaction as an Intervening Variable in Residential Mobility." *Demography* 11(2), 173-188.

Spilimbergo, A. and L. Ubeda. (2004). "Family Attachment and the Decision to Move by Race." *Journal of Urban Economics* 55(3), 478-497.

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

Straszheim, M. R.. (1975) *An Econometric Analysis of the Urban Housing Market*. Urban and Regional Studies Number 2, National Bureau of Economic Research, New York.

Train, K. E. (2003). *Discrete Choice Methods with Simulation*. Cambridge, UK: Cambridge University Press.

Vandell, K.D., W.C. Barnes, D.J. Hartzell, D. Kraft and W. Wendt. (1993). "Commercial Mortgage Defaults: Proportional Hazards Estimation Using Individual Loan Histories." *Journal of American Real Estate and Urban Economics Association* 20(4), 451-480.

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

Watkins, Craig. (2001) "The definition and identification of housing submarkets." *Environment and Planning A*, 33 2235-2253.

Whitehead, C. M. E. and J. C. Odling-Smee. (1975) "Long-run Equilibrium in Urban Housing – a Note." *Urban Studies*, 12, 315-318.

Table 1: Means and standard deviations of variables at loan origination

Variables	Means (STDs)	Description
Original loan balance ($000)	167.63 (121.83)	Face amount of the mortgage at the date of origination (in 1993 or 1994), in thousands of dollars.
15-year loan indicator	0.31 (0.46)	Indicator variable is one if the loan has a 15 year maturity, zero if a 30 year maturity.
Original refinance indicator	0.79 (0.41)	One if the mortgage at the date of origination (in 1993 or 1994) was to refinance a previous mortgage, zero if it was to purchase the home.
Borrower age	46.74 (11.18)	Age of the borrower in years, from the loan application.
Minority indicator	0.23 (0.42)	Indicator variable equal to one if the application classifies borrowers into any one of three minority groups, otherwise zero.
Borrower income ($000)	8.08 (9.03)	Monthly household income at the time of origination.
Obligation ratio (%)	30.12 (9.60)	The ratio of fixed expenses to borrower income. This is the standard ratio used by lenders when evaluating loan applications.
High credit score indicator	0.64 (0.48)	The high score indicator flags borrowers with credit scores greater than 1000 — a group that includes approximately 50% of the borrowers.
Low credit score indicator	0.103 (0.30)	The low score indicator flags borrowers with credit scores less than 800 — a group that includes approximately 10% of the borrowers.
Number of observations	1985	Number of loans with data on all variables.

Notes: 1. Standard deviations are in parentheses.
2. All loans are originated in the year of 1993 or 1994 and traced through 12/31/1998.
3. The credit score here is not the FICO score. It is an internally developed score used by the data provider.

Table 2: Descriptive Statistics for Continuous Variables Used in Our Models

Variable	Mean	S.T.D	Min	Max	Med.	Description
Market price of loan* ($ per $100 of principal)	98.30	3.33	88.27	110.29	98.32	The present value of the remaining payments at the current interest rate: an adjustment is made for the option to terminate the loan early.
Original loan balance ($00,000)	167.63	121.83	14.00	1,452.00	131.20	Same as in table 1.
Estimated Points	2.03	1.39	-8.44	7.94	2.02	We regress the individual loan coupon rate on treasury rates, loan and borrower characteristics. The residuals from this equation provide a measure of points since a borrower paying a rate substantially below the predicted rate must have "bought down" the rate by paying above average points.
Current loan-to-value* (%)	57.26	24.16	5.13	175.55	58.21	Current loan balance was estimated from the original balance and amortization. The local regression model described in Clapp *et al.* (2001) estimated the house value and its standard deviation at each point in time The estimated value of the house was divided into the estimated current loan balance to get the current loan-to-value ratio.
House price appreciation × Age ≥ 40 indicator ($,000)	3,553	27,755	-405,994	613,024	0.00	The appreciation variables were constructed as follows: house price appreciation in thousands of dollars is multiplied by an indicator of borrower age and by borrower age.
House price appreciation × Age < 40 indicator ($,000)	317	13,013	-179,628	619,165	0.00	Same as above.
Unemployment rate* (%)	6.63	1.87	2.80	10.05	6.57	The unemployment rate is for the county of residence in each quarter.
Borrower age	46.88	11.36	18.00	89.00	46.00	Same as in table 1.

Obligation ratio	29.92	9.62	0.00	85.00	31.00	Same as in table 1.
Number of Observations				38,301		The data is constructed as an event-history dataset, with one observation for each quarter for each loan during the observation period. 1,985 Loans were observed from origination to termination, or 12/31/1998, whichever was earlier. Number of records expands from 1,985 to 38,301.

Note: Variables with * are those used in our main model, which has a simplified specification with only six explanatory variables.

Table 3: Frequency Tables for Dummy Variables Used in Our Models

	Value	Frequency	Percent	Description
15-year loan indicator	0	26,167	68.32	Same as in table 1.
	1	12,134	31.68	Same as in table 1.
Original refinance indicator	0	7,429	19.40	Same as in table 1.
	1	30,872	80.60	Same as in table 1.
Prob. Negative Equity > 90 percentile indicator*	0	34,457	89.96	The house value was compared to the current loan balance and the normal distribution was used to estimate the probability of negative equity.
	1	3,844	10.04	
Minority indicator*	0	29,011	75.74	Same as in table 1.
	1	9,290	24.26	Same as in table 1.
High borrower income indicator	0	30,676	80.09	Monthly household income at the time of origination greater than $10,000.
	1	7,625	19.91	
High credit score indicator	0	13,303	34.73	Same as in table 1.
	1	24,998	65.27	Same as in table 1.
Low credit score indicator*	0	34,781	90.81	Same as in table 1.
	1	3,520	9.19	Same as in table 1.
Termination choice				
Refinance	1	321	0.84	
Sale	1	252	0.66	
Default	1	27	0.07	
Continue to pay	1	37,701	98.43	
Number of Observations		38,301	100	

Note: Variables with * are those used in our main model, which has a simplified specification with only six explanatory variables.

Table 4: Means and standard deviations of time-varying variables at loan origination and termination

Variables	At Origination				At Termination		
	All Loans	Refinance	Sale	Default	Refinance	Sale	Default
Market price of loan	100.00	99.85	99.89	99.65	101.36	98.53	101.55
($ per $100 of principal)	(1.86)	(1.43)	(1.33)	(2.68)	(3.18)	(3.12)	(3.46)
Prob. Negative Equity > 90	0.15	0.17	0.09	0.42	0.09	0.04	0.15
percentile indicator	(0.36)	(0.37)	(0.28)	(0.50)	(0.29)	(0.20)	(0.37)
Current loan-to-value (%)	60.17	52.93	54.03	88.69	54.63	55.38	85.18
	(22.57)	(22.03)	(21.34)	(12.55)	(23.41)	(22.14)	(14.74)
Unemployment rate (%)	8.28	8.22	8.35	8.51	5.69	6.01	6.40
	(1.67)	(1.64)	(1.62)	(1.61)	(1.80)	(1.79)	(1.42)
Number of Observations	1,985	321	252	27	321	252	27

Notes: 1. Standard deviations are in parentheses.
2. Descriptions of the variables are in table 1 through table 3.

Table 5: Estimation Results for the Main Model Specifications of Multinomial Logit Model (MNL) and Nested Logit Model (NMNL)

Parameter	Multinomial Logit Model			Nested Logit Model		
	Refinance	Sale	Default	Refinance	Sale	Default
Intercept	-5.41***	-4.97***	-8.55***	-5.41***	-4.97***	-7.97***
	(0.1)	(0.09)	(0.59)	(0.1)	(0.1)	(1.35)
Market price of loan	0.89***	-0.06	0.44*	0.89***	-0.05	0.36
	(0.07)	(0.09)	(0.21)	(0.07)	(0.1)	(0.21)
Current loan-to-value	-0.04	0.20*	0.67*	-0.04	0.20*	0.66*
	(0.07)	(0.08)	(0.35)	(0.07)	(0.08)	(0.35)
Prob. Negative Equity >	0.11	-0.87*	-1.57*	0.11	-0.89**	-1.51**
90 percentile indicator	(0.22)	(0.34)	(0.61)	(0.22)	(0.34)	(0.57)
Unemployment rate	-0.24***	-0.23**	0.01	-0.24***	-0.23**	0.02
	(0.07)	(0.07)	(0.29)	(0.07)	(0.07)	(0.27)
Minority indicator	-0.73***	-1.07***	-0.36	-0.73***	-1.06***	-0.43
	(0.16)	(0.23)	(0.43)	(0.16)	(0.22)	(0.41)
Low credit score	-0.37	-0.42	2.78***	-0.37	-0.28	2.36*
indicator	(0.23)	(0.33)	(0.51)	(0.23)	(0.42)	(0.91)
Nested Logit Model					0.82*	
correlation parameter (ρ)					(0.34)	
Log Likelihood		-3,289.05			-3,289.39	
B.I.C.		3,420.97			3,426.58	

Notes: 1. Standard deviations are in parentheses. “*” for $p < 0.05$; “**” for $p < 0.01$; “***” for $p < 0.001$.

2. The data are structured as an event-history dataset, with one observation for each quarter for each loan during the observation period. Loans were observed from origination to termination, or 12/31/1998, whichever was earlier. There are 38,301 observations.

3. Continuous variables are standardized before model estimation.

4. We use loan age dummies, whose coefficients are omitted here. They can be provided upon request.

5. Both the MNL and the Nested Logit Models are estimated through maximum likelihood estimation (MLE). The Nested Logit Model is estimated following the simultaneous estimation approach rather than two-stage estimation with inclusive value.

Table 6: Estimation Results for the Expanded Model Specifications of the Multinomial Logit Model (MNL) and the Nested Logit Model (NMNL)

Parameter	Multinomial Logit Model			Nested Logit Model		
	Refinance	Sale	Default	Refinance	Sale	Default
Intercept	-4.91***	-5.40***	-10.51***	-4.91***	-5.40***	-9.83***
	(0.19)	(0.25)	(1.38)	(0.19)	(0.25)	(2.04)
Market price of loan	0.81***	-0.13	0.57	0.81***	-0.12	0.49
	(0.08)	(0.11)	(0.3)	(0.08)	(0.11)	(0.31)
Original loan balance	0.24***	-0.11	-0.17	0.24***	-0.11	-0.17
	(0.06)	(0.1)	(0.93)	(0.06)	(0.1)	(0.89)
15-year loan indicator	-0.07	-0.30*	-0.93	-0.07	-0.30*	-0.88
	(0.15)	(0.16)	(1.87)	(0.15)	(0.16)	(1.71)
Estimated Points	-0.05	-0.18*	0.20	-0.05	-0.17*	0.17
	(0.07)	(0.07)	(0.26)	(0.07)	(0.07)	(0.26)
Original refinance indicator	-0.57***	0.23	1.37	-0.57***	0.24	1.27
	(0.15)	(0.23)	(0.96)	(0.15)	(0.23)	(0.93)
Current loan-to-value	-0.34**	0.08	0.79	-0.34**	0.08	0.73
	(0.1)	(0.1)	(0.96)	(0.1)	(0.11)	(0.9)
Prob. Negative Equity > 90 percentile indicator	0.20	-0.89*	-1.71	0.20	-0.91*	-1.57
	(0.23)	(0.36)	(0.96)	(0.23)	(0.36)	(0.91)
House price appreciation $\times$Age $\geq$ 40 indicator	-0.07	-0.18*	-0.20	0.01	0.03	0.08
	(0.05)	(0.07)	(1.04)	(0.03)	(0.07)	(0.68)
House price appreciation $\times$Age < 40 indicator	0.01	0.04	0.07	-0.07	-0.18**	-0.19
	(0.03)	(0.07)	(0.72)	(0.05)	(0.07)	(1)
Unemployment rate	-0.25***	-0.27***	0.05	-0.25***	-0.27***	0.03
	(0.07)	(0.08)	(0.3)	(0.07)	(0.08)	(0.28)
Borrower age	-0.04	-0.18	-0.24	-0.04	-0.18*	-0.23
	(0.07)	(0.1)	(0.27)	(0.07)	(0.1)	(0.25)
Minority indicator	-0.66***	-1.04***	-0.30	-0.66***	-1.03***	-0.36
	(0.17)	(0.24)	(0.51)	(0.17)	(0.24)	(0.47)

Table 6 (Continue): Estimation Results for the Expanded Model Specifications of the Multinomial Logit Model (MNL) and the Nested Logit Model (NMNL)

Parameter	Multinomial Logit Model			Nested Logit Model		
	Refinance	Sale	Default	Refinance	Sale	Default
High borrower income indicator	0.05 (0.16)	0.41* (0.21)	-0.15 (1.61)	0.05 (0.16)	0.41* (0.21)	-0.14 (1.51)
Obligation ratio	0.07 (0.06)	0.03 (0.07)	0.35 (0.54)	0.07 (0.06)	0.03 (0.07)	0.33 (0.51)
High credit score indicator	-0.16 (0.14)	0.22 (0.16)	0.96 (1.74)	-0.16 (0.14)	0.23 (0.16)	0.91 (1.58)
Low credit score indicator	-0.50* (0.24)	-0.17 (0.36)	3.61* (1.78)	-0.50* (0.24)	-0.06 (0.44)	3.21* (1.91)
Nested Logit Model correlation parameter (ρ)					0.84* (0.35)	
Log Likelihood		-3,247.26			-3,247.21	
B.I.C.		3,537.48			3,542.70	

Notes: 1. These are expanded specifications with sixteen variables as used in the Clapp et al (2001) study.

2. Standard deviations are in parentheses. "*" for $p < 0.05$; "**" for $p < 0.01$; and "***" for $p < 0.001$.

3. The data are structured as an event-history dataset, with one observation for each quarter for each loan during the observation period. Loans were observed from origination to termination, or 12/31/1998, whichever was earlier. There are 38,301 observations.

4. Continuous variables are standardized before model estimation.

5. We use loan age dummies, whose coefficients are omitted here. They can be provided upon request.

6. Both the MNL and the Nested Logit Models are estimated through maximum likelihood estimation (MLE). The Nested Logit Model is estimated following the simultaneous estimation approach rather than two-stage estimation with inclusive values.

Table 7: Out-of-sample Predictive Performance Comparison between Multinomial Logit Model (MNL) and Nested Logit Model (NMNL)

Panel 1
Pool Choice Prediction Error

	MNL	NMNL	Difference	t-statistics
Refinance	7.85%	7.87%	-0.02%	
	6.10%	6.11%	1.93%	-0.01
Sale	6.68%	6.35%	0.32%	
	4.97%	4.85%	1.55%	0.21
Default	42.86%	25.89%	16.97%	
	20.91%	12.45%	5.44%	3.12

Panel 2
R-Square

	MNL	NMNL	Difference	t-statistics
Refinance	0.0171	0.0171	0.0000	
	0.0032	0.0032	0.0010	0.00
Sale	0.0022	0.0022	0.0000	
	0.0007	0.0007	0.0002	-0.08
Default	0.0018	0.0018	0.0000	
	0.0014	0.0014	0.0004	0.08

Notes: 1. These are results from 200 bootstrapped test samples. The dataset (38,301 observations) is randomly split in half. The in-sample half was used to estimate the model and these coefficients were used to predict the probability of each termination hazard for each test sample bootstrapped from the out-of-sample half. Standard deviations are in parenthesis.
2. Pool choice prediction error means that the actual probability of termination for the entire out-of-sample "pool" was divided by the predicted probability of termination and the ratio was converted to a percentage error. The R-squares track prediction accuracy at the individual loan level. Please see Appendix B for more detailed explanation of the out-of-sample prediction and performance comparison criteria.
3. The t-values use standard formulas to test differences between means.

Table 8: Termination Choices of Simulated Data

Termination choice	Frequency	Percent
Refinance	907	2.37
Sale	2,330	6.08
Default	666	1.74
Continue to pay	34,398	89.81
Number of Observations	38,301	100

Notes: . This reflects a different market environment from that of the real data: interest rates are increasing, house prices are falling and overall economy is worse than that in the 1990s.

Table 9: Estimation Results for the Main Model Specifications of the Multinomial Logit Model (MNL) and the Nested Logit Model (NMNL) Based on the Simulated Data

Parameter	Multinomial Logit Model			Nested Logit Model		
	Refinance	Sale	Default	Refinance	Sale	Default
Intercept	-3.52*** (0.29)	-2.39*** (0.23)	-7.00*** (0.36)	-3.61*** (0.29)	-3.18*** (0.23)	-6.48*** (0.63)
Market price of loan	0.26*** (0.05)	-0.09* (0.04)	0.59*** (0.05)	0.26*** (0.05)	-0.06* (0.04)	0.20*** (0.05)
Current loan-to-value	0.10 (0.06)	0.02 (0.04)	1.27*** (0.08)	0.08 (0.06)	0.05 (0.04)	0.41*** (0.08)
Prob. Negative Equity > 90 percentile indicator	0.16 (0.17)	-0.21 (0.13)	-1.61*** (0.19)	0.15 (0.17)	-0.29* (0.12)	-0.99*** (0.17)
Unemployment rate	0.05 (0.05)	-0.07* (0.03)	0.02 (0.06)	0.06 (0.05)	-0.08* (0.03)	0.03 (0.05)
Minority indicator	-0.21 (0.12)	-0.39*** (0.08)	0.09 (0.12)	-0.22* (0.12)	-0.41*** (0.08)	-0.23* (0.1)
Low credit score indicator	-0.13 (0.18)	-1.46*** (0.22)	0.36 (0.82)	0.02 (0.18)	-0.11 (0.27)	2.08*** (0.25)
Nested Logit Model correlation parameter (ρ)					0.46*** (0.10)	
Log Likelihood		-7593.17			-7331.16	
B.I.C.		7711.49			7459.33	

Notes: 1. The estimation sample (19,142 observations) is half the full sample of simulated data (sub-sample 1).

2. Standard Deviations are in parentheses. "*" for $p < 0.05$; "**" for $p < 0.01$; and "***" for $p < 0.001$.

3. The data are structured as an event-history dataset, with one observation for each quarter for each loan during the observation period. Loans were observed from origination to termination, or 12/31/1998, whichever was earlier.

4. Continuous variables are standardized before model estimation.

5. We use loan age dummies, whose coefficients are omitted here. They can be provided upon request.

6. Both the MNL and the Nested Logit Models are estimated through maximum likelihood estimation (MLE). The Nested Logit Model is estimated following the simultaneous estimation approach rather than two-stage estimation with inclusive values.

Table 10: Out-of-sample Predictive Performance Comparison between Multinomial Logit Model (MNL) and Nested Logit Model (NMNL) Based on the Simulated Data

Panel 1
Pool Choice Prediction Error

	MNL	NMNL	Difference	t-statistics
Refinance	21.12%	19.91%	1.21%	
	3.13%	3.18%	1.00%	1.21
Sale	5.07%	4.06%	1.01%	
	1.37%	2.05%	0.55%	1.83
Default	36.72%	15.07%	21.65%	
	6.00%	3.27%	1.53%	14.17

Panel 2
R-Square

	MNL	NMNL	Difference	t-statistics
Refinance	0.0018	0.0018	0.0000	
	0.0006	0.0006	0.0002	0.12
Sale	0.0070	0.0083	-0.0014	
	0.0013	0.0012	0.0004	-3.49
Default	0.1328	0.1551	-0.0223	
	0.0136	0.0140	0.0044	-5.11

Notes: 1. These are results from 200 bootstrapped test samples. The dataset (38,301 observations) is randomly split in half. The in-sample half was used to estimate the model and these coefficients were used to predict the probability of each termination hazard for each test sample bootstrapped from the out-of-sample half. Standard deviations are in parenthesis.

2. Pool choice prediction error means that the actual probability of termination for the entire out-of-sample "pool" was divided by the predicted probability of termination and the ratio was converted to a percentage error. The R-squares track prediction accuracy at the individual loan level. Please see Appendix B for more detailed explanation of the out-of-sample prediction and performance comparison criteria.

3. The t-values use standard formulas to test differences between means.