Institute of
Business and
Economic Research

Fisher Center for
Real Estate and
Urban Economics

PROGRAM ON HOUSING AND URBAN POLICY

DISSERTATION AND THESES SERIES

DISSERTATION NO. D07-002

MODELING RESIDENTIAL MORTGAGE PERFORMANCE WITH RANDOM UTILITY MODELS

By

Carolina Marquez

Spring 2007

These papers are preliminary in nature: their purpose is to stimulate discussion and comment. Therefore, they are not to be cited or quoted in any publication without the express permission of the author.

UNIVERSITY OF CALIFORNIA, BERKELEY

Modeling Residential Mortgage Performance with Random Utility Models

by

Carolina Marquez

BS (University of Chicago) 1997
MFE (University of California, Berkeley) 2004

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Economics

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, BERKELEY

Committee in charge:

Professor Paul Ruud, Chair
Professor John Quigley
Professor Nancy Wallace

Fall 2007

The dissertation of Carolina Marquez is approved:

Chair Date

Date

Date

University of California, Berkeley

Fall 2007

Modeling Residential Mortgage Performance
with Random Utility Models

Copyright 2007
by
Carolina Marquez

Abstract

Modeling Residential Mortgage Performance
with Random Utility Models

by

Carolina Marquez
Doctor of Philosophy in Economics

University of California, Berkeley

Professor Paul Ruud, Chair

The papers in this dissertation explore a random utility framework for improving models used to analyze the payment trends of residential mortgage borrowers. With growth in the issuance of mortgage-backed securities, analysis of borrower behavior is an important area of study. Accurate models of borrower behavior are rare because of the complicated dynamics that incorporate both financial and behavioral elements. This dissertation explores three issues that are essential for accurately modeling the borrower's decision process: time-varying covariates, correlated competing risks, and correlation over time.

The first chapter focuses on the implementation of time-varying covariates and proposes two alternatives to the popular Cox proportional hazard rate model for predicting prepayment probabilities. Both alternatives exploit the natural panel structure of mortgage data in order to effectively incorporate time-varying covariates. A comparison of the three models reveals that a logit model, based on a random utility framework, yields prepayment rates that are closer to the

observed probabilities than a Cox proportional hazard rate model or a log-logistic proportional hazard rate model.

The second chapter introduces an ordered logit model as an appropriate tool for analyzing crucial features of the behavior of borrowers, who often avail themselves of options other than full prepayment and default, such as partial prepayment and delinquency. An ordered logit model allows for a large set of correlated choices, while circumventing the independence from irrelevant alternatives property found in the more common multinomial logit model. In addition, the random utility functions used in this model allow borrower behavior to be conditioned on past decisions.

The final chapter demonstrates the importance of random coefficients for accurately predicting borrower behavior. The choices made by a borrower are influenced by past events and attitudes, leading to correlation in the borrower's behavior over time. Random coefficients are used to account for unobserved heterogeneity and link the choices made by the borrower throughout the life of the loan. The results in this chapter indicate that behavior predictions vary substantially with the model specification. Random coefficients, although not previously applied to mortgages, allow for a more accurate representation of the decision process than fixed coefficients.

Professor Paul Ruud
Dissertation Committee Chair

To my parents.

I would like to thank my family and friends for their support and unwavering faith. Special thanks to Paul Ruud, Nancy Wallace, John Quigley, and Jim Wilcox for their many years of guidance and patience. Thanks, also, to Linda Kreitzman and all of those involved with the Master's in Financial Engineering program at the Haas School of Business, without which I may not have entered the world of finance.

Contents

List of Figures

List of Tables

Introduction

This dissertation examines several important issues associated with modeling the payment behavior of residential mortgage borrowers. Improvements to existing methodologies are made through the application of random utility models, which provide a structural link between the borrower's decision process and economic variables. The papers in this dissertation focus on three main issues: implementing time-varying covariates; modeling correlated competing risks; and accounting for unobserved heterogeneity and correlation over time. In each chapter, elements of a random utility model are introduced to address the failings of existing models, building up to an ordered logit model with random coefficients.

Mortgage-backed securities make up the largest proportion of bond market debt. The fast-paced growth of this sector has heightened the role of the housing market in the U.S. economy, and increased the demand for accurate mortgage valuation models. An important area of research is the performance of residential mortgages. This is a challenging field, however, because the dynamics that underlie the borrower's decision process incorporate both financial and behavioral elements. Reduced-form models, mostly derived from models of survival analysis, have the flexibility to capture the high degree of heterogeneity observed among residential mortgage borrowers, but they do not always account for the relevant financial dynamics, a defining feature of structural contingent-claims models. The random utility framework, introduced in this dissertation for modeling borrower behavior, displays the potential for bridging the reduced-form and structural modeling methodologies. The ground work for future research in this direction is laid out in the chapters that follow.

One of the contributions of this dissertation is a loan-level analysis of the payment trends of non-conforming, prime mortgages issued and securitized between 2000 and 2005.[1] This analysis is useful for several reasons. First, this is a period that has not been frequently studied despite economic trends that have created a rich environment for mortgage activity in both the primary and secondary markets. Second, little emphasis has been placed on non-conforming, prime mortgages even though they represent the fastest growing segment of the market for mortgage-backed securities (MBS). Third, a loan-level analysis uncovers payment trends that are not frequently studied but are highly relevant to the valuation of a mortgage.

The first chapter examines the implementation of time-varying covariates in the analysis of prepayment. Two alternatives to the popular Cox proportional hazard rate model are proposed: a log-logistic proportional hazard rate model, and a logit random utility model. While the focus on prepayment presents a

[1] Non-conforming loans are those that cannot be guaranteed by Freddie Mac or Fannie Mae, usually because the loan balance exceeds the prescribed limit. Prime mortgages are those issued to borrowers with good credit profiles.

simplified framework, it is useful because many studies of prime mortgages — both academic and in industry — treat prepayment separately from default. In addition, this approach facilitates a direct comparison between the three models, and serves as a necessary introduction to mortgage modeling within a random utility framework.

It is shown that the two alternative models presented in the first chapter are better suited for modeling borrower behavior with time-varying covariates than the Cox proportional hazard rate model. The advantages of the log-logistic proportional hazard rate model and the logit random utility model are that they use the natural panel structure of mortgage data to effectively incorporate time-varying covariates, and they allow for the estimation of a parametric hazard rate function that is essential for pricing algorithms. The Cox proportional hazard rate model is found to predict behavior with respect to time-varying variables that is counter to economic intuition, calling into question the conclusions of previous studies that have used this model. In a comparison of forecasted probabilities, the logit random utility model outperforms the log-logistic proportional hazard rate model, which is constrained by the functional form of the baseline hazard rate function. It can be concluded, from the results in the first chapter, that a random utility model is preferable to models based on survival analysis, even for simple applications such as the study of prepayment trends.

The second chapter extends the analysis to consider a large set of competing risks, which includes non-termination events. Two models are compared: a multinomial logit model along the lines of that used in the subprime literature, and an ordered logit model that has not been previously used with respect to mortgages. An aspect of borrower behavior that is often overlooked is the degree to which the borrower deviates from the amortization schedule. As the status of the mortgage changes, so does the borrower's decision process. To accurately model mortgage borrower behavior it is necessary to employ a model that considers an expanded set of mortgage choices and conditions the borrower's utility on past decisions. The models presented in the second chapter, both based on a random utility framework, meet these requirements.

Although an ordered logit model has not previously been used with respect to mortgages, it is well suited for this analysis since all of the mortgage events can be defined relative to the amortization schedule and, thus, can be set on an ordered continuum. Through the ordering of the choice set, the ordered logit model accounts for correlation among alternatives, in contrast to the multinomial logit model, which assumes independence between choices through the independence from irrelevant alternatives (IIA) property. The results in the second chapter indicate that, although a multinomial logit model has a higher degree of parameterization, which allows it greater flexibility to track in-sample probabilities, the IIA property leads to questionable predictions on borrower behavior. The ordered logit model generates more favorable estimates, and is likely to perform better out-of-sample since it accounts for correlated competing risks, which is a fundamental aspect of the borrower's decision process.

The third chapter presents a key innovation of this dissertation: the use of random coefficients in mortgage analysis. The decision made by a borrower with respect to his mortgage is influenced by past events and attitudes, as well as the historical path of economic variables. This causes correlation in the borrower's behavior over time, a feature that is ignored by many models. Random coefficients, often associated with the modeling of unobserved heterogeneity, provide an ordered logit model with a borrower-specific error component that is constant over time and can be used to induce correlation in borrower behavior.

In the third chapter, it is shown that predicted borrower behavior varies substantially with the model specification. In the comparison of a fixed coefficients ordered logit model and an independent random coefficients ordered logit model, it is found that the random coefficients model has a better fit, and predicts behavior that is in line with economic intuition. This is particularly true for delinquent loans, and it is an important result because these loans have not previously been studied in depth. There is also evidence that it is necessary to account for correlation between coefficients, particularly between coefficients corresponding to different initial states. This allows the model to take into account all of the borrower's past choices, and more accurately represents the borrower's decision process.

This dissertation shows that a random utility framework, in particular an order logit model, is well suited to modeling the various dynamics of borrower behavior. Just as it has been shown that this model has the flexibility to incorporate time-varying covariates, model correlated competing risks, and account for unobserved heterogeneity and correlation over time, the straightforward structure of an ordered logit model lends itself to many additional extensions, from enhancing pricing algorithms with empirical probability distributions to modeling a dynamic choice process. There are many interesting areas of future research along these lines. The models introduced in this dissertation show great potential for substantially improving the accuracy of valuation models, and increasing our understanding of borrower behavior.

Chapter 1

Focus on Time-Varying Covariates

1.1 Introduction

With the growth in the issuance of mortgage-backed securities, the role of the housing market in the U.S. economy has come under the spotlight, and the performance of residential mortgages has become an object of interest. A fundamental aspect of this analysis is the modeling of prepayment, a common practice in which a borrower pays all of the remaining principal balance on his mortgage prior to the scheduled term. This practice creates uncertainty in the security cash flows, and it causes an interest shortfall that shortens the duration of the security. Accurate prepayment models are fundamental for both investors in the market for mortgage-backed securities and originators of individual mortgages.

The most common causes of prepayment are refinancing and property sale. When interest rates fall, it is to the borrower's advantage to lower the total mortgage cost by contracting a new mortgage to pay off the original higher rate mortgage. For most borrowers this is a relatively low-cost transaction. Prepayment will also occur when the borrower sells his home and is required to pay off his existing mortgage in order to transfer ownership. Both events are driven by fluctuations in economic variables, namely interest rates and housing prices. Accurate prepayment models must capture the *paths* of these variables and appropriately link them to the borrower's prepayment decision.

The development of prepayment models is an active area of research. Many models, especially in recent studies, acknowledge the dependence on the path of economic variables through the use of time-varying covariates. The most common approach for incorporating time-varying covariates is through the Cox (1972) proportional hazard rate model, but this is an open issue that has not been adequately resolved. In this chapter, I propose two fully parametric alternatives that take advantage of the natural panel structure of mortgage data. This feature allows the researcher to effectively incorporate time-varying covariates, easily deal with tied events, and account for censoring. The first alternative uses episode splitting to extend the proportional hazard rate model of Schwartz and Torous (1989). The second alternative applies a random utility model to the prepayment decision.

Early prepayment models, such as Dunn and McConnell (1981b) and Brennan and Schwartz (1985), were based on the contingent claims model of Black and Scholes (1973) and Merton (1973). The focus of these "structural" models is on how the movement of interest rates affects the prepayment option embedded in a mortgage. Stochastic processes are assumed for the key factors underlying the mortgage, and an expression for the mortgage value is derived. The time to prepayment is given by a set of boundary conditions that trigger the exercise of the termination option.

Reduced-form models of prepayment were introduced by Green and Shoven (1986) and Schwartz and Torous (1989). These models, like many that followed, are based on the Cox (1972) proportional hazard rate model, which links the probability distribution of a duration variable to a set of covariates. Regression methods are employed to study the effect of observable characteristics on the probability of prepayment. Proxy variables are often used to represent the value of the termination option.

Reduced-form models are attractive because they allow the researcher to account for heterogeneity among borrowers, and determine how loan characteristics and exogenous economic factors contribute to the borrower's behavior. To partially account for the path of interest rates and housing prices, the covariates of the model are allowed to vary with time. Cox (1972) showed that a proportional hazard rate model could easily be adapted to incorporate time-varying covariates through episode splitting and partial likelihood estimation. The drawbacks of the Cox proportional hazard rate model are that it requires a balanced panel, which is often not available with loan-level mortgage data, and that the baseline hazard rate function, which determines how prepayments vary with loan age, is factored out of the model.

An alternative to prepayment models based on survival analysis is the use of a random utility model. Even though discrete choice theory has been used in the study of prepayment (Green and LaCour-Little (1999) and Clapp et. al. (2001)), an explicit link has not been made to the random utility maximization that underlies the prepayment decision. Since a borrower undertaking the prepayment decision is attempting to minimize the cost of his mortgage, or alternatively maximize his total portfolio holdings, a random utility model is a natural framework. This approach also presents an alternative structural model, and provides greater flexibility for introducing time-varying covariates and other elements that are integral to an accurate representation of the borrower's decision process.

Random utility models were first introduced by Thurstone (1927) in the context of psychological stimuli. A formal link to utility maximization was made by Marschak (1960). Dagsvik (1994) and McFadden and Train (2000) have shown that any random utility model can be approximated by a mixed logit model. This finding makes random utility models an attractive tool for a variety of questions in which a choice is to be made between discrete alternatives. Random utility models have been used widely in consumer decision making, labor economics, industrial organization, and transportation science, among many other fields that share features present in mortgage analysis.

This chapter compares three reduced-form models for analyzing the prepayment decision with time-varying covariates: the Cox proportional hazard rate model, a log-logistic proportional hazard rate model, and a logit random utility model. Section 1.2 presents the details of the three models. Section 1.3 discusses the dataset, a panel of about 100,000 non-conforming, fixed-rate, prime, residential mortgages issued between 2000 and 2005. This is an interesting dataset because it covers recently issued non-conforming mortgages, whereas most studies look at conforming mortgages securitized by Fannie Mae, Freddie Mac, or Ginnie Mae. It is valuable to examine this time period because of the importance of the housing market over this period, and the implications it holds for current economic forecasts.

The results reported in section 1.4 indicate that both alternatives introduced in this chapter perform better than the more commonly used Cox proportional hazard rate model. Comparisons are made on the basis of coefficient estimates, hazard rate elasticities, and forecasting power. There is evidence that the Cox model does not properly capture the time dynamics of the prepayment decision, calling into question the conclusions of previous studies in which the Cox model is used. While the log-logistic and logit models predict the same general trends, the logit model has greater flexibility and is better able to capture borrower heterogeneity and accurately forecast hazard rates. This chapter concludes that a random utility model is a preferable framework for several reasons: it uses the natural structure of mortgage data in order to properly capture time trends; it allows for the estimation of a fully parametric hazard rate function with sufficient flexibility to capture borrower heterogeneity; and it serves as an alternative structural framework by establishing an explicit relationship between the borrower's decision process and the economic factors of interest.

1.2 Models

1.2.1 Cox Proportional Hazard Rate Model

One of the most commonly used reduced-form models for analyzing mortgage performance is the Cox (1972) proportional hazard rate model. The benefits of this model include the ability to associate loan characteristics with prepayment rates, and to easily adapt the model to include time-varying covariates. A useful measure for mortgage valuation is the hazard rate, which is defined as the probability of prepayment in a time period, given survival to that period. Since prepayment leads to the termination of the mortgage, it is reasonable to use a hazard rate model where the underlying variable is the time to mortgage termination, T, measured as the number of months from origination. Estimation of the hazard rate is based on the probability distribution of this variable. The hazard rate is defined as the probability of failure at time t given survival to time t,

$$h(t) = \lim_{\Delta t \to 0} \frac{\Pr(t \leq T < t + \Delta t | T \geq t)}{\Delta t} = \frac{f(t)}{S(t)} = -\frac{d \ln S(t)}{dt}$$

where the survival function, $S(t)$, is defined as the probability of surviving to a given time, and is equal to one minus the cumulative distribution function, $F(t)$.

$$S(t) = \Pr(T \geq t) = 1 - F(t)$$

The probability density of a failure is given by

$$f(t) = \lim_{\Delta t \to 0} \frac{\Pr(t \leq T < t + \Delta t)}{\Delta t} = -\frac{dS(t)}{dt}.$$

By integration we can derive the following relationships between the hazard rate, the survival function, and the probability density function:

$$S(t) = \exp\left(-\int_0^t h(u)du\right) \quad f(t) = h(t)\exp\left(-\int_0^t h(u)du\right).$$

In the Cox model, the hazard rate is a function of the baseline hazard rate, h_0, and loan characteristics, v.

$$h(t; v) = h_0(t)exp(\beta v)$$

The baseline hazard rate function describes how termination varies over time. Covariates serve to shift the baseline hazard rate to account for heterogeneity among borrowers. Often proportional hazard rate models are written in log-hazard form to facilitate the interpretation of the coefficients.

$$\log h(t) = \kappa + \beta v \quad \kappa = \log h_0(t)$$

One of the features of the Cox model is that, by using a partial likelihood estimation method, the baseline hazard rate is factored out of the model.[1] In many applications this is beneficial because assumptions on the shape of the baseline hazard rate function can be avoided. The partial likelihood is calculated as the product of the individual likelihoods for all observed events. The likelihood for each event is given by the probability that the event occurred to the corresponding

[1]The baseline hazard rate can be estimated non-parametrically after the estimation of the coefficient vector β.

individual, rather than any other individual in the dataset. In other words, it is the hazard rate for the target individual divided by the sum of the hazard rates for all individuals at risk at that time.[2]

$$PL = \prod_{i=1}^{n} L_i = \prod_{i=1}^{n} \left[\frac{e^{\beta v_i}}{\sum_{k=1}^{n} d_{ik} e^{\beta v_k}} \right]^{\delta_i} \qquad \begin{array}{l} d_{ik} = 1 : t_k \geq t_i \\ \\ d_{ik} = 0 : t_k < t_i \end{array}$$

t_i and t_k are termination times for individuals. i represents the target individual, and k is an index that cycles through all individuals to determine the observations that are still at risk when individual i terminates. δ_i is a censoring index; $\delta_i = 0$ for a right-censored observation so that those observations are excluded from the partial likelihood.

Cox assumes continuous time, but with mortgage data we observe termination at discrete intervals, leading to the problem of ties. If it is assumed that the termination decision is undertaken at discrete intervals, rather than just recorded at discrete intervals, an exact partial likelihood with ties can be calculated.[3] This is a reasonable assumption in the case of mortgages because a termination is not effective until the end of the billing cycle. In the previous expression the partial likelihood is compiled by multiplying over individuals. In discrete time the partial likelihood is calculated by multiplying over unique event times.[4] In the equations below, j is an index for event time, and N_j is the number of individuals with event time j.

$$PL = \prod_{j=1}^{J} L_j \quad L_j = \frac{\psi_j}{\psi_j + \psi_{j+1} + ... + \psi_J}, \quad \psi_j = \prod_{i=1}^{N_j} O_{ij}$$

[2]This is the procedure followed by the SAS command PHREG.
[3]Breslow (1974) and Efron (1977) suggest approximations.
[4]An event time is defined as the number of months from origination to prepayment.

$O_{ij} = e^{\beta v_i}$ where individual i has a termination time of j.

In many applications of survival models, the relevant factors are static over the life of the individual. With mortgage termination, however, many of the key factors, such as interest rates and housing prices, vary over time. The path of these variables is often more relevant than the level. In the Cox model, the partial likelihood is a product over all termination times. To incorporate time-varying covariates the ratios that make up the partial likelihood are evaluated at each risk period, rather than solely at the termination period. A risk period is a time period in which the individual is still "alive", in other words still at risk for termination. In this way, each individual is observed in multiple periods.[5]

$$PL = \prod_{j=1}^{J} L_j \quad L_j = \frac{\psi_j(j)}{\psi_j(j) + \psi_{j+1}(j) + ... + \psi_J(j)}$$

$$\psi_j(k) = \prod_{i=1}^{N_j} O_{ij}(k) \quad O_{ij}(k) = e^{\beta v_{ik}}$$

1.2.2 Parametric Proportional Hazard Rate Model

While the Cox proportional hazard rate model is one of the most popular methods for incorporating time-varying covariates, it has several drawbacks. First, since the baseline hazard rate has been factored out of the model, this method is not useful for mortgage valuation in which it is necessary to forecast an out-of-sample hazard rate function. Second, estimation of the Cox model with time-varying covariates requires a balanced panel. Loan-level mortgage data does not

[5] Guidelines for estimating a Cox model with time-varying covariates in SAS can be found in Allison (1995).

conform to this structure, and so the Cox model leads to loss of data. Third, when covariates are allowed to vary with time, the model can no longer be considered a proportional hazard rate model because the covariates for different individuals vary at different rates. These problems can be addressed by extending the Schwartz and Tourous (1989) log-logistic proportional hazard rate model to include time-varying covariates.

Schwartz and Tourous (1989) proposed one of the first reduced-form models for valuing residential mortgages. They use a parametric proportional hazard rate model to price pools of Ginnie Mae mortgages. To capture *a priori* beliefs on the movement of the hazard rate over time, the baseline hazard rate is assumed to follow a log-logistic function.

$$h_0(t;\gamma,p) = \frac{\gamma p(\gamma t)^{p-1}}{1+(\gamma t)^p}$$

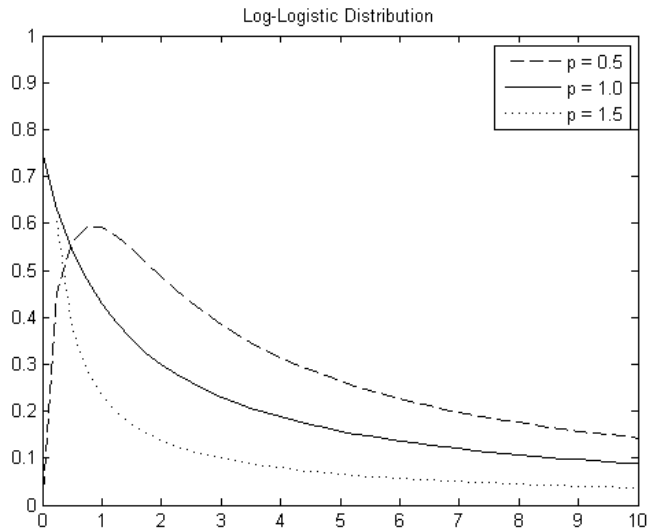
Under this assumption, a closed-form solution for the probability density of failure times can be derived.

$$f(t) = \frac{\gamma p(\gamma t)^{p-1}}{1+(\gamma t)^p}\exp(\beta v)\exp\left(-\exp(\beta v)\ln(1+(\gamma t)^p)\right)$$

$$S(t) = \exp\left(-\exp(\beta v)\ln(1+(\gamma t)^p)\right)$$

The log-logistic baseline hazard rate function is a useful specification because it captures the trends we observe with respect to seasoning on mortgage pools. Prepayment rates are low at first, and then grow rapidly, eventually slowing down as the pool ages and most of the loans that are prone to prepayment have exited. The parameter p characterizes the shape of the hazard rate over time. When p

is greater than one, the hazard rate first increases and then decreases. When p is between zero and one, the hazard rate decreases over time. It is important to note that the log-logistic proportional hazard rate model proposed in this chapter differs from that estimated by many statistical packages.[6]



Unlike the Cox model, the log-logistic proportional hazard rate model is fully parametric and can be estimated using maximum likelihood. The likelihood func-

[6]In many cases the log-logistic model is not characterized as a true proportional hazard rate model. Instead the structure is a proportional odds model that is derived from a linear log-survival time. This is the model estimated by the log-logistic option in SAS with PROC LIFEREG.

$$logT = \beta v + \sigma\epsilon$$

To obtain a log-logistic model it is assumed that the residual has a logistic distribution.

$$f(\epsilon) = \frac{e^{\epsilon}}{(1+e^{\epsilon})^2}$$

The hazard rate function is derived as

$$h(t) = \frac{\lambda p(\lambda t)^{p-1}}{1+(\lambda t)^p} \quad p = \frac{1}{\sigma} \quad \lambda = \exp\left[-(\beta v)\right].$$

The most convenient way to interpret this model is in terms of odds-ratios. The survival function is given by $S(t) = \frac{1}{1+(\lambda t)^p}$ and the odds-ratio is $\log\left[\frac{S(t)}{1-S(t)}\right] = p(\beta v - \log t)$. Under this specification, the coefficients do not have the natural interpretation given by the Schwartz and Torous model.

tion must take into account three events: mortgages that prepaid prior to the observation period (left-censored); mortgages that prepaid during the observation period (observed events); and mortgages that have not prepaid by the end of the observation period (right-censored). For left-censored mortgages, the relevant probability is the cumulative probability of prepayment prior to the observation period, $F(t)$, where t is the time from origination to the start of the observation period. For observed events the likelihood is given by the probability of prepaying at time t, $f(t)$, where t is the time from origination to prepayment. Finally, for right-censored observations we are concerned with the probability of not prepaying by the end of the observation period, $S(t)$, where t is the time from origination to the end of the observation period. If we let η be the set of left-censored mortgages, ε the set of observed events, and μ the set of right-censored mortgages, then we have the following log-likelihood function.

$$\ln L(\theta) = \sum_{i \in \eta} \ln(1 - S(t_i)) + \sum_{i \in \varepsilon} \ln(f(t_i)) + \sum_{i \in \mu} \ln(S(t_i))$$

Often with the available loan-level mortgage data there are no left-censored observations so we can simplify the log-likelihood.

$$\ln L(\theta) = \sum_{i \in uncensored} \ln h(t_i) + \sum_{i \in all} \ln S(t_i)$$

In Schwartz and Torous (1989) the covariates are static, measured at the time of termination. While it is clear that this is a poor assumption with respect to prepayment, it is relatively easy to extend the model to include time-varying covariates. The key is in the definition of the hazard rate. Since the hazard rate is a conditional probability, it can be identically defined in a panel dataset. The method used to extend survival analysis models to a panel dataset is known as episode splitting. An episode is defined by a start time, an end time, an origination state, and a destination state. Each episode is accompanied by a set of covariates. For mortgages, an episode is normally defined as a one month payment period. In each episode a borrower decides whether or not to prepay. The benefit of this approach is that, owing to the monthly payment cycle, mortgage data is naturally structured for methods that require episode splitting.

Under this framework, it is necessary to adjust the probability definitions that enter the likelihood function. By episode splitting we divide T into q pieces with $T_0 = 0$ and $T_q = T$. Since the hazard rate is defined as a conditional probability, it requires no modification. The survival function can also be defined in terms of conditional probabilities.

$$S(t_k|t_{k-1}) = \frac{S(t_k)}{S(t_{k-1})} = \frac{\exp\left(-\int_0^{t_k} h(u)du\right)}{\exp\left(-\int_0^{t_{k-1}} h(u)du\right)}$$

$$= \exp\left(-\int_0^{t_k} h(u)du + \int_0^{t_{k-1}} h(u)du\right) = \exp\left(-\int_{t_{k-1}}^{t_k} h(u)du\right)$$

Using Bayes theorem we have that for each loan $S(t) = \prod_{k=1}^{q} S(t_k|t_{k-1})$ and so the likelihood function can be defined as:

$$\ln L(\theta) = \sum_{i \in uncensored} \ln h(t_i) + \sum_{i \in all} \sum_{k=1}^{q} \ln S(t_{ik}|t_{ik-1}).$$

Since the covariates are assumed to be invariant between t_{k-1} and t_k, the integral in the conditional survival function has a closed form.

Several assumptions are necessary for this derivation. The first, mentioned above, is that the covariates are constant over the episode. While this is not strictly true for continuous variables, such as interest rates, the path of these variables can be smoothed without an adverse effect. The second assumption is that the covariates are not dependent on the response variable. Formally, the hazard rate function must be conditioned on the path of the covariates, in addition to survival to time t.

$$h(t; \mathrm{V}(t)) = \lim_{\Delta t \to \infty} \frac{\Pr(t \leq T < t + \Delta t | T \geq t, \mathrm{V}(t))}{\Delta t}$$

This introduces the path probability of the covariates into the equation for the survival function,

$$S(t; V(t)) = \prod_{k=1}^{q} S(t_k, v(t_k) | t_{k-1}, V(t_{k-1}))$$

$$= \prod_{k=1}^{q} S\left(t_k | t_{k-1}, V(t_{k-1})\right) \times \Pr\left(v(t_k) | t_{k-1}, V(t_{k-1})\right),$$

where $v(t)$ represents the contemporaneous values of the covariates and $V(t)$ represents the full path of covariate values.

Kalbfleisch and Prentice (2002) show that in the case of time-invariant covariates, $V(t) = v$, or path independent covariates, $V(t) = v(t)$, the path probability, $\Pr\left(v(t_k) | t_{k-1}, V(t_{k-1})\right)$, does not play a role. In many cases, however, the variables of interest are both time-varying and path dependent, so the path probability must be considered. Kalbfleisch and Prentice (2002) separate time-varying covariates into three groups: external, exogenous, and endogenous. For external covariates — measurement can be made independent of the failure process — or exogenous covariates — covariate values are not influenced by the failure process — the path probability can be factored out of the equation through episode splitting. Problems with the derivation arise when covariates are endogenous, for which the path probability is linked to the distribution of the termination variable, and a closed-form solution for the survival function cannot be obtained.

Most of the covariates that are considered in mortgage analysis are external or exogenous, so episode splitting is a viable procedure. There is, however, an additional assumption that is implicit in most mortgage performance models, but is not often discussed, namely, that the episodes for a given loan are independent over time. Conditioning on survival and the path of covariates addresses some of the correlation in behavior over time. However, many of the factors that enter a borrower's decision process are unobservable. Examples are preference for housing, income potential, and financial savvy. Strictly speaking, the survival function should be conditioned on the path of unobservable factors as well. If these factors are independent, then the path probability does not play a role in the derivation. However, many of the unobservable factors are consistent over time and endogenous to the failure process. This presents a significant challenge for modeling borrower behavior with time-varying covariates. A potential solution to this problem is the subject of the third chapter.

1.2.3 Random Utility Logit Model

Several studies have used episode splitting with regard to mortgages but in the context of discrete choice modeling. (See Green and LaCour-Little (1999) and Clapp et. al. (2001).) This follows Jenkins (1995), who showed that duration models could easily be estimated as discrete choice models, and in particular that the Cox proportional hazard rate model could be written equivalently as a logit model.

The basis for this equivalence is that in a proportional hazard rate model the odds ratio can be written as a linear function of covariates.

$$\log h(t) = \beta v$$

Under this specification, the assumption is that the time to termination, T, follows a logistic distribution. Jenkins proposes an alternative definition where the hazard rate is defined in terms of a logistic function

$$\log\left[\frac{h(t)}{1-h(t)}\right] = \beta v,$$

and makes the claim that the two specifications converge as the hazard rate becomes increasingly small.[7] Despite similarities, the assumptions underlying the two specifications are quite different. Implicit in the latter definition is a random utility model, whereas the first specification makes distributional assumptions on the duration variable.

To apply a random utility model to the prepayment decision, we again use

[7]In the analysis of mortgages the hazard rate is not necessarily small since prepayment is a relatively frequent event.

episode splitting. In each episode, the borrower makes the decision to prepay or to maintain his mortgage schedule based on which choice yields the highest utility. Let the utility associated with prepaying in a given period be a linear function of loan characteristics and economic variables.

$$U_{it} = \beta v_{it} + \epsilon_{it}$$

Utility is a latent variable that is not observed, but we do observe the choice made by the borrower in that period. Prepayment corresponds to a response variable, y, with a value equal to one. If prepayment is observed, this implies that the latent utility is positive.

$$y_{it} = 1 \Rightarrow U_{it} > 0$$

Under the assumption that the residual follows an i.i.d. Gumbel, or extreme value, distribution, the probability of prepayment is given by a logistic function.

$$Pr(y_{it} = 1) = Pr(\beta v_{it} + \epsilon_{it} > 0) = Pr(\epsilon_{it} > -\beta v_{it}) = \frac{1}{1 + exp(\beta v_{it})}$$

An alternative is to assume the residual follows an i.i.d normal distribution,

$$Pr(y_{it} = 1) = \Phi(\beta v_{it}),$$

but in the binomial, i.e. two-choice, case there is little difference between the two assumptions.

The coefficients of the model are estimated by maximum likelihood. The likelihood function is calculated by aggregating the probability of the observed choice stream for each individual.

$$\log L(\theta; v) = \sum_{n=1}^{N} \ln \left[\Pr \left(\{ y_{nt} : y_{nt} = 1 \forall t = 1 ... T_n \} \right) \right]$$

Because prepayment leads to termination, the observed choice stream has a special form. All periods prior to prepayment have a response variable equal to zero. The last element in the choice stream represents either the month of prepayment or the last month in the observation period, leading to an unbalanced panel. If prepayment is observed, this element is equal to one, otherwise it is set to zero and the loan is right-censored. The structure of the choice stream implies that the probability of a choice in any given period is equivalent to the hazard rate for that choice.[8] If the error term of the utility function is assumed to be independent over time, the log-likelihood function can be simplified.

$$\log L(\theta; v) = \sum_{n=1}^{N} \sum_{t=1}^{T_n} y_{nt} \log Pr(y_{nt} = 1)$$

1.2.4 Elasticities

While the random utility model preserves many of the important elements of proportional hazard rate models, i.e. the dependence of prepayment on loan age and the ability to forecast hazard rates, this chapter aims to show that there is a clear difference between the three models presented in this section. Since the functional forms of the models are different, hazard rate elasticities are used to facilitate a comparison. The table below gives the relevant elasticity formulas for each model. The elasticities corresponding to the logit random utility model differ in that they are a function of the hazard rate. This creates a nonlinearity that allows the random utility model to better represent observed behavior. It is

[8] When tracking loans in this way, a choice of prepayment can only be made if the loan has survived to that period. The probability of observing a response variable equal to one in any period is equivalent to the hazard rate.

	Time	**Other Covariates**
Cox Proportional Hazard	-	βv
Log-logistic Proportional Hazard	$(p-1)-th$	βv
Logit Random Utility	$\beta v(1-h)$	

interesting to note that, although age enters the random utility model linearly, the elasticity with respect to age is on the same order as that of the log-logistic proportional hazard rate model. This point will be pursued further in the results section.

1.3 Data

The three models discussed in the previous section are applied to a set of loans that serve as collateral for securities issued by Wells Fargo Mortgage Backed Securities Trust. In accordance with the guidelines of the trust, all loans are residential, first-lien, non-conforming mortgages issued by Wells Fargo Home Mortgage. The loans in this dataset are considered "jumbo loans" because, for the most part, the balance at origination exceeds the conforming limit. The source dataset contains loans that correspond to 97 collateral pools for 157 securities series issued since 2000, and comprise both fixed-rate and adjustable-rate mortgages. Performance history is available starting in October 2000. The data was obtained from Wells Fargo Corporate Trust Services, which serves as administrator and trustee to Wells Fargo Mortgage Backed Securities Trust as well as over 200 other issuers. The data is publicly available and released by request of the securities issuers.[9]

[9] The data was obtained from www.ctslink.com.

1.3.1 Origination Trends

The analysis focuses on approximately 100,000 fixed-rate mortgages issued between 2000 and 2005 on single-family, owner-occupied housing. 34 fixed-rate mortgages issued in 2000 and 2001 were dropped because they were contracted with a prepayment penalty; all the other loans in the dataset had no prepayment penalty. Summary statistics are reported in Tables 1.1 and 1.2. On average, the loans in the dataset follow the characteristics of jumbo loans. The average balance is $454,000, 95% of the loans have a loan to value (LTV) ratio below 80%, and 42% of the loans were contracted on property in California. All of the mortgages are conventional, but less than 2% carry Private Mortgage Insurance (PMI). Most of the mortgages have a term of 15 or 30 years. In 2000 and 2001 over 80% of the loans issued were 30-year mortgages but, as interest rates fell, 15-year mortgages gained in popularity, making up nearly 60% of the mortgages issued in 2003 and 2004. It is striking that only 5% of the fixed-rate mortgages were issued in 2004, in contrast with 10% to 30% for the other years.

It is important to note that, although the dataset contains a large number of loans with substantial variation in terms and performance, the loans are selected as members of pools to use in securitization and, therefore, may display different trends than the mortgage market at large. Most of the borrowers have excellent credit,[10] with 65% having a FICO score above 720, and just 1% with a FICO score below 620, which is often considered a breakpoint for poor credit. The

[10]FICO scores measure credit quality on a scale of 300 to 850.

coupon rates for the loans in the dataset follow the trends observed for conventional mortgages, with average rates falling below that of conventional mortgages towards 2005. The standard deviation in each year is on average 40 basis points. Over time the dispersion of the LTV ratio at origination increases, but in the direction of lower LTV ratios. LTV ratios are lowest for loans issued in 2003 with an average of 59%. The average appraisal value at origination of the properties mortgaged is high at nearly $800,000, but ranges from $24,000 to over $16 million. 80% of the loans are in the range of $300,000 to $1.5 million.

Since all loans are issued by Wells Fargo, one could expect regional concentration to be high in states where Wells Fargo has a strong presence, and in areas with high housing prices and hence a large proportion of jumbo loans. Nearly half of the loans are from California, but the dataset contains loans for properties in all 50 states and the District of Columbia. States with a strong representation are New York, New Jersey, Maryland, Virginia, Florida, Minnesota, Illinois, and Texas. Table 1.3 lists the region definitions, Table 1.4 reports the distribution of loans by region and issue year, and Table 1.5 displays summary statistics stratified by region. In general, loan characteristics are uniform across regions with trends in property values at origination following the OFHEO Housing Price Index (HPI).[11]

[11]This index is published by the Office of Federal Housing Enterprise Oversight (OFHEO).

1.3.2 Prepayment Trends

A standard measure for investors and originators are prepayment rates given as a function of loan age. In proportional hazard rate models, this relationship is represented by the baseline hazard rate function. A measure commonly used in industry is the Public Securities Association (PSA) Standard Prepayment Model (Figure 1.1). This model takes into account the seasoning trends that are observed in mortgage pools by assigning an expected prepayment rate according to loan age.[12]

The PSA model is a simplified convention that does not fully represent observed prepayment trends. To gain a better understanding of the relationship between loan age and prepayment rates, it is useful to calculate the empirical hazard rate function. Based on the relationship between the hazard rate, the survival function, and the probability density function, Kaplan and Meier (1958) propose an empirical survival function

$$\widehat{S}(t) = \prod_{j:t_j \leq t} \left[1 - \frac{d(t_j)}{n(t_j)}\right],$$ [13]

where $n(t)$is the number of loans that survive to time t and $d(t)$is the number of loans that fail during the interval $t+\Delta t$. Δt is taken to be equal to the billing cycle of the mortgage, one month. This expression takes into account right censoring because fractions are taken with respect to the loans that are at risk at time t.

[12] The PSA benchmark model (also called "100% PSA") assumes that the prepayment rate starts at a 0.2% constant prepayment rate (CPR) in the first month following origination of the mortgage loan (not the pool) and increases 0.2% per month in each succeeding month. The prepayment rate is assumed to level off at 6% in month 30 and beyond. The CPR is the annualized equivalent of the single monthly mortality (SMM), which measures the fraction of the beginning-month pool balance that prepays during the month.

[13] These calculations can be performed in SAS using PROC LIFETEST.

Figure 1.2 shows the empirical hazard rate function for the loans under analysis. The curve is derived by applying the Ramlau-Hansen (1983) kernel smoothing technique to the Kaplan-Meier hazard ratios calculated at each month.[14] This figure illustrates the seasoning behavior we expect with residential mortgages. The hazard rate increases with age, peaking at around 40 months, and then falls sharply. This decline in prepayment rates for seasoned loans, which is not captured by the PSA curves, is known as "burnout", and it is in order to capture this effect that a log-logistic parametric hazard rate function is preferred. For this sample, however, we also observe a dip at around 24 months, indicating that it may be necessary to allow for greater flexibility in the relationship between prepayment rates and loan age. As will be shown below, this is one of the reasons to turn to a random utility model.

Figure 1.3 shows that these trends also hold at the regional level and, in general, there is little difference in the observed prepayment behavior between regions. Prepayment rates are slightly lower in California, the New York metropolitan area, and the South. The greatest deviations from the national trends are found in the Great Lakes region and the Plains/Midwest region. Prepayment rates for the Great Lakes region fail to peak and are more stable, almost following the PSA convention. Prepayment rates for the Plains/Midwest region are more volatile than for any other region. Several regional variables are included in the analysis, but greater emphasis is placed on national factors.

[14]SAS code is available in Allison (1995).

1.3.3 Explanatory Variables

The beginning of the period observed in this dataset (2000-2005) is marked by a sharp decline in stock market prices in April 2000, propelled by the failure of many of the internet and technology ventures that grew in the 1990s (Figure 1.4). This was followed by an increase in the unemployment rate particularly in the West Coast, New England, and the Mid-Atlantic where many of these ventures were located (Figure 1.5). In order to stimulate recovery, the Federal Reserve began easing monetary conditions in January 2001. This continued until June 2004, when the federal funds rate reached a low of 1% (Figure 1.4). Growth was slow and recovery was not apparent in many regions until late in 2002 (Figure 1.6). The strength of the economy over this period was largely due to the growth in housing prices. By 2000 housing prices were already strong as a result of the expansion stimulated by the growth of the technology sector (Figure 1.7). When stock market prices declined, housing prices did not follow. The behavior of the loans in this dataset is consistent with an environment of low interest rates, high housing prices, and hesitant growth.

Interest rates enter the model through the refinance incentive. As this is a reduced-form model, this incentive is measured as the difference between the contracted coupon rate on the mortgage and the prevailing yield on the ten-year Treasury bond, the "coupon gap". For a sufficiently high coupon gap, the borrower will prepay. Lags of the coupon gap are used to account for transaction costs, and the cube of the coupon gap is included to capture nonlinear effects. One of the shortfalls of using the coupon gap as a proxy for the refinance incentive is that it does not take into account interest rate expectations. Due to transaction costs, a borrower may delay a refinancing if he expects interest rates to fall in the future. In a static model it is not possible to account for such expectations directly. However, the slope of the yield curve, calculated as the difference between the yields on the ten-year and two-year Treasury bonds, indicates the direction in which interest rates are expected to move. This factor also controls for the relative popularity of fixed-rate versus adjustable-rate mortgages.

In addition to refinancing, it is also necessary to account for property sales. Unfortunately, many of the factors that influence a borrower's decision to sell his home, such as preferences, family structure, and personal finances, are unobservable. Instead we include two factors that are correlated with the strength of the economy and, thus, may lead to higher property turnover rates. These factors are state productivity, measured as the annual growth rate in gross state product, and house price appreciation, calculated from the OFHEO House Price Index (HPI) corresponding to the property's metropolitan statistical area (MSA). These factors also account for regional variation in prepayment rates and proxy for worker mobility. Quarterly dummy variables are included to control for the spring and summer months in which property sales are high.

One of the benefits of a reduced-form model is the ability to account for heterogeneity among borrowers. We would like to observe characteristics of the borrower that determine available refinancing options. Although many of the relevant borrower characteristics are unobservable, certain features of the loan can proxy for these factors. The FICO score is used to measure the borrower's credit quality, with higher scores implying better credit. These scores are grouped into five buckets: 750-850, 720-750, 660-720, 620-660, and less than 620. The term and LTV ratio at origination are included to control for the borrower's financial flexibility and preference for debt.

1.3.4 Mortgage Purpose

Previous studies (Stanton (1995), Downing, Stanton, and Wallace (2005), Dunn and Spatt (2005)) have shown that prepayment rate forecasts from structural models can be improved by accounting for transaction costs. Dunn and Spatt (2005), in particular, find that due to transaction costs borrowers may delay an optimal refinancing if they expect a more profitable opportunity in the future. The implication is that past mortgage-related choices influence the borrower's prepayment decision.

With loan-level data it is possible to observe the purpose for which the *current* mortgage was contracted.[15] Including dummy variables for standard refinancing, home purchase, equity refinancing, and employer-sponsored relocation allows the model to control for the effects of transaction costs, and to account for the borrower's previous mortgage decisions. It might be that, due to transaction costs, a borrower who recently purchased his home will display different behavior than a

[15]Note that this does not reveal whether an observed prepayment on the current mortgage is due to property sale or refinancing.

borrower that has owned his home for several years but recently refinanced. It is also useful to distinguish between a cash-out refinancing and a standard refinancing. A cash-out refinancing occurs when a homeowner borrows from the equity in his property by increasing his loan balance. This may be an indication of an income constraint, or may reduce the refinancing opportunities available to the borrower in the future.

Figure 1.8 shows the empirical hazard rate functions stratified by mortgage purpose. Visual inspection reveals that these curves are indeed different. The Wilcoxon test of equality over strata rejects the null hypothesis that all hazard rates are identical. There are similarities, however, between the two types of refinancing and the two types of purchase. A more refined Wilcoxon test also rejects the null hypothesis that the hazard rate function for home purchase is the same as that for relocation, but it does not reject the null hypothesis that the hazard rate function for equity refinancing is the same as that for a standard refinancing. The results in the following section show that the differences observed in Figure 1.8 translate into significantly different predictions in prepayment rates for borrowers with different mortgage purposes.

1.4 Results

Each of the three models discussed above — the Cox proportional hazard rate model, the log-logistic proportional hazard rate model, and the logit random utility model — are estimated using the full sample of fixed-rate mortgages.

Comparisons are made on the basis of the coefficient estimates, elasticities for continuous variables, and hazard rate forecasts. Given the theory underlying these models, we would expect the results to be consistent across the three models, but this is not the case. The log-logistic proportional hazard rate model and logit random utility model make similar predictions, but the results of the Cox proportional hazard rate model are quite different, and in some cases are counter to economic intuition. In the log-logistic model, loan age and the one-month lag of the coupon gap are the primary drivers for prepayment. In the logit model, effects are balanced across the continuous variables, making this model a better tool for capturing borrower heterogeneity. The results in this section lead to the conclusion that the logit model has better forecasting power than either of the proportional hazard rate models.

1.4.1 Covariate Effects

The coefficient estimates for each model are shown in Table 1.6. Recall that, in estimating a Cox proportional hazard rate model, the baseline hazard rate function is factored out of the model. No explicit form for the baseline hazard rate function has been assumed for the Cox model. Hazard rate elasticities, reported in Table 1.7, are used to facilitate the comparison between the three models. Since the elasticity formulas depend on the level of the covariates and, in the case of the logit model, the forecasted hazard rate, elasticities are calculated individually for each observation. The values in the table are the averages over all observations.

The variable that has the strongest effect on the probability of prepayment in all three models is the coupon gap. In the log-logistic and logit models, the estimates are consistent with economic intuition; the coefficient on the level of the one-month lag of the coupon gap is positive, indicating that prepayment is more likely as the refinance incentive increases. The negative coefficient on the cube of the one-month lag of the coupon gap implies that the effect of the refinance incentive levels off as the coupon gap continues to increase. This combined effect leads to an elasticity of 4.18 under the log-logistic model and 2.52 under the logit model. This is a significant difference that would lead to deviations in the mortgage prices derived under the two models. In both models, the effect of the three-month lag of the coupon gap is positive, but weaker than that of the one-month lag. This is consistent with the belief that transaction costs have fallen due to increased competition among originators and innovations in online lending.

The Cox model also predicts a positive relationship between the coupon gap and the probability of prepayment, but this relationship exists only for the three-month lag of the coupon gap. The elasticity with respect to the one-month lag is actually negative. This contradicts the hypothesis that transaction costs, measured as the time between making the prepayment decision and closing the loan, have decreased. Contradictory effects are also estimated with respect to the slope of the yield curve. The log-logistic and logit models predict positive elasticities with respect to the yield curve slope. This is reasonable because an increase in the slope of the yield curve is caused by an increase in long term rates versus short term rates. In other words, there is an expectation that interest rates will rise, and borrowers will refinance to lock-in low rates. However, under the Cox model this elasticity has the same magnitude but is negative. This implies that as the slope of the yield curve increases prepayment rates decline, a puzzling result.

Differences between the Cox model, and the log-logistic and logit models are also observed with respect to state productivity. The assumption is that as state productivity increases greater economic activity leads to more frequent property sales. This is supported by positive elasticities calculated under the log-logistic and logit models.[16] Under the Cox model, however, this effect is negative, which can also be explained by a feasible argument. When state productivity declines, higher unemployment causes workers to move out of the state, leading to an increase in property sales.

The Cox model also predicts effects that are contrary to those of the log-logistic and logit models for mortgage term, seasonality, and mortgage purpose. The log-logistic and logit models predict that prepayment rates are higher among lower term mortgages and mortgages contracted for a standard refinancing. The Cox model, on the other hand, predicts higher prepayment rates for 30-year mortgages and mortgages contracted for property sale. The log-logistic and logit models confirm the hypothesis that prepayment is higher during the spring and summer months, while the Cox model predicts more frequent prepayment between October and December.

Most studies that look solely at prepayment do not consider the effects of

[16] The elasticity under the log-logistic model is low.

housing prices. However, this factor does play an important role in the prepayment decision through housing equity and affordability. Although housing price appreciation, as measured by the OFHEO HPI, is not a significant factor under the Cox model and is weak under the log-logistic model, the logit model predicts that prepayment rates rise as housing prices increase and borrowers have access to greater home equity. All three models predict that prepayment is less frequent among borrowers that are equity constrained through high LTV ratios at origination, a previous equity refinancing, or an employer-sponsored relocation. The three models also agree with respect to the borrower's FICO score which, like home equity, can influence refinancing opportunities. For all models, the coefficients on the FICO score buckets, which are similar in magnitude across the three models, are increasingly negative as the FICO score decreases.

It is interesting to note that the three models agree with respect to credit quality, which does not differ greatly over the life of a prime mortgage, and differ substantially in regards to the refinance incentive, which is highly variable over time. This suggests that the source of the problem may be in the methods used to incorporate time-varying covariates in a Cox model. Since the log-logistic and logit models simply exploit the natural structure of the data, it appears that the integrity of the Cox model is not maintained under time-varying covariates. This conclusion is supported by observed trends which align with the results of the log-logistic and logit models.

1.4.2 Goodness of Fit

In order to fully evaluate the models presented in this chapter, we would like to make comparisons based on a goodness of fit measure. One possibility is the likelihood ratio (LR) index, which is calculated by comparing the log-likelihood function for the fully specified model with the log-likelihood function for a model in which all the coefficients are equal to zero.

$$LRIndex = 1 - \frac{LL_\beta}{LL_0}$$

A low value for the likelihood ratio index implies that the fully specified model is not an improvement over the assumption of "no model". The likelihood ratio index is reported for each model in the fourth row of Table 1.6. Not surprisingly, the likelihood ratio index for the Cox model is very low. The likelihood ratio indices for the log-logistic and logit models are similar, at 0.17 and 0.15, respectively.

While it is useful to have a concise goodness of fit measure, the likelihood ratio index is just a statistical value, and does not yield sufficient information on the forecasting ability of the models. A true test of predictive power is how closely the forecasted hazard rates match the observed prepayment probabilities. Figure 1.9 plots the average forecasted hazard rates for the log-logistic and logit models, and the empirical hazard rates calculated using the Kaplan-Meier ratio. These values are plotted against the age of the loan to underscore the time dependence inherent in each model. All curves are calculated by averaging over observations. The Cox model is omitted from this comparison because hazard rate predictions are not readily available from the partial likelihood estimation used for this model.[17]

Despite the similarities between the log-logistic model and the logit model, the figure shows that the the logit model more accurately predicts observed trends than the log-logistic model. There is virtually no difference between the predictions of the logit model and the empirical hazard rates below 40 months, the period over which prepayment activity is highest. In contrast, the values estimated by the log-logistic model are as much as eight percentage points higher than the empirical rates. This is a surprising result because it is generally believed that proportional hazard rate models are interchangeable with logit models. This figure shows, however, that forecasts of borrower behavior can be improved substantially by using a random utility framework that explicitly links the borrower's decision process to utility maximization.

One of the reasons the log-logistic model performs poorly is the constraint placed by the baseline hazard rate function, shown in the bottom panel of Figure 1.9.[18] With a shape factor of 1.3, the peak of the baseline hazard rate is at 15 months, which captures only the first hump of the empirical hazard rate function. In order to optimize the overall fit, which includes high prepayment rates in months 30 to 40, the forecast of the baseline hazard rate is too high overall.

After 40 months the log-logistic model is a better fit than the logit model, because the baseline hazard rate function under the logit model is monotonic by

[17]The baseline hazard rate could be estimated non-parametrically to generate comparable hazard rate predictions, but that has not been done in this chapter.

[18]Note that the scale of the baseline hazard rate functions is lower than we would expect, because intercept terms have been included in order to make the curves comparable.

definition, and cannot capture the burnout effect observed under the log-logistic model. This failure can be remedied by allowing age to enter the logit specification in a piecewise manner, so that separate coefficients are estimated for each region of the baseline hazard rate function.

1.5 Conclusion

This chapter presents two alternatives to the Cox proportional hazard rate model for forecasting prepayment rates of residential mortgages. The focus is on accurately estimating the effects of loan characteristics and economic variables within the context of a model that incorporates time-varying covariates. The three models compared in this chapter are the Cox proportional hazard rate model, a log-logistic proportional hazard rate model with time-varying covariates, and a panel data random utility model.

We find that the Cox model performs rather poorly, making predictions with respect to key factors, such as the refinance incentive, that are contradictory to those forecasted by the log-logistic and logit models. This is a surprising result since use of the Cox model is prevalent, and theoretical arguments indicate a strong similarity between the three models. It appears that, at least for the application to mortgage prepayment, the Cox model does not properly capture the dynamics of time-varying covariates, and the results of previous studies using this model must be questioned.

While the log-logistic and logit models produce similar coefficient and elasticity estimates, only the logit model forecasts hazard rates that are in line with the observed behavior. In addition, the elasticities calculated under the logit model are balanced across the continuous variables, allowing the model to capture a higher degree of heterogeneity among borrowers. In the log-logistic model, most of the explanatory power is attributed to the coupon gap and the baseline hazard rate function.

This chapter shows that the logit random utility model is a superior framework for analyzing mortgage prepayment. By using the natural structure of mortgage data, it is straightforward to incorporate time-varying covariates and account for censoring. The model properly accounts for time trends and borrower heterogeneity, forecasting accurate prepayment rates that can easily be incorporated into an out-of-sample valuation framework. In addition, a random utility model presents an alternative structural framework with sufficient flexibility to address advanced modeling issues such as competing risks and unobserved heterogeneity, which are integral to an accurate representation of the borrower's behavior and will be discussed in the following chapters.

Table 1.1: Distribution of Loans by Issue Year, Purpose, and Term

Issue Year	2000	2001	2002	2003	2004	2005	All
	Percent of Total Sample (103,347 Loans)						
	13.0	30.9	17.7	20.5	5.1	12.8	100.0
	Purpose at Origination - Percent by Issue Year						
Standard Refinancing	12.7	39.3	50.3	68.9	50.9	21.9	42.2
Cash-Out Refinancing	7.1	20.4	21.0	20.8	23.0	33.3	20.7
Property Purchase	55.4	31.0	17.9	9.7	22.4	39.7	28.2
Employer Relocation	24.8	9.2	10.8	0.6	3.7	5.1	8.9
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	Term at Origination - Percent by Issue Year						
10-Year	0.07	0.06	0.19	0.29	0.29	0.02	0.14
15-Year	10.01	12.92	30.74	58.70	54.61	14.84	27.45
20-Year	0.25	0.50	0.88	0.54	0.10	0.02	0.46
30-Year	89.27	85.76	67.59	39.81	44.32	84.96	71.36
Other	0.40	0.76	0.60	0.66	0.68	0.16	0.59
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0

This table shows the distribution of loans by issue year, purpose at origination, and term. Only a small percent of the fixed-rate mortgages considered in the analysis were issued in 2004 compared to a more even distribution over the other years. Of the loans issued in 2000 over 75% were contracted for property purchase - either through a normal purchase or an employer-sponsored relocation. In contrast, in 2003 nearly 70% of the loans were contract for a standard refinancing, and over 20% were contracted for a cash-out refinancing. Prior to 2003 the majority of mortgages issued had a term of 30 years. 15-year loans gained in popularity over 2003 and 2004. Borrowers returned to 30-year mortgages in 2005.

Table 1.2: Summary Loan Statistics by Issue Year

	Mean	Std. Dev.	Percentiles 25	50	75
Coupon Rate at Origination (%)					
2000	8.08	0.45	7.75	8.13	8.38
2001	7.17	0.44	6.88	7.13	7.50
2002	6.51	0.42	6.25	6.50	6.88
2003	5.43	0.45	5.00	5.50	5.75
2004	5.46	0.39	5.13	5.50	5.75
2005	5.71	0.32	5.63	5.75	5.88
All	6.47	1.03	5.63	6.5	7.25
Property Appraisal at Origination ($)					
2000	520,738	267,535	379,900	452,000	599,900
2001	652,631	359,626	449,000	555,000	735,000
2002	766,351	485,666	500,000	640,000	859,000
2003	935,872	592,247	600,000	775,000	1,100,000
2004	938,365	652,196	575,000	750,000	1,100,000
2005	844,908	531,176	562,500	712,000	951,120
All	769022	504974	483,795	635,000	875,000
Loan To Value Ratio at Origination (%)					
2000	75	13	70	80	80
2001	70	14	62	74	80
2002	65	16	55	68	79
2003	59	16	48	61	72
2004	61	16	49	64	75
2005	67	15	59	71	80
All	66	16	55	69	80
Loan Balance at Origination ($)					
All	450,345	165,491	350,000	410,000	510,000
FICO Score at Origination					
All	729	48	697	738	769

On average, the loans in the dataset follow the characteristics of jumbo loans. The average balance is $454,000 and 95% of the loans have a LTV ratio below 80%. Since these loans are selected for securitization, they display different trends than the overall mortgage market. Most of the borrowers have excellent credit with 65% having a FICO score above 720, and just 1% with a FICO score below 620. Over time the dispersion of the LTV ratio at origination increases, but in the direction of lower LTV ratios. LTV ratios are lowest for loans issued in 2003. The average appraisal value at origination is high at nearly $800,000 but ranges from $24,000 to over $16 million. 80% of the loans are in the range of $300,000 to $1.5 million.

Table 1.3: Region Definitions

REGION	STATES
CA-HI	CA, HI
NY-NJ-CT	NY, NJ, CT
Midatlantic	DC, MD, VA, DE
Texas	TX
Florida	FL
South	LA, MS, AL, GA, SC, NC, TN, AR, KY, WV
Lakes	IL, PA, OH, WI, MN, MI
Mountain	AZ, CO, ID, MT, NM, NV, UT
Northwest	OR, WA, AK
Plains/Midwest	ND, SD, IN, IA, NE, KS, OK, MO
New England	MA, ME, NH, RI, VT

This table lists the region definitions used in the tables and figures that follow. States have been grouped according to regional proximity and economic similarities. States with high representation in the sample have been given their own category.

Table 1.4: Distribution of Loans by Region and Issue Year (Percent of Total Sample)

	2000	2001	2002	2003	2004	2005	Total
CA-HI	3.43	12.38	7.38	10.49	2.41	5.07	41.16
NY-NJ-CT	2.12	3.63	2.32	2.49	0.55	1.36	12.47
Midatlantic	1.26	3.02	1.61	1.41	0.42	1.49	9.21
Texas	0.91	1.31	0.72	0.61	0.13	0.38	4.06
Florida	0.44	0.90	0.47	0.48	0.15	0.37	2.81
South	0.98	1.58	0.86	0.79	0.18	0.73	5.12
Lakes	1.76	3.56	1.96	2.01	0.51	1.25	11.05
Mountain	0.99	1.94	0.94	0.93	0.26	0.82	5.88
Northwest	0.37	0.81	0.41	0.46	0.17	0.52	2.74
Plains/Midwest	0.33	0.71	0.38	0.44	0.11	0.20	2.17
New England	0.52	1.07	0.70	0.52	0.13	0.41	3.35
Total	13.11	30.91	17.75	20.63	5.02	12.60	100.0

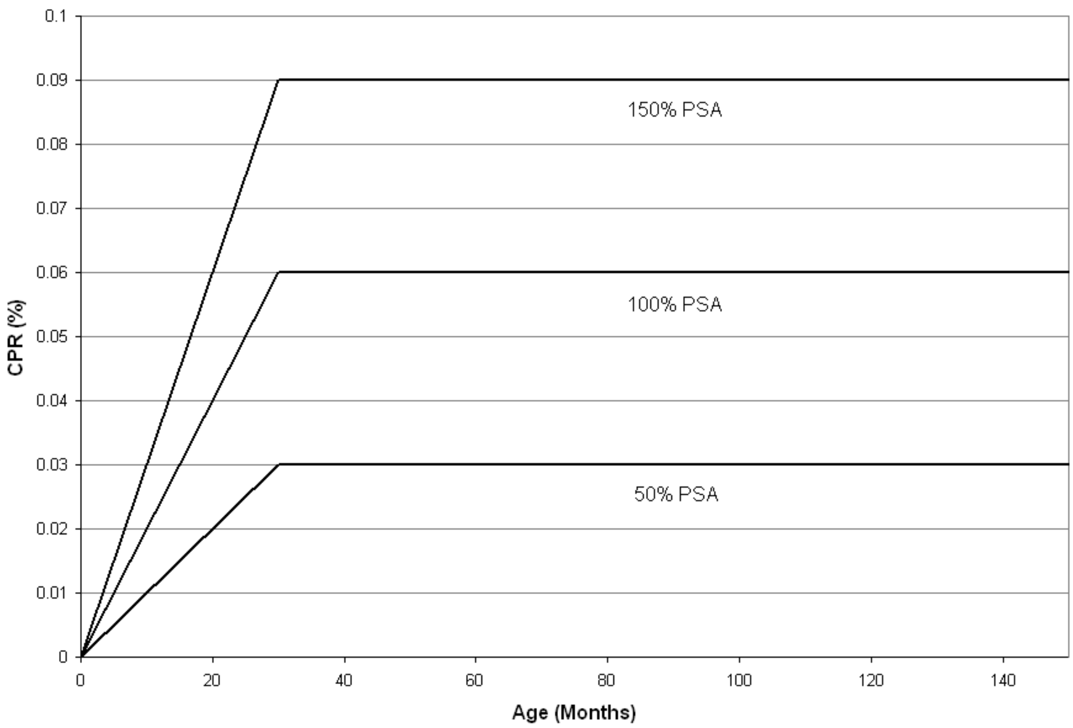
Since all loans are issued by Wells Fargo, one could expect regional concentration to be high in states where Wells Fargo has a strong presence and in areas with high housing prices and hence a large proportion of jumbo loans. Nearly half of the loans come from California, but the dataset contains loans for properties in all 50 states and the District of Columbia. States with a strong representation are New York, New Jersey, Maryland, Virginia, Florida, Minnesota, Illinois, and Texas.

Table 1.5: Summary Statistics by Region

	Contract Rate at Origination		Property Appraisal at Origination		LTV Ratio at Origination		Loan Balance at Origination		FICO Score at Origination	
	Mean	St.D.	Mean	St.D.	Mean	St.D.	Mean	St.D.	Mean	St.D.
CA-HI	6.46	0.98	884,926	594,171	61.97	16.26	483,936	184,086	734.46	45.49
NY-NJ-CT	6.68	1.03	762,506	473,280	66.02	15.87	455,431	171,295	730.25	49.51
Midatlantic	6.66	0.96	672,669	336,164	70.94	12.93	448,393	152,407	732.69	49.07
Texas	6.78	1.05	672,350	364,160	71.28	13.04	449,620	171,836	725.67	50.68
Florida	6.69	1.04	710,579	486,742	68.00	15.72	433,723	174,002	722.5	49.82
South	6.67	1.01	638,265	370,240	72.15	13.00	430,413	158,676	726.41	49.37
Lakes	6.67	1.06	689,824	406,412	69.56	13.63	444,687	159,842	730.85	47.97
Mountain	6.74	1.02	668,205	376,444	69.91	13.92	434,685	164,287	728.37	48.65
Northwest	6.57	1.01	692,609	423,860	69.41	14.28	441,153	164,743	730.87	48.34
Plains/Midwest	6.62	1.06	640,580	307,914	71.59	11.83	436,690	158,292	733.26	47.37
New England	6.69	0.98	736,414	421,654	66.45	15.15	446,954	159,503	735.91	47.30

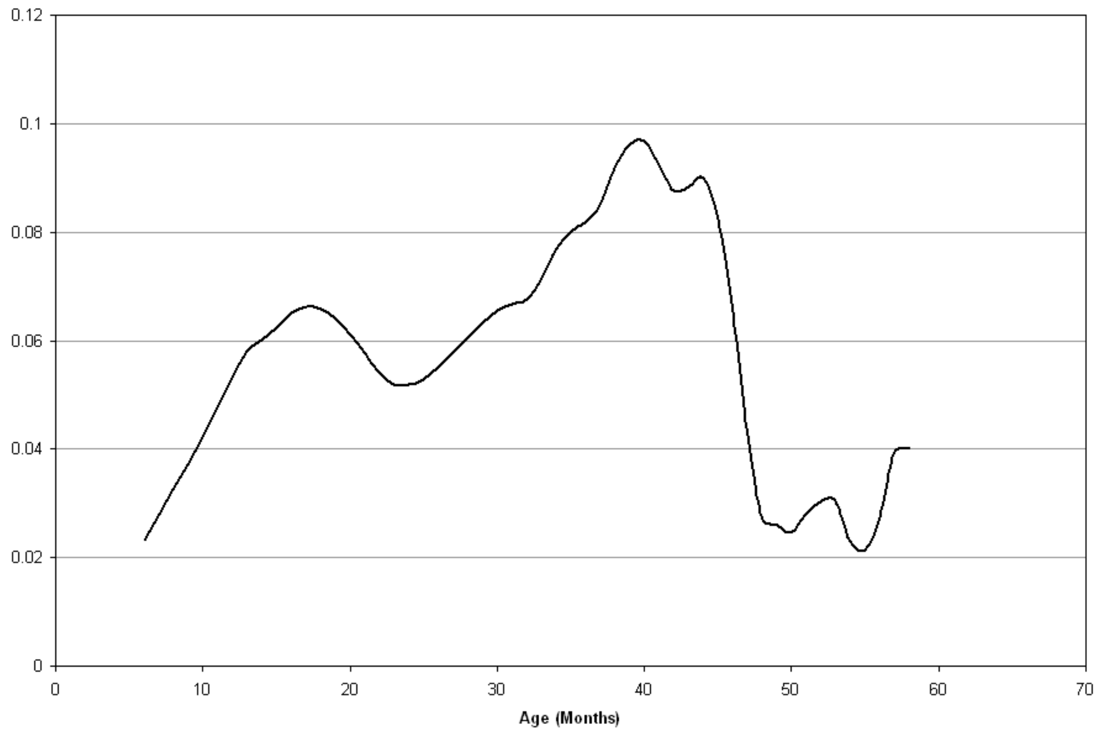
This table displays summary statistics by region. In general, loan characteristics are uniform across regions with trends in property values at origination following the OFHEO Housing Price Index (HPI).

Figure 1.1: The PSA Standard Prepayment Model



The PSA benchmark model (also called "100% PSA") assumes that the prepayment rate starts at a 0.2% constant prepayment rate (CPR) in the first month following origination of the mortgage loan (not the pool) and increases 0.2% per month in each succeeding month. The prepayment rate is assumed to level off at 6% in month 30 and beyond. The CPR is the annualized equivalent of the single monthly mortality (SMM), which measures the fraction of the beginning-month pool balance that prepays during the month.

Figure 1.2: Empirical Prepayment Function - Kaplan-Meier Ratio by Age



This figure shows the empirical hazard rate function observed in the data. The curve is derived by applying the Ramlau-Hansen kernel smoothing technique to the Kaplan-Meier hazard ratios calculated at each month. The figure illustrates the seasoning behavior we expect with residential mortgages. The hazard rate increases with age, peaking at around 40 months, and then falls sharply. This decline in prepayment rates for seasoned loans, which is not capture by the PSA curves, is known as "burnout". It is in order to capture this effect that a log-logistic parametric hazard rate is preferred. However, for this sample we also observe a dip at around 24 months indicating that it may be necessary to allow for greater flexibility in the relationship between prepayment rates and loan age. Figure 1.3 shows that the same trends are displayed at the regional level. Prepayment rates are slightly lower in California, the New York metropolitan area, and the South. The greatest deviations from the national trends are in the Great Lakes region and the Plains/Midwest region. Prepayment rates for the Great Lakes region fail to peak and are more stable, almost following the PSA convention. On the other hand, prepayment rates for the Plains/Midwest region are more volatile than for any other region.

Figure 1.3: Kaplan-Meier Ratio by Age and Region

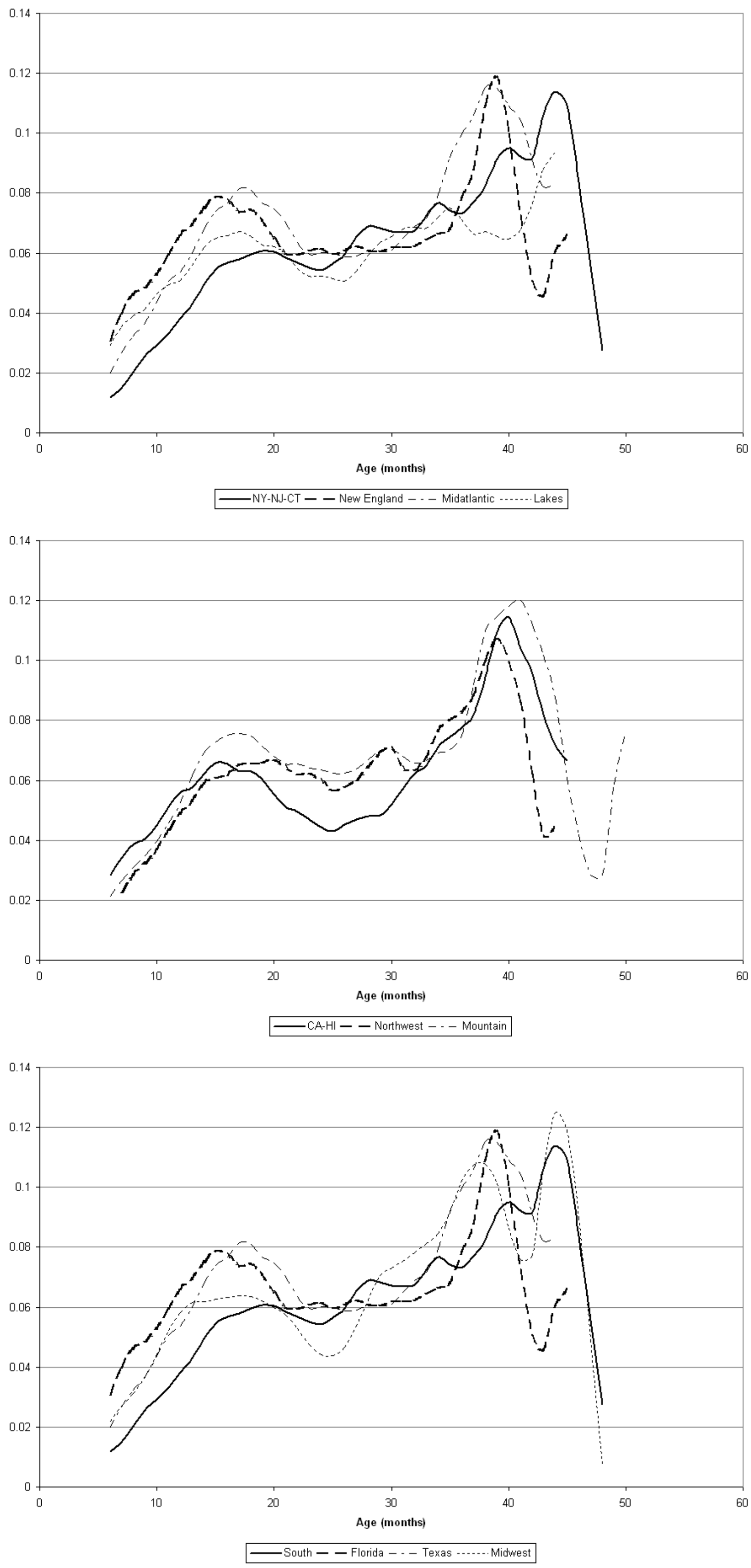
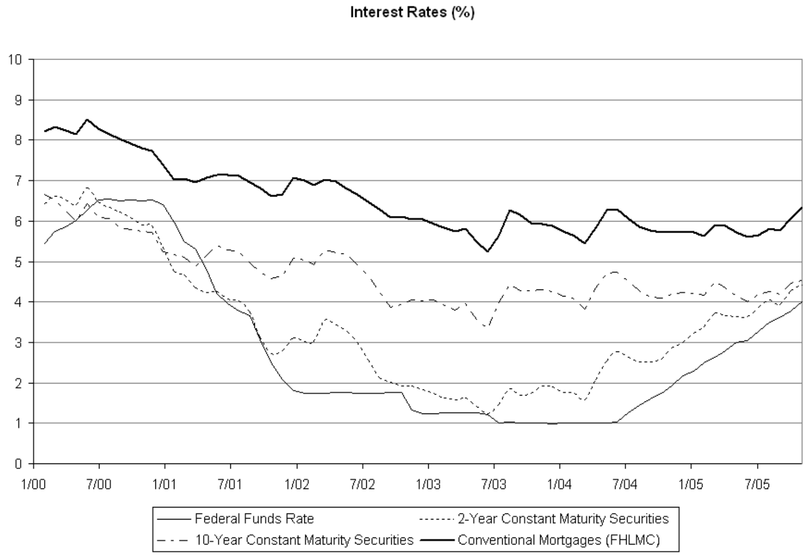


Figure 1.4: Historical Stock Market and Interest Rate Trends



The beginning of this period is marked by a sharp decline in stock market prices in April 2000, propelled by the failure of many of the internet and technology ventures that grew in the 1990s. This was followed by an increase in the unemployment rate particularly in the West Coast, New England, and the Mid-Atlantic where many of these ventures were located. In order to stimulate recovery, the Federal Reserve began easing monetary conditions in January 2001, and this continued until June 2004, when the federal funds rate reached a low of 1%. Growth was slow and recovery was not apparent in many regions until late in 2002. The strength of the economy over this period was largely due to the growth in housing prices. By 2000 housing prices were already strong as a result of the expansion stimulated by the growth of the technology sector. When stock market prices declined, housing prices did not follow.

Figure 1.5: Historical Unemployment Rates (%) by Region

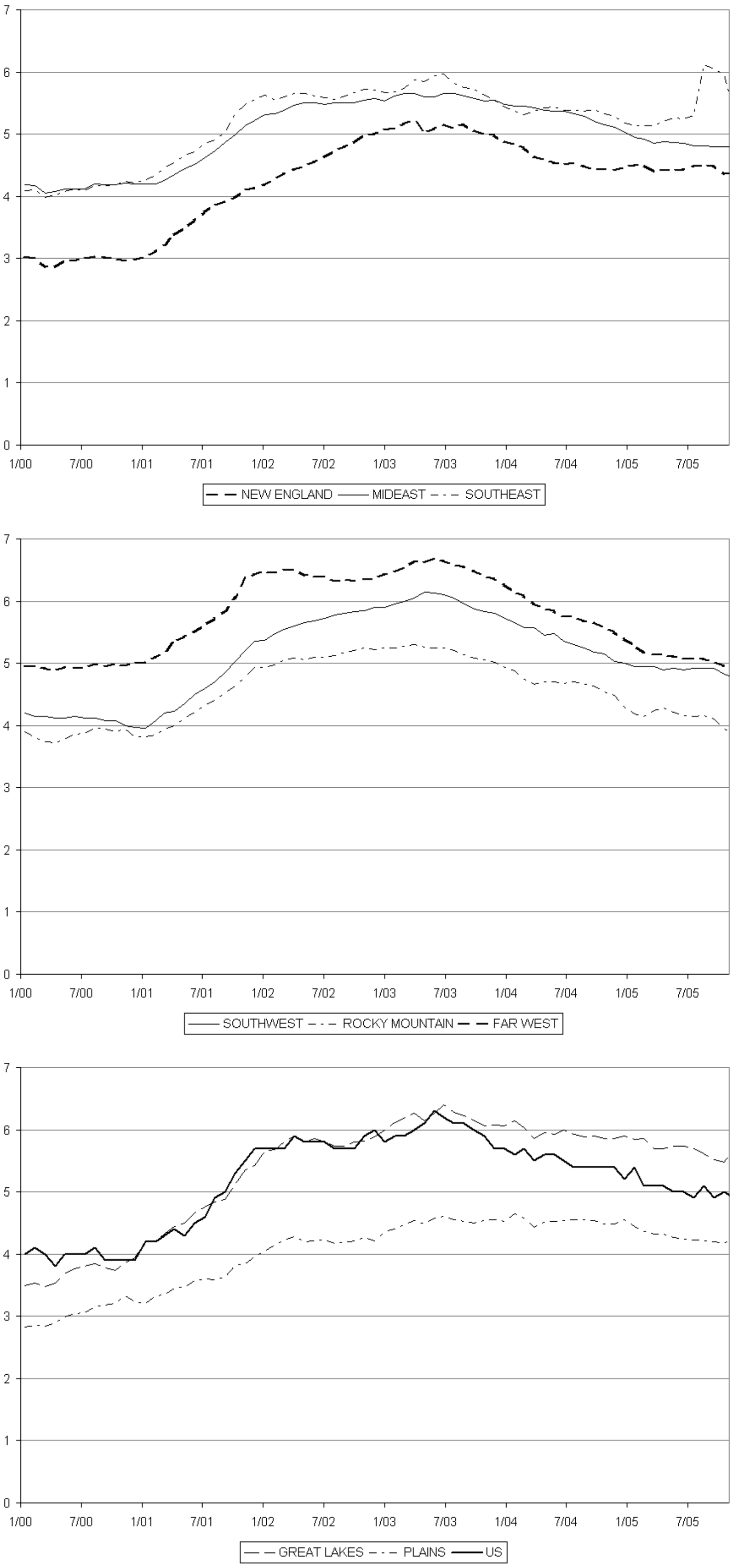


Figure 1.6: Annual Growth Rate (%) in Real GSP by Region

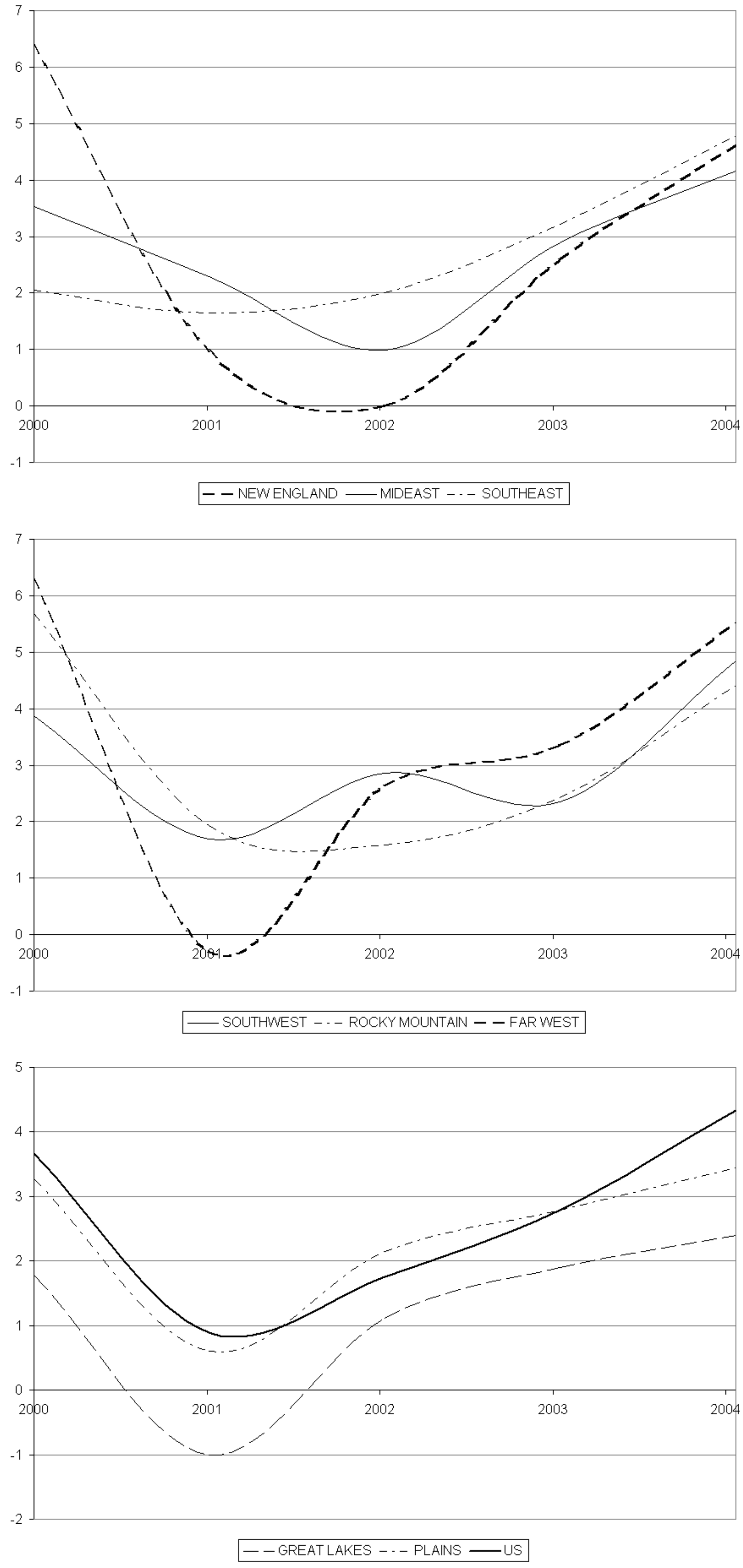


Figure 1.7: Annual Growth Rate (%) in the OFHEO HPI by Region

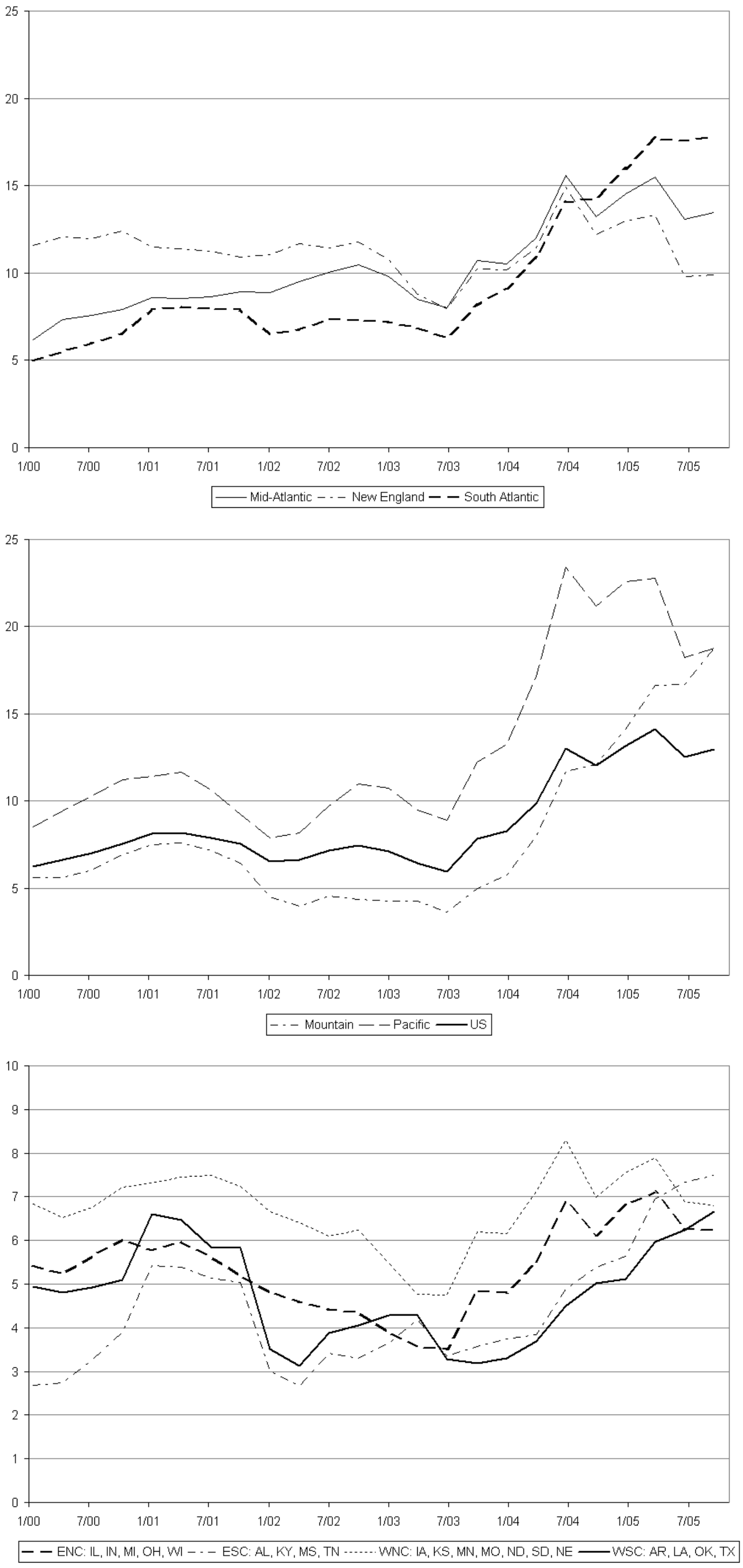
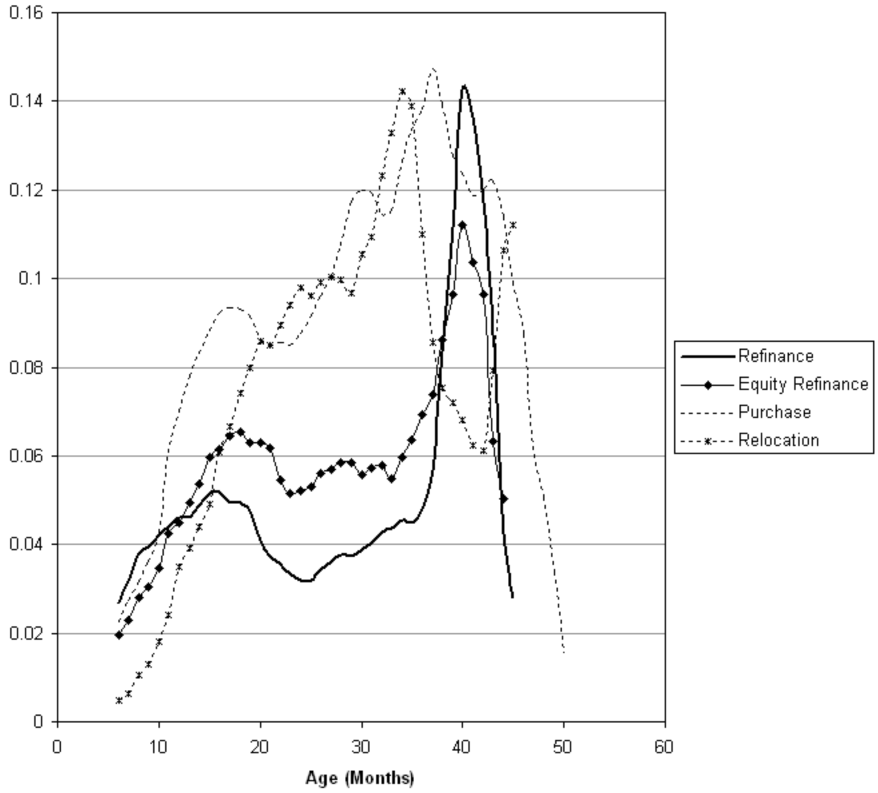


Figure 1.8: Empirical Prepayment Function - Kaplan-Meier Ratio by Purpose



Wilcoxon Test for Equality Over Strata

Test 1: h(t|Refinance) = h(t|Equity Refinance) = h(t|Purchase) = h(t|Relocation)
test statistic: 345.60 Chi-Sq. Prob.: <.0001

Test 2: h(t|Purchase) = h(t|Relocation)
test statistic: 338.71 Chi-Sq. Prob.: <.0001

Test 3: h(t|Refinance) = h(t|Equity Refinance)
test statistic: 0.3395 Chi-Sq. Prob.: 0.5601

Table 1.6: Coefficient Estimates

	Cox Model		Log-logistic		Logit Model	
Number of Loans	103,347					
Number of Observations	1,727,742					
Log-likelihood	-62,441		-188,519		-237,906	
LR Index	0.03		0.17		0.15	
γ	-		0.022	***	-	
			(0.001)			
p	-		1.322	***	-	
			(0.007)			
Intercept	-		-4.393	***	-9.174	***
			(0.046)		(0.046)	
Age (Months)	-		-		0.048	***
					(0.001)	
LTV at	-0.005	***	-0.005	***	-0.006	***
Origination	(0.001)		(0.000)		(0.000)	
Mortgage Term	-0.175	***	0.391	***	0.316	***
$<$ 30 years	(0.030)		(0.010)		(0.011)	
Coupon Gap	-0.898	***	2.143	***	2.287	***
(1-month lag)	(0.048)		(0.020)		(0.021)	
Coup. Gap	0.010	***	-0.047	***	-0.055	***
Cubed (1-month lag)	(0.002)		(0.001)		(0.001)	
Coupon Gap	1.231	***	0.219	***	0.046	***
(3-month lag)	(0.042)		(0.011)		(0.013)	
Slope of Yield	-0.501	***	0.695	***	0.586	***
Curve (1-month lag)	(0.019)		(0.008)		(0.008)	
HPI	0.005		0.055	***	0.023	***
Appreciation	(0.008)		(0.003)		(0.003)	
Baseline FICO: 850 - 750						
FICO: 750 - 720	-0.045		-0.051	***	-0.053	***
	(0.034)		(0.010)		(0.011)	
FICO: 720 - 660	-0.159	***	-0.175	***	-0.185	***
	(0.031)		(0.009)		(0.010)	
FICO: 660 - 620	-0.433	***	-0.371	***	-0.409	***
	(0.058)		(0.015)		(0.017)	
FICO: 620 - 300	-0.615	***	-0.581	***	-0.660	***
	(0.095)		(0.024)		(0.026)	

continued on next page

Standard errors are reported in parenthesis.
*** 1% Significance; ** 5% Significance; * 10% Significance

Table 1.6: Coefficients Estimates *(continued)*

	Cox Model		**Log-logistic**		**Logit Model**	
Baseline: Jan - Mar						
Apr - Jun	-0.103	**	0.165	***	0.163	***
	(0.047)		(0.011)		(0.012)	
Jul - Sep	0.076	*	0.160	***	0.125	***
	(0.046)		(0.011)		(0.012)	
Oct - Dec	1.002	***	-0.022	**	-0.015	
	(0.039)		(0.011)		(0.012)	
State Productivity	-5.385	***	5.965	***	1.140	***
	(0.311)		(0.197)		(0.220)	
Baseline: Standard Refinancing						
Equity Refinancing	-0.122	***	-0.170	***	-0.187	***
	(0.034)		(0.011)		(0.012)	
Purchase	0.066	**	-0.067	***	-0.077	***
	(0.034)		(0.010)		(0.011)	
Relocation	-0.437	***	-0.055	***	-0.327	***
	(0.050)		(0.013)			

Standard errors are reported in parenthesis.
*** 1% Significance; ** 5% Significance; * 10% Significance

Table 1.7: Average Hazard Rate Elasticities

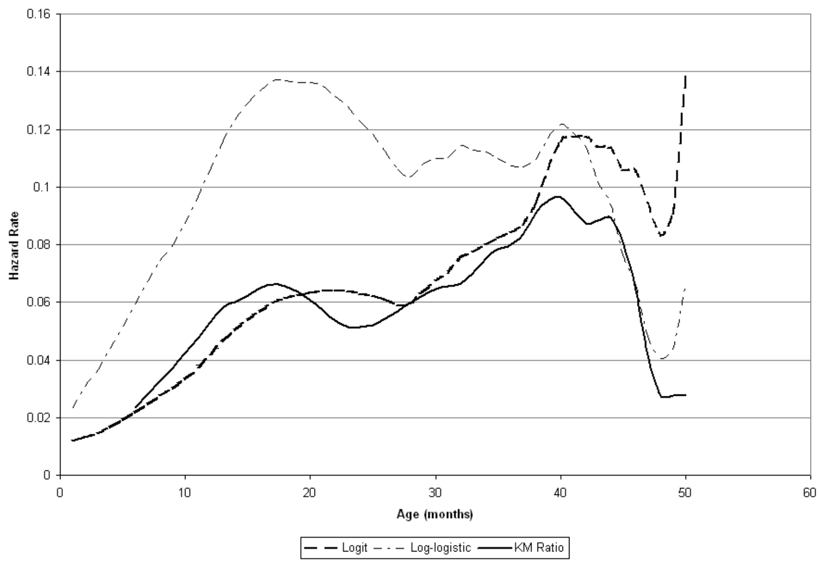
	Cox Model	Log-logistic	Logit Model
Age	-	0.547	0.109
LTV at Origination	-0.348	-0.358	-0.361
Coupon Gap (1-month lag)	-1.436	4.180	2.524
Coupong Gap (3-month lag)	2.414	0.085	0.430
Slope of Yield Curve	-0.780	0.867	1.08
HPI Appreciation	0.014	0.064	0.157
State Productivity	-0.214	0.044	0.238

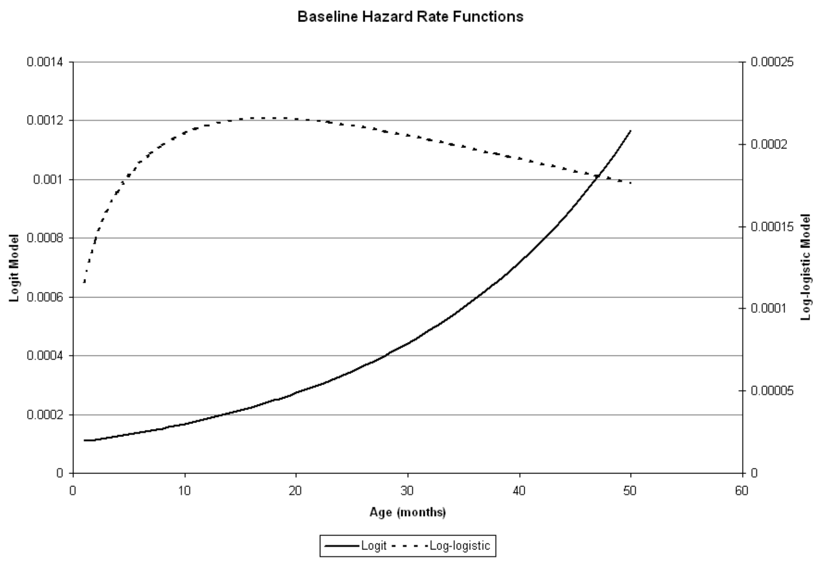
Each of the three models discussed above — the Cox proportional hazard rate model, the log-logistic proportional hazard rate model, and the logit random utility model — are estimated using the full sample of fixed-rate mortgages. Comparisons are made based on the coefficient estimates, elasticities for continuous variables, and hazard rate forecasts. The elasticities reported in this table and the curves shown in Figure 1.9 are calculated by taking the average over all observations in the dataset.

The log-logistic proportional hazard rate model and logit random utility models predict similar effects, but, surprisingly, the results of the Cox proportional hazard rate model are quite different. In the log-logistic model, loan age and the one-month lag of the coupon gap are the primary drivers for prepayment. In the logit model, the elasticities are more uniform across the continuous variables. This gives the model greater flexibility to capture borrower heterogeneity, and thus the hazard rate forecasts from the logit model are closer to the empirical hazard rate function.

The accuracy of the hazard rate function predicted by the log-logistic model is diminished because of the constraint placed by the baseline hazard rate. With a shape factor of 1.3 the peak of the baseline hazard rate is at 15 months which captures only the first hump of the empirical hazard rate function. After 40 months the log-logistic model is a better fit. This is because the baseline hazard rate function under the logit model is monotonic by definition, and therefore cannot capture the burnout effect we observe. Note that the scale of the baseline hazard rate functions do not follow our intuition because the intercept terms have been included in order to make the curves comparable.

Figure 1.9: Forecasted vs. Empirical Hazard Rates





Chapter 2

Focus on Competing Risks

2.1 Introduction

Recent studies of residential mortgages (Deng, Quigley, and Van Order (2000) and Downing, Stanton, and Wallace (2005)) have extended their analysis beyond prepayment to include default. They have found that, although default is a rare event among prime mortgages, it is important through its correlation with prepayment. This chapter shows that it is also necessary to consider non-termination events, such as partial prepayment (curtailment) and delinquency, in order to accurately predict borrower behavior and the corresponding cash flows that enter the valuation of mortgage-backed securities.

This chapter examines loan-level payment trends for a large set of non-conforming, prime mortgages issued and securitized between 2000 and 2005. A key feature of this data is the degree to which borrowers deviate from their amortized payment schedule, even prior to prepayment or default. As the status of the mortgage changes, so does the borrower's decision process. For example, a borrower that has chosen to curtail his mortgage will have a different attitude towards missing a payment than one who is already delinquent. The choice made by the borrower is influenced by the other alternatives available to him in that period, also known as the competing risks, and the utility obtained from each alternative. To accurately model mortgage borrower behavior it is necessary to employ a model that considers an expanded set of mortgage choices, conditions the borrower's utility on past decisions, and allows for correlation among the competing risks. This chapter considers two models that meet these requirements: a multinomial logit (MNL) model along the lines of that used in the subprime literature, and an ordered logit (OLG) model that has not been previously used with respect to mortgages.[1]

In basic hazard rate models, competing risks are estimated independently, with alternative choices being treated as censored events. By contrast, in contingent claims models the borrower is assumed to hold a joint termination option that encompasses both prepayment and default. Han and Hausman (1990) propose a hazard rate model that allows for correlated competing risks,[2] thereby addressing one of the weaknesses of reduced-form models relative to structural contingent claims models. This model takes advantage of the link between hazard rate models and discrete choice models discussed in the previous chapter.

The Han and Hausman (1990) model is fundamentally a binomial probit model, a type of discrete choice model. In many ways, discrete choice models are superior to hazard rate models because they offer greater flexibility for in-

[1]Ordered logit models are common in the analysis of survey data and in transportation science. See, for example, Bhat (1999), Daly and Bierlaire (2006), and Wang and Kockelman (2005).

[2]An application of this model can be found in Deng, Quigley, and Van Order (2000).

corporating elements — such as time-varying covariates, correlated competing risks, and unobserved heterogeneity — required to accurately characterize the borrower's decision process. An implicit use of this type of model is common in the literature on subprime mortgages (DeFranco (2002), Capozza and Thomson (2005), Pennington-Cross (2006)).

In the study of subprime mortgages, it is essential to account for delinquency in addition to prepayment and default. The Han and Hausman model is not adequate for this type of analysis. First, several of the events under consideration do not lead to termination of the mortgage, so a model based on the distribution of survival times is not appropriate. Second, the probit specification is difficult to implement with a large choice set because of the difficulties that arise in simulation through a multivariate normal distribution. For these reasons, a multinomial logit model is often used to study subprime mortgages. The multinomial logit model is based on a latent process that is abstracted from time, and allows the borrower's choice process to be conditioned on past behavior. This is in contrast to a hazard rate model, where the mortgage is assumed to be *current* at the beginning of each time period.

While a multinomial logit model that includes non-termination events has not previously been applied to prime mortgages, this type of analysis proves to be useful even for borrowers with good credit. In this chapter, it is shown that borrowers with prime mortgages frequently deviate from their amortization schedule through curtailment and delinquency. Consequently, a great deal can be learned from a model that considers non-termination events and conditions on past behavior. In addition, with a multinomial logit model it is possible to study delinquency separately from default. This is fundamental for prime mortgages because they rarely experience default, so it is necessary to model delinquency independently, perhaps as an income-driven event, rather than as a step towards default.

In this chapter, an ordered logit model is introduced to address several of the failings of the multinomial logit model. While the latter model provides a good starting point for tracking a borrower's behavior over time, it suffers from two important drawbacks. First, it displays the independence from irrelevant alternatives (IIA) property, which implies an independence between competing risks that is not appropriate for mortgages. Second, problems arise with the small event samples that are often observed with respect to delinquency and default. While the multinomial logit model has greater flexibility for capturing in-sample trends, the IIA property is shown to be a serious drawback, leading to predictions on borrower behavior that are counter to economic intuition.

The next section presents the details of both a multinomial logit model and an ordered logit model as applied to the analysis of mortgage borrower behavior. The dataset of approximately 100,000 non-conforming, prime, fixed-rate mortgages is discussed in section 2.3 with attention given to the observed payment trends. Section 2.4 compares the two models on the basis of estimated probability derivatives and forecasted transition matrices. This chapter concludes that an ordered logit model is the preferred approach because it accounts for correlation between mortgage events, handles small event samples, and provides an intuitive structural framework for characterizing the borrower's decision process.

2.2 Models

Most fixed-rate, residential mortgages are amortized by means of equal payment amounts due in each month of the contracted term. At every payment period, the borrower must decide whether or not to make the required payment, pay more than the scheduled amount in order to shorten the duration of the loan, or pay all of the remaining balance for a full prepayment. If the scheduled payment is made, the loan is said to be current. If a partial prepayment is made, the loan is considered curtailed. When a payment is missed the loan is delinquent.

Each month that a mortgage payment is missed the loan falls further into delinquency, and the creditor can begin foreclosure proceedings when the loan has been delinquent for more than 90 days. In practice, when a loan is in foreclosure, the borrower can make up the past due balance or even prepay the loan. Otherwise, the bank takes ownership of the property and issues a "real estate offering" (REO) to sell the property. Note that a 90-day delinquency means that three mortgage payments have been missed, not that the loan has been delinquent for three months. For example, a borrower may miss one payment and in the next month continue with the payment schedule, but not make up the missed payment. The borrower would then be 30 days delinquent for several months, rather than falling deeper into delinquency.

The key to modeling the full range of mortgage borrower behavior is to recognize that the utility space and choice set faced by the borrower change over time. For example, a borrower that is delinquent will have a different propensity towards making his mortgage payment than one who is current or curtailed. As a result, it is important to model non-termination events in addition to prepayment and default. This is the common practice with studies of subprime mortgages, but it is also essential among prime mortgages, where it is shown that delinquency is not as rare as commonly believed. A random utility model presents a convenient framework for this type of analysis. With a random utility model applied to panel data it is possible to encompass a large set of competing risks, track sequential choices, and condition the decision process on past behavior.

2.2.1 Multinomial Logit Model

In order to apply a random utility model to mortgage data, we employ a procedure known as "episode splitting". An episode is defined by a start time, an end time, an origination state, and a destination state, and each episode is accompanied by a set of covariates. For mortgages we normally define one episode per month of loan life, and a borrower is assumed to make a choice concerning his mortgage payment in each episode. Each choice is associated with a latent utility that is assumed to be a linear function of observed and unobserved factors, which include loan- and borrower-specific characteristics as well as economic variables. Factors can be time-varying, such as current interest rates, or constant over time, such as the term of the loan.

Since the model tracks a borrower's behavior over time, the utility function is conditioned on the status of the mortgage at the beginning of the period.

$$U_{ij,nt} = \beta_{ij} v_{nt} + \varepsilon_{ij,nt}$$

In this equation, n denotes an individual, i denotes an initial state, j denotes the destination state, and t represents a given time period. In many discrete choice applications, the deterministic component of the latent process contains choice-specific factors. In this model, as is common in mortgage analysis, the factors are assumed to be the same for all choices from a given initial state.

The borrower is assumed to select the choice in each period that is associated with the highest utility. The observed choice is denoted by an indicator variable, $y_{ij,nt}$, that is equal to one if choice j is selected by individual n in time period t, conditional on being in state i at the beginning of the period. In a static model, like the one used here, the optimal choice is made in each period independent of future periods.

Prepayment ($j = 0$), for example, is chosen when the utility obtained from this choice is greater than the utility obtained from any other choice.

$$y_{i0,nt} = 1 \Rightarrow U_{i0,nt} > U_{i1,nt} \wedge U_{i0,nt} > U_{i2,nt} \cdots$$

More generally, the probability of a transition from state i to state j is derived from the joint distribution of the error terms in the latent utility functions. In this specification, the errors are independent over individuals and over time. If we assume, also, that the errors follow an extreme value distribution and are independent over alternatives, the probability of each choice follows a logit model.

$$P_{ij,nt} = Prob(y_{ij,nt} = 1) = \frac{e^{\beta_{ij} v_{nt}}}{\sum\limits_{k=0}^{J} e^{\beta_{ik} v_{nt}}}$$

Since the mortgage events faced by a borrower in any given time period are mutually exclusive, it is important to examine the substitution patterns implied by the model. As indicated above, the multinomial logit model displays the IIA property, which implies that the ratio of any two event probabilities does not depend on any other alternative.

$$\frac{P_{ij,nt}}{P_{ik,nt}} = \frac{\frac{e^{\beta_{ij} v_{nt}}}{\sum\limits_{l=0}^{J} e^{\beta_{il} v_{nt}}}}{\frac{e^{\beta_{ik} v_{nt}}}{\sum\limits_{l=0}^{J} e^{\beta_{il} v_{nt}}}} = \frac{e^{\beta_{ij} v_{nt}}}{e^{\beta_{ik} v_{nt}}}$$

As an example, imagine that curtailment had not been previously permitted and is suddenly introduced as a potential choice. Borrowers who would like to shorten the duration on their mortgage will begin to make larger mortgage payments, and move away from maintaining their amortization schedule. Under the IIA property, the same proportion of borrowers is expected to move away from 30-day delinquency towards curtailment in order to maintain a constant probability ratio between staying current and entering 30-day delinquency. This is not, however, a reasonable conclusion because it is unlikely that borrowers who were previously unable to make their mortgage payments would, suddenly, find the excess income to curtail their mortgage. This demonstrates that a multinomial logit model may not be an appropriate framework for mortgage analysis because there is a strong correlation between alternatives that is not captured owing to the IIA property.

2.2.2 Ordered Logit Model

An alternative to the multinomial logit model is an ordered logit model. An ordered logit model is particularly well suited to the analysis of a borrower's behavior because all of the relevant mortgage events can be defined relative to the amortization schedule: prepayment is caused by making a payment equal to the remaining balance; curtailment is the result of a payment in excess of that dictated by the amortization schedule; maintaining the schedule requires a fixed payment in each period; entering delinquency is the result of missing a scheduled payment; and so on. As a result, the relevant mortgage events, or states, can be set on a continuum.[3] The states are ordered either by the payment required or by a natural time dependence.

Figure 2.1 illustrates this continuum. The events on either end — prepayment and liquidation — are termination events. To the right of prepayment, the events range from curtailment to delinquency. 60-day delinquency follows 30-day delinquency, and 90-day delinquency follows 60-day delinquency. Any loan that is more than 90 days delinquent, but not yet in foreclosure, is considered to be in 120-day delinquency. To the right of delinquency are the events as they are dictated by regulation: foreclosure, REO, and liquidation.

[3]This term is used throughout to convey fluidity in a borrower's choice stream. It is not used in the literal sense of a continuous scale since the choice set is discrete.

The goal is to model the behavior of the borrower as he moves along the continuum. A decision is made in every payment period, and each period is assigned an initial state and a termination state. The borrower can choose to stay in the same state, for example continue to make the scheduled payments, or he can choose to move to an alternative state. The borrower is not constrained to move to an adjacent state, but not all states are available from every initial state. When the ends of the continuum are reached, the mortgage terminates. From a curtailed or current state the borrower can choose to move to any point between prepayment and 30-day delinquency. It is not possible to reach 60-day delinquency during that payment period. Since the borrower can recover or even prepay from any interior state, the choice set expands as the borrower falls further into delinquency. From an initial state of 30-day delinquency the choice set expands to include 60-day delinquency, at 60-day delinquency the choice set includes 90-day delinquency, and at 90-day delinquency the choice set includes foreclosure. The full continuum is available to a borrower that begins the period in a foreclosed state.[4]

Given an initial state, i, all events available over that period are linked by a single latent utility process, $U_{i,nt}$, which can be thought of as a propensity for mortgage payment. The researcher observes a response variable, $y_{i,nt}$, that indicates which state on the continuum is chosen over that period. A choice towards the left of the continuum corresponds to a higher value of the latent process. For J possible choices, the latent process is divided into J contiguous

[4]In the empirical application, foreclosure is permitted from 60-day delinquency, as well, since foreclosure regulations vary across jurisdictions.

segments separated by $J-1$ unobserved thresholds. The threshold values are estimated along with the coefficients.

$$y_{i,nt} = j \quad if \quad k_{ij+1} < U_{i,nt} < k_{ij}, \quad j = 0...J-1, \quad k_{i0} = \infty, k_{iJ} = -\infty$$

As in the multinomial logit model, the latent utility is assumed to be a linear function of observed and unobserved factors.

$$U_{i,nt} = \beta_i v_{nt} + \varepsilon_{i,nt}$$

The key difference with the multinomial logit model is that in an ordered logit model there is a single set of coefficients that links all of the choices available from a given initial state. It is this dependence on a single utility process that induces the correlation among choices.

The unobserved component of the latent process is assumed to follow an independent extreme value distribution. In a single payment period, the probability of moving from state i to an interior state j is given by:

$$\begin{aligned}
P_{ij,nt} &= Prob(k_{ij+1} < U_{i,nt} < k_{ij}) \\
&= Prob(k_{ij+1} - \beta_i v_{nt} < \varepsilon_{i,nt} < k_{ij} - \beta_i v_{nt}) \\
&= \frac{\exp\left(k_{ij} - \beta_i v_{nt}\right)}{1 + \exp\left(k_{ij} - \beta_i v_{nt}\right)} - \frac{\exp\left(k_{ij+1} - \beta_i v_{nt}\right)}{1 + \exp\left(k_{ij+1} - \beta_i v_{nt}\right)}.
\end{aligned}$$

The probability of prepayment is equal to the probability in the right tail.

$$\begin{aligned}
P_{i1,nt} &= Prob(k_{i2} < U_{i,nt}) \\
&= Prob(k_{i2} - \beta_i v_{nt} < \varepsilon_{i,nt})
\end{aligned}$$

$$= 1 - \frac{\exp\left(k_{i2} - \beta_i v_{nt}\right)}{1 + \exp\left(k_{i2} - \beta_i v_{nt}\right)}$$

Similarly, the probability of the most delinquent event available from state i is given by the probability in the left tail.

$$P_{iJ,nt} = Prob(U_{i,nt} < k_{iJ})$$

$$= Prob(\varepsilon_{i,nt} < k_{iJ} - \beta_i v_{nt})$$

$$= \frac{\exp\left(k_{iJ} - \beta_i v_{nt}\right)}{1 + \exp\left(k_{iJ} - \beta_i v_{nt}\right)}$$

2.2.3 Estimation

Both models are estimated using maximum likelihood. The likelihood function is derived by aggregating over each borrower the probability of his observed choice stream. Because this is a static model, the probability of the choice stream is simply the product of the probabilities of the event observed in each period. By definition, the initial state in a given period is the choice selected in the previous period ($i_t = j_{t-1}$).

$$L(\theta|v) = \prod_{n=1}^{N} Prob(\vec{y}_n)$$

$$= \prod_{n=1}^{N} Prob(\{y_{i_1 j_1,n1}, y_{i_2 j_2,n2}, ...y_{i_T j_T,nT}\})$$

$$= \prod_{n=1}^{N} P_{i_1 j_1,n1} P_{i_2 j_2,n2} ... P_{i_T j_T,nT}$$

Since both models are based on a logit model, the likelihood function has a closed-form solution, and can be estimated using pre-programmed methods available in most statistical packages. The differences in the two models enter through the functional forms of the probabilities. As a consequence, the models are non-nested and the parameter set is larger under the multinomial logit model, which serves as an advantage for that model.

2.3 Data

The data used in this chapter is the same as that discussed in the first chapter. The dataset is made up of loans that serve as collateral for securities issued by Wells Fargo Mortgage Backed Securities Trust. In accordance with the guidelines of the trust, all loans are residential, first-lien, non-conforming mortgages issued by Wells Fargo Home Mortgage. The loans in this dataset are considered "jumbo loans" because, for the most part, the balance at origination exceeds the conforming limit. The source dataset contains loans that correspond to 97 collateral pools for 157 securities series issued since 2000, and comprise both fixed-rate and adjustable-rate mortgages. Performance history is available starting in October 2000. The data was obtained from Wells Fargo Corporate Trust Services, which serves as administrator and trustee to Wells Fargo Mortgage Backed Securities Trust as well as over 200 other issuers. The data is publicly available and released by request of the securities issuers.[5]

The analysis focuses on approximately 100,000 fixed-rate, single-family, owner-occupied mortgages issued between 2000 and 2005. Summary statistics are re-

[5]The data was obtained from www.ctslink.com.

ported in Tables 1.1 and 1.2 (Chapter 1). On average, the loans in the dataset follow the characteristics of jumbo loans. The average balance is $454,000, 95% of the loans have a loan to value (LTV) ratio below 80%, 65% have a FICO score above 720,[6] and 42% of the loans were contracted on property in California. Most of the mortgages have a term of 15 or 30 years. It is important to note that, although the dataset contains a large number of loans with substantial variation in terms and performance, the loans are selected as members of pools to use in securitization and, therefore, may display different trends than the mortgage market at large.

2.3.1 Explanatory Variables

Two of the most important factors for explaining mortgage payment trends are interest rates and housing prices. Interest rates enter the model through the refinance incentive. This incentive is measured as the "coupon gap", which is defined as the difference between the contracted coupon rate on the mortgage and the prevailing yield on the ten-year Treasury bond. Lags of the coupon gap are used to account for transaction costs, and the cube of the coupon gap is included to capture nonlinear effects. The slope of the yield curve, calculated as the difference between the yields on the ten-year and two-year Treasury bonds, indicates the direction in which interest rates are expected to move. Loan age is included to account for seasoning.

We are also interested in factors that influence delinquency and default. When

[6]FICO scores measure credit quality on a scale of 300 to 850.

the value of the property falls below the value of the mortgage, the borrower has an incentive to stop payment on the mortgage. This incentive is measured by the current LTV ratio, calculated as the ratio of the current balance to the estimated current value of the property. The current value of the property is estimated by applying growth rates calculated from the housing price index that corresponds to the metropolitan statistical area of the property.[7]

While delinquency is required for default, delinquency is not necessarily tied to the value of the property. When a borrower fails to make a mortgage payment, it is normally due to an income constraint, rather than a conscious decision to enter into default. Unfortunately, the income and employment expectations of the borrower are unobserved. State productivity is included to proxy for the strength of the local economy and control for relative income effects.

To accurately measure a borrower's refinance incentive, it is necessary to account for variation in refinancing opportunities. Although many of the relevant borrower characteristics are unobservable, certain loan characteristics can serve as proxy variables. These factors are also correlated with a borrower's propensity for delinquency, which is one of the key elements originators consider in constructing menus of mortgage terms. FICO scores are used to measure the borrower's credit quality, with a higher score implying better credit. The term and LTV ratio at origination are included to control for the borrower's financial flexibility and preference for debt. With loan-level data it is also possible to observe the purpose

[7]This index is published by the Office of Federal Housing Enterprise Oversight (OFHEO).

for which the *current* mortgage was contracted.[8] Including dummy variables for standard refinancing, home purchase, equity refinancing, and employer-sponsored relocation allows the model to control for the effects of transaction costs, and account for the borrower's previous mortgage decisions.

2.3.2 Payment Trends

The state of a loan is defined by comparing the actual balance to the scheduled balance at the end of the period. This is taken to be the initial state for the following period's decision process. If the actual balance is equal to the scheduled balance, then the loan is current. If the actual balance is less than the scheduled balance, then the loan is curtailed. Note that a payment in excess of the scheduled amount is not necessary in every period for the loan to remain curtailed. A sufficiently large payment in one period may keep the actual balance above the scheduled balance for several months regardless of the size of the payments that follow. When the actual balance falls to zero, the loan is prepaid and no further activity is observed on the loan.

The delinquency status is not determined by how many months the loan has been delinquent, but by how many months the balance is past due. If the current balance is greater than the scheduled balance from the *previous* period, the loan is at least 30 days delinquent. If the current balance is greater than the scheduled balance from two months prior, the loan is at least 60 days delinquent, and so on

[8] Note that this does not reveal whether an observed prepayment on the current mortgage is due to property sale or refinancing.

for 90 and 120 days delinquent. Any delinquency above 120 days is labeled as a 120-day delinquency. Under this definition, a borrower that misses one payment and then resumes making his scheduled payments without making up the missed payment will remain at 30-day delinquency for several periods.[9] Bankruptcy, foreclosure, REO, and liquidation are assigned by a code available in the raw data. The code changes over time tracking the progression to default.

Bankruptcy is not usually studied with respect to mortgages, but it is relevant to the mortgage decision because homestead protection can be used by the borrower to delay a foreclosure on his home. Since the dataset contains information on whether a loan is in bankruptcy, the ordered choice model can be used to determine the effect of bankruptcy on the mortgage decision. On the continuum of choices this event is placed between 120-day delinquency and a foreclosure. A bankruptcy is an extreme event so a borrower would most likely be delinquent for several months before making the decision to enter bankruptcy. On the other hand, bankruptcy delays a foreclosure so we can assume that, for the borrower, bankruptcy is preferred to foreclosure.

Table 2.1 shows the percentage of the 103,347 loans that enter a non-current state for at least one period. Two-thirds of the loans prepay prior to the end of the five-year observation period, and only 0.3% default. A default is defined as a liquidation, or a prepayment from a state that is to the right of 60-day delinquency

[9] It is assumed that a borrower must pass through 30-day delinquency before entering 60-day delinquency, that a borrower must pass through 60-day delinquency before entering 90-day delinquency, and so forth. In this way, accumulated interest cannot cause a borrower to jump from 30-day to 90-day delinquency or from 60-day to 120-day delinquency.

on the continuum. Nearly all the loans issued prior to 2003 have been prepaid. This is in contrast to only 30% prepayment among the loans issued in 2003, the year interest rates fell to their lowest point. Perhaps surprisingly, loans issued in 2003 have the highest frequency of curtailment at 53% in contrast to the 42% experienced by the loans overall.

While 35% percent of the loans enter 30-day delinquency, only 4% progress to 60-day delinquency. The proportion of loans entering 90-day and 120-day delinquency is even lower at 1.7% and 0.4%, respectively. Delinquency appears to be more common among loans issued in 2000 and 2005, possibly indicating poor economic conditions for housing investment during those years. Small numbers of loans experience bankruptcy (146), foreclosure (250), or REO (55). While these are small subsamples compared to the total sample of 103,347 loans, this is consistent with the expected frequency of default events, which are not common because of high transaction costs. In addition, the incentive to default is expected to be low during a period of rapid growth in housing prices. By using an ordered logit model it is possible to account for small event samples better than with a hazard rate model or a multinomial logit model because the sample used in estimation is determined by the initial state, not the termination state.

One of the features of the observed payment trends is a high degree of persistence in the same state. As an example, the histograms in Figure 2.2 show profiles of the spell length among borrowers that experience curtailment or 30-day delinquency.[10] Spells of curtailment are generally long: over 75% of the borrowers that

[10]Percentages are taken with respect to all loans that have a spell in the target state of at

curtail are in this state for four months or more. By contrast, about 60% of the borrowers that enter 30-day delinquency remain in this state for three months or less, although a good number experience longer spells. The loans in this dataset also exhibit incidences of recurrence. Delinquent loans that recover after a few periods may fall into delinquency again, and curtailed loans may become current but revert to curtailment at a later date.

Table 2.2 provides information on such patterns of persistence and recurrence. It shows the average and maximum spell *lengths* for each event, as well as the average and maximum *number* of spells over the life of a loan. The average spell of curtailment is long at 11 months, and on average loans that curtail experience more than one spell of curtailment. The average length of a spell of 30-day delinquency is three months, while the average frequency of spells is two, greater than for any other event. The average spell length for 60-day and 90-day delinquency is short, only one month, largely because recovery from these events is quite high. Loans may spend several months in REO as the sale proceedings are concluded, while the long spells observed for 120-day delinquency reflect the practice of keeping loans in this category until foreclosure is declared.

A loan is defined to be in "recovery" if it reverts to a current, curtailed, or 30-day delinquency status at some point after the first instance of a delinquency event. Recovery rates are reported in the last column of Table 2.2.[11] Nearly

least one month.

[11]Prepayment from a delinquent state is *not* considered a recovery. Recovery from 30-day delinquency is defined as a move to a current or curtailed state only. The recovery rates reported do not take into account a potential relapse into delinquency after recovery.

70% of the loans that enter 30-day delinquency recover from the first spell of this event. Recovery is also high for loans that enter 60-day delinquency, 43%. Even recovery from more severe delinquency is relatively frequent with rates of 24% for 90-day delinquency, 33% for 120-day delinquency, and 36% for foreclosure. This indicates that severe delinquency, even foreclosure, may not lead to the loss of principal in the case of prime mortgage loans, a clear departure from the delinquency experience of subprime mortgages.

2.4 Results

The estimation results from both the multinomial logit model and the ordered logit model provide evidence to support the claim that it is important to model non-termination events, and to condition the choice set and the utility space on the state of the mortgage at the beginning of the payment period. Tables with parameter estimates can be found in the Appendix. In a multinomial logit model, there is a separate utility function for each transition. As a result, the coefficients of the model are specific to the destination state as well as the initial state. In an ordered logit model, by contrast, a single set of coefficients links all of the destination states available from a given initial state, and threshold parameters are used to determine the utility associated with the optimal choice. This distinction is a key element to the comparison of the two models. A higher degree of parameterization gives the multinomial logit model greater flexibility to capture heterogeneity among borrowers, but the utility specification in the ordered logit model encompasses correlation between competing risks that is essential to accurately model the borrower's decision process.

Under both models the coefficients are different across initial states, indicating that the utility space varies with the choices made by the borrower in previous periods. A likelihood ratio statistic is used to test whether the coefficients are jointly equal to zero. For all initial states the null hypothesis is rejected, confirming the overall significance of the specification. The significance of individual coefficients differs markedly over the initial states. For the states most likely to be represented in hazard rate models for prepayment (curtailed, current, and 30 days delinquent), most of the factors are significant. As the states enter further into delinquency, however, fewer and fewer of the explanatory variables remain significant. This outcome could simply be a matter of sample size, a greater number of observations fall into the non-delinquent states leading to smaller standard errors, or it could be an indication that delinquent borrowers are mostly influenced by factors that cannot be observed, such as income expectations and wealth flexibility.[12]

Tables 2.3 through 2.12 show the average estimated derivatives for each transition event, with respect to the covariates. These derivatives are used to facilitate comparison between the two models. The formulas for the derivatives are given

[12]The low significance among the delinquent states is not because borrowers in those states tend to move towards more severe forms of delinquency rather than towards prepayment, where the factors considered may be more relevant. Indeed, empirical probabilities show that a borrower in 90-day delinquency is just as likely to prepay as a borrower who is current. It is true, however, that the factors influencing a prepayment will be different for those two borrowers.

below.[13]

$$MNL: \quad \frac{\partial P_{ij}(v_{nt})}{\partial v} = \beta_{ij}\frac{\exp(\beta_{ij}v_{nt})}{\sum_k \exp(\beta_{ik}v_{nt})} - \frac{\exp(\beta_{ij}v_{nt})}{\sum_k \exp(\beta_{ik}v_{nt})}\sum_l \left(\beta_{lj}\frac{\exp\left(\beta_{lj}v_{nt}\right)}{\sum_k \exp(\beta_{ik}v_{nt})}\right)$$

$$OLG: \quad \frac{\partial P_{ij}(v_{nt})}{\partial v} = -\beta_i\left[\frac{\exp(k_{ij}-\beta_i v_{nt})}{(1+\exp(k_{ij}-\beta_i v_{nt}))^2} - \frac{\exp(k_{ij+1}-\beta_i v_{nt})}{(1+\exp(k_{ij+1}-\beta_i v_{nt}))^2}\right]$$

Since these formulas depend on the covariate values and are, therefore, observation specific, the tables show the average over all the observations that fall into a given initial state. Since the events in the choice set are mutually exclusive, the transition probabilities must add up to unity and the derivatives with respect to a single variable must add to zero. The results below focus on the tradeoff between transition probabilities. If a given factor has a positive effect on the probability of prepayment, i.e. a positive probability derivative, the probability of another event must decrease. The tradeoff between event probabilities is determined by the definition of the event continuum in the ordered logit model. In the multinomial logit model, the this tradeoff is driven by the IIA property.

2.4.1 Transition From a Current Status

Transition from a current state is the most common event, and is the focus of the majority of studies that look at prime mortgages. Table 2.3 reports the probability derivatives estimated for loans that begin the period in a current

[13]The expression given for the ordered logit model corresponds to an interior event. Similar equations can be derived for the events on the end of the continuum.

state. Under both models the LTV ratio at origination has a positive effect on prepayment and curtailment, and has a negative effect on remaining current and entering 30-day delinquency. Borrowers with higher LTV ratios at origination carry a higher loan burden and gain the most from refinancing. This is also the cause of the positive correlation between the mortgage term, and the probability of prepayment and curtailment. On the other end of the spectrum, borrowers with lower LTV ratios tend to have greater financial stability and are, thus, less likely to fall into delinquency. The contemporaneous LTV ratio, on the other hand, has a negative correlation with prepayment and curtailment, and a positive correlation with staying current and entering 30-day delinquency. This indicates that, controlling for the LTV ratio at origination, borrowers with higher contemporaneous LTV ratios are less likely to prepay and more likely to fall into delinquency, perhaps because of a slower rate of property appreciation or previous episodes of delinquency.

The pairing of prepayment with curtailment versus staying current and entering 30-day delinquency is a natural result of the ordered logit model. When a borrower is current, prepayment and curtailment require a payment that is larger than the scheduled amount. Staying current or entering 30-day delinquency require making a payment that is equal to or less than the scheduled amount. When there is a shift in a covariate that increases the utility associated with making a mortgage payment, the borrower will move towards prepayment or curtailment and away from delinquency and staying current. Although this tradeoff structure is not inherent to a multinomial logit model, the pairing of the effects for prepayment and curtailment versus the effects for staying current and entering 30-day delinquency are observed under the multinomial logit model for the LTV ratio at origination, the contemporaneous LTV ratio, and the mortgage term. However, for several key factors the probability derivatives are different for the two models, leading to a different tradeoff structure under the multinomial logit model than under the ordered logit model, and implying different predictions in borrower behavior.

The factor that has the largest effect on the probability of prepayment is the one-month lag of the coupon gap. As interest rates fall the coupon gap widens and borrowers can minimize their mortgage costs by refinancing. Under the multinomial logit model, an increase in the coupon gap also leads to an increase in the probability of entering 30-day delinquency, which raises questions as to the adequacy of the multinomial logit model. These increases are offset by a decrease in the probability of remaining current, which is the reference category. There is no effect on the probability of curtailment under the multinomial logit model.

Under the ordered logit model, an increase in the coupon gap leads to increases in the probabilities of prepayment *and* curtailment, which are offset by a decrease in the probabilities of staying current and entering 30-day delinquency. The predictions made under the ordered logit model indicate that the other factors associated with interest rates, namely, the three-month lag of the coupon gap and the slope of the yield curve, have similar effects to the one-month lag of the coupon gap, but with smaller magnitudes. Under the multinomial logit model, these factors have a positive impact on both the probability of prepayment and curtailment while having little effect on the probability of entering 30-day delinquency.

An interesting factor to examine is the purpose for which the current mortgage was contracted. This is not an indicator of why a prepayment may be triggered, but rather a way to condition on past behavior. In particular, this factor allows the model to control for the circumstances of the borrower's last prepayment. Under both models the probability of prepayment is lower, and the probability of 30-day delinquency is higher for borrowers whose current mortgage was contracted for an equity refinancing rather than a standard refinancing. This may be because a reduction in home equity will limit future refinancing opportunities, or point to an income constraint that could contribute to an increase in the probability of becoming delinquent. Under the ordered logit model, a borrower who recently purchased his home is more likely to prepay than a borrower who recently refinanced. This is consistent with previous studies (Dunn and Spatt (2005)) that find that transaction costs delay multiple refinancing. Under the multinomial model, however, home purchase is associated with a lower probability of prepayment than a previous refinancing.

Transition to 30-day delinquency is driven primarily by poor credit quality. Under both models the probability of 30-day delinquency increases as a borrower's FICO score decreases. Borrowers with poor credit, a FICO score below 660, are less likely to prepay, curtail, *or stay current*, indicating a strong propensity towards delinquency. Low FICO scores are also associated with a reduction in the probability of prepayment because borrowers will face fewer refinancing opportunities. Under the ordered logit model, borrowers with lower FICO scores are also less likely to curtail their mortgage. This is a reasonable result because a lower FICO score suggests a disinclination towards timely payment of liabilities and a preference for debt. The multinomial logit model, however, predicts no relationship between a decrease in FICO scores and the probability of curtailment.

While some of the differences between the multinomial logit model and the ordered logit model are inconsequential, other differences are quite puzzling. For example, we would expect credit quality to have some relationship with the propensity for curtailment as is predicted by the ordered logit model. Another difference between the two models is found with respect to property appraisal. Under the ordered logit model, higher property appraisal implies a higher probability of curtailment, which is reasonable if we believe that this factor proxies for higher income. The multinomial logit model, by contrast, predicts a lower probability of curtailment to be associated with higher property appraisal.

These differences are a consequence of the substitution patterns underlying the models. In the multinomial logit model, due to the IIA property, the probability ratios of two alternatives are independent of any other choice. A change in a factor that contributes to one or two events (e.g. property appreciation, which has a positive effect on the probability of prepayment and a negative effect on the probability of 30-day delinquency under both models) will be predicted to affect a third event (in the example, curtailment) in such a way as to preserve the independence between alternatives, rather than with any regard to the true economic effect of that factor. In the order logit model, by contrast, there is an explicit correlation between alternatives. Similar choices, such as prepayment and curtailment, will be predicted to exhibit similar effects.

2.4.2 Transition From a Curtailed Status

For loans that are curtailed at the beginning of a period, the probability trade-off structure for the ordered logit model is centered around curtailment itself. A borrower that is in curtailment at the beginning of the period can remain in curtailment by simply making the scheduled payment; falling back to a current or 30-day delinquent status requires making a payment less than the scheduled amount. Only prepayment requires a payment larger than the scheduled amount. Under the ordered logit model, the factors that increase the refinance incentive, such as the coupon gap and the slope of the yield curve, only have a positive effect on the probability of prepayment. The probabilities of maintaining a curtailed status, returning to a current status, and entering 30-day delinquency move together and opposite to the probability of prepayment.

Table 2.4 reports the probability derivatives estimated for loans in a curtailed state. For the ordered logit model, the effects on the probability of prepayment are similar to those predicted for loans in a *current* state. The LTV ratio at origination, the property appraisal at origination, the coupon gap, the slope of the yield curve, and the mortgage term all have a positive effect on the probability of prepayment. The contemporaneous LTV ratio, the FICO score, and the first quarter dummy variable all have a negative effect on the probability of prepayment. The magnitude of the derivatives on the probability of prepayment are larger for loans in a curtailed state than in a current state. This suggest that by curtailing the borrower may have compensated for transaction costs that had prevented a prepayment from a current state. This result indicates that curtailment may be a strong signal for prepayment.

An interesting difference between the results from a current and curtailed status is that from a curtailed status age has a significant effect, increasing the probability of prepayment. From a current status age was not a significant factor. Another difference is that all mortgage purposes — equity refinancing, purchase, and relocation — have a lower probability of prepayment than a standard refinancing. This is consistent with the results observed under the multinomial logit model.

For loans that begin the period in a curtailed state, as with *current* loans, the substitution patterns under the multinomial logit model do not match those observed under the ordered logit model. For many of the key variables — age, property appraisal, the coupon gap, slope of the yield curve, and mortgage term — an increase in the probability of prepayment is offset primarily by a decrease in the probability of maintaining a curtailed status. For the LTV ratio at origination and the contemporaneous LTV ratio, the probability derivatives match those observed for loans in a current status. A higher LTV ratio at origination implies higher probabilities of both prepayment and curtailment, and lower probabilities of staying current and entering 30-day delinquency, while a higher contemporaneous LTV ratio leads to a lower probability of prepayment and curtailment.

A puzzling result, which may be linked to the influence of the IIA property, is that under the multinomial logit model the borrower's FICO score has no effect on the probability of prepayment from a curtailed state. This result raises questions because it is commonly believed that FICO scores are one of the key factors that originators consider when determining mortgage terms and rates. As a result, we would expect lower FICO scores to reduce the refinancing opportunities available to the borrower even though he has curtailed his loan. It is possible that the increase in home equity caused by the curtailment compensates for lower credit quality, but it is not likely that an originator of a prime mortgage would overlook a truly bad FICO score.

2.4.3 Transition From 30-Day Delinquency

While 35% of the loans enter 30-day delinquency, only 4% go on to 60-day delinquency. The dominant propensity is for a borrower to miss a payment in one month, and then resume his payment schedule in the following month. Over a one-month period nearly 60% of the borrowers in 30-day delinquency will continue in that state, and this probability increases with age. One of the hypotheses behind this trend is that, when borrowers find themselves in delinquency, they may choose prepayment as a way of resolving a liability they can no longer handle. However, the results reported in Table 2.7 show that for the ordered logit model an increase in the coupon gap leads to a *decrease* in the probability of prepayment from 30-day delinquency. This suggests that prepayments observed from 30-day delinquency may not be the result of refinancing but rather property sale.

Evidence supporting the hypothesis that borrowers in 30-day delinquency may turn to prepayment in order to resolve a liability they cannot handle is given by the effect of the borrower's FICO score. Under the ordered logit model, borrowers with good (750-660) credit are more likely to prepay or curtail their mortgage than borrowers with excellent (850-750) credit. Borrowers with good credit are also *less* likely to remain in delinquency than borrowers with excellent credit. For borrowers with poor credit (below 660) the expected results hold, i.e poor credit leads to an increase in the probability of delinquency. For current and curtailed loans, by contrast, any reduction in credit rating leads to a lower probability of prepayment and a higher probability of delinquency. A possibility is that borrowers with excellent credit, who find themselves in delinquency, have greater financial flexibility for generating the funds to make up missed payments and do not need to sell their property to recover. It could also be that borrowers with good, but not excellent credit, are delinquent in other liabilities in addition to their mortgage. They may have a greater need to sell their property or otherwise adjust the terms on their mortgage.

It is not unreasonable to suppose that a borrower in 30-day delinquency, which is not a severe credit event, can refinance to a mortgage with terms that are more manageable. This is particularly true in an environment of high housing prices, when such borrowers may face mobility constraints, and during periods of low interest rates, when increased competition among mortgage originators might expand the refinancing opportunities available to a delinquent borrower. Table 2.5 shows that, under the multinomial logit model, there is a positive relationship between the coupon gap and the probability of prepayment for loans in 30-day delinquency. This result suggests that refinancing opportunities may, in fact, be available for borrowers in delinquency. In addition, both the ordered logit model and the multinomial logit model predict that an increase in the slope of the yield curve will lead to an increase in the probability of prepayment, curtailment, and staying current, and a decrease in the probability of falling further into delinquency. For loans that are in delinquency, the slope of the yield curve may be a better proxy for the refinance incentive than the coupon gap.[14]

As was seen with respect to current and curtailed mortgages, for loans that begin the period in 30-day delinquency, the multinomial logit model and the ordered logit model agree on the effects of both the LTV ratio at origination and the contemporaneous LTV ratio. Controlling for the LTV ratio at origination, a loan with a higher contemporaneous LTV ratio will be more likely to remain in

[14]We could also consider this to be a delinquency effect. An inverted yield curve is a signal for a recession, and so the slope of the yield curve may be correlated with income uncertainty. Since state productivity has little effect under either model, it is possible that the slope of the yield curve is a better indicator of economic conditions.

delinquency under both models. This could represent a higher loan burden, or the effects of housing depreciation. However, over this period such an effect is more likely to represent a long spell in a delinquent state, since housing prices are consistently appreciating.

For loans in 30-day delinquency, the negative consequences of the IIA property are particularly strong. A decrease in the FICO score leads to an increase in the probability of entering 60-day delinquency *and* an increase in the probability of curtailing the loan. These positive effects are offset with a decrease in the probability of recovering to a current status; there is little effect on the probability of remaining at 30-day delinquency. It seems doubtful that a borrower with poor credit would be equally likely to miss an additional payment as to receive sufficient extra income to, not only recover, but actually curtail their mortgage. Another puzzling result is that an increase in the coupon gap leads to a very large increase in the probability of staying in 30-day delinquency, offset almost entirely by a decrease in the probability of recovering or entering 60-day delinquency, two very different events. This is further evidence that the substitution patterns implied by the multinomial logit model may not be appropriate for the analysis of mortgage payment trends.

2.4.4 Transition From Severe Delinquency

Since it is observed that most loans recover from 30-day delinquency, it is not accurate to consider these results to be representative of delinquent loans. Loans that are in 30-day delinquency do share certain characteristics with delinquent loans, but mortgages in 60-day delinquency are more relevant for understanding the trends that lead to severe delinquency and ultimately default. A borrower that is 60 days delinquent faces the critical choice of whether to pay off the past due amount or enter 90-day delinquency and be threatened with foreclosure.

Table 2.6 shows that, under the multinomial logit model, for loans that are in 60-day delinquency, only age, the coupon gap, the mortgage term, the quarterly dummy variables, and the indicator for relocation are relevant to the probability of prepayment. Comparing with Table 2.8, the only factor that has the same effect on the probability of prepayment under the two models is the coupon gap. This indicates that borrowers who are 60 days delinquent are quite interest-rate sensitive. The models do agree, however, on the factors that influence a borrower to enter 90-day delinquency. The one-month lag of the coupon gap, the slope of the yield curve, and the mortgage term all have large negative effects on the probability of entering 90-day delinquency. Borrowers with good, but not excellent credit, and borrowers whose current mortgage was contracted for relocation, have lower probabilities of entering 90-day delinquency and staying in 60-day delinquency. The only factors that contribute to an increase in the probability of entering 90-day delinquency are an increase in the three-month lag of the coupon gap and the fourth quarter dummy variable. There is little agreement between the two models as to what may cause a borrower to recover from 60-day delinquency.

Tables 2.6 and 2.8 also show the derivatives for the probability of foreclosure from 60-day delinquency. However, the significant results that are observed appear to be a result of the model's substitution pattern, rather than true effects. An event that has not been discussed up to this point is bankruptcy. It is assumed that bankruptcy can be entered from any initial state. However, since it is not truly a mortgage event, we find that few of the factors considered in the analysis have any effect on the probability of entering bankruptcy.

A serious drawback of the multinomial logit model is that it cannot handle small event samples. In this analysis, the multinomial logit model was not able to estimate coefficients for any state below 60-day delinquency, because of a quasi-complete separation of data points. A quasi-complete separation means that there is a linear combination of the covariates that separates the response variables into unique regions:[15]

$$\begin{cases} \gamma v \leq c_0 : y_0 = 1 \\ c_0 \geq \gamma v \geq c_1 : y_1 = 1 \\ \dots \\ \gamma v \geq c_J : y_J = 1 \end{cases}.$$

In such cases, the parameter estimates approach infinity, and the log-likelihood function diminishes to a non-zero constant. This is problematic because in the study of mortgages it is very important to use a model that can handle small event samples. Delinquency and default are not frequent events, but it is advantageous to accurately forecast the likelihood of these events since their occurrence has serious consequences for the borrower, originator, and investor.[16]

[15]A complete separation is defined by a strict inequality.

[16]This has played out to dire results in the subprime market. Between 2000 and 2005 there

One of the advantages of an ordered logit model is its ability to handle small event samples. This benefit is realized because the limiting constraint on the sample size is determined by the number of observations in the initial state, not the termination state.[17] Tables 2.9 through 2.12 report the probability derivatives for transition from a state below 60-day delinquency under the ordered logit model. For loans that are in 90-day delinquency, the relevant factors are the one-month lag of the coupon gap, the slope of the yield curve, the contemporaneous LTV ratio, and the term of the mortgage. At 120-day delinquency the transition probabilities are influenced by loan age, the property appraisal at origination, and the one-month lag of the coupon gap. While age has a positive effect on the probability of recovery, versus falling deeper into delinquency, property appraisal at origination has a negative effect. Loans in bankruptcy are more likely to prepay as the coupon gap *falls* and the slope of the yield curve rises. For loans in foreclosure, the only significant factors are the slope of the yield curve and the contemporaneous LTV ratio. It is important to note that, since most of the factors that affect delinquency and default are unobserved, these significant factors are likely to proxy for economic shocks that are not directly accounted for.

---

were few defaults, even among subprime mortgages, due to rising housing prices. A model that was unable to handle small event samples would be unable to predict the dramatic increase in foreclosures that occurred when housing prices began to fall.

[17]In a multinomial logit model, since a separate utility function is estimated for each transition, the limiting constraint is determined by the number of observations that transition *to* the state in question. For an ordered logit model, only a single utility function is estimated from a given initial state; the estimation procedure is affected only by the number of loans in that initial state, not by how many loans transition to each event.

2.4.5 Forecasted Transition Probabilities

A feature of models that consider non-termination events is the ability to forecast transition matrices. A transition matrix is made up of the probabilities of all potential transition pairs at a given point in time. Under the Markov property the transition matrices for an individual can be multiplied together to generate the probability of a borrower's choice stream. This is useful in predicting the payment path of an individual over time, and is essential for accurate pricing algorithms.

Figures 2.3 through 2.8 graph the *average* forecasted transition matrices stratified by loan age. Despite the differences observed in the derivative estimates between the multinomial logit model and the ordered logit model, the forecasted transition probabilities are quite similar. While both models track the empirical probabilities relatively well, the multinomial logit model generates predicted transition probabilities that are closer to the observed values. This may be a result of a higher degree of parameterization in the multinomial logit model, which provides this model with greater flexibility for matching in-sample trends. However, this feature must be weighed against the negative effects of the IIA property.

Although the ordered logit model has fewer parameters, it has a distinct parametric feature that is not available in a multinomial logit model: the threshold parameters. These parameters define the favorable substitution patterns observed under the ordered logit model, and impose the correlation between competing risks that is necessary for accurately modeling the borrower's decision process.

While the multinomial logit model may be more accurate with respect to in-sample probabilities, the discussion in this section suggests that the multinomial logit model is fundamentally misspecified, and that an ordered logit model may perform better out-of-sample, where economic conditions will differ from those observed in-sample.

2.5 Conclusion

This chapter compares a multinomial logit model with an ordered logit model. Both models, based on a random utility framework, have the capability to model the full range of borrower behavior, including non-termination events, and to condition the borrower's utility space on past choices. The results in this chapter show that the borrower's response varies significantly with the status of the mortgage, and confirm the importance of tracking a borrower's behavior over time.

Under the multinomial logit model, a separate utility function is defined for each transition, granting the multinomial logit model the flexibility to capture the variation observed in in-sample probabilities. Although the ordered logit model has a more constrained parametrization, it has the distinct advantage of accounting for correlation between the mortgage events. This chapter finds that this is a fundamental aspect of properly modeling the borrower's decision process, and that, as a result, the IIA property is a severe drawback of the multinomial logit model.

An interesting result of this chapter is that it is valuable to study delinquency separately from default. Although delinquency is closely tied to default, delinquency is really an income-driven event. This chapter shows that delinquency does not necessarily lead to default, and that delinquent behavior has as much bearing on the probability of prepayment as it does on the probability of default. This result highlights an additional failing of the multinomial logit model, problems with small event samples. Although severe delinquencies are infrequent, it is necessary to consider them in an analysis of borrower behavior. An ordered logit model is better able to handle small event samples, and the straightforward structure of an ordered logit model makes it attractive for dealing with other complexities of borrower behavior.

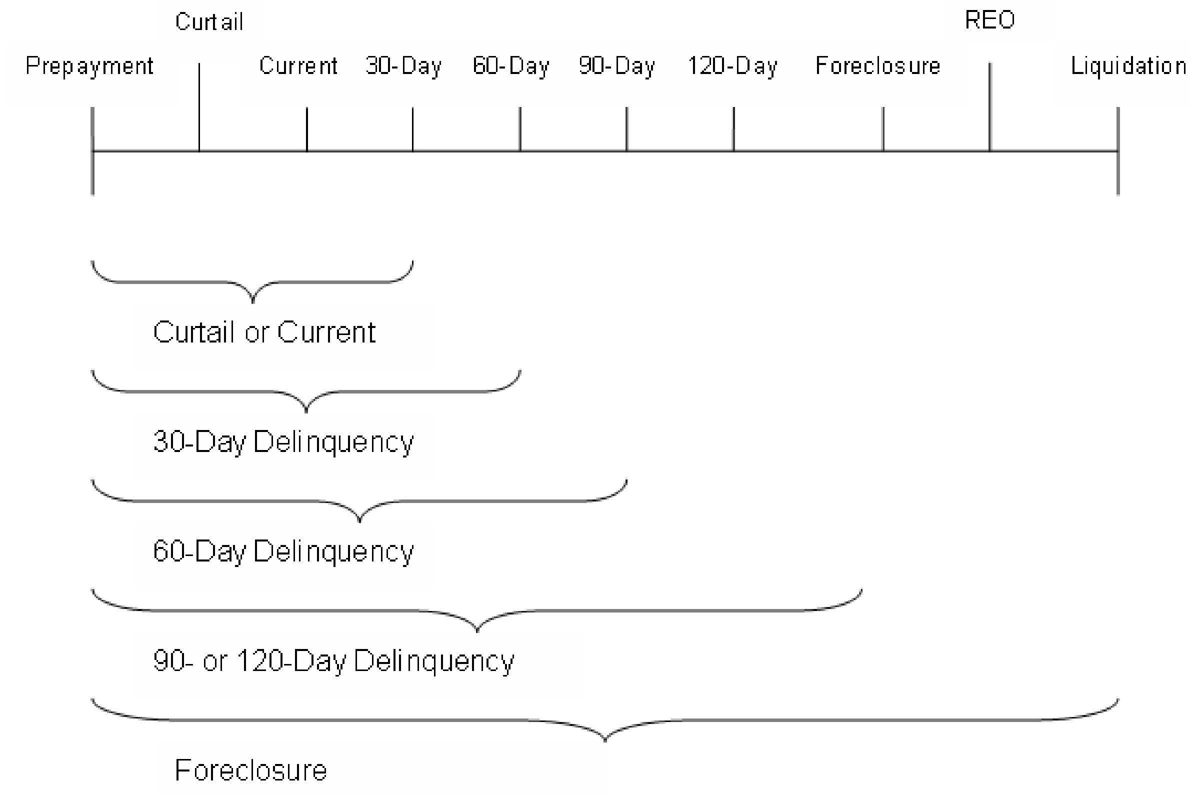


Figure 2.1: Continuum of Choices Available to a Borrower

This figure shows the continuum of mortgage choices. The choices available to a borrower over a one month payment period depend on the state of the mortgage at the beginning of the period; initial states are given below the continuum. As the mortgage falls deeper into delinquency, the choice continuum expands.

Table 2.1: Number of Loans that Enter a Non-Current State

Issue Year	2000	2001	2002	2003	2004	2005	All
Prepayment	13,293	31,721	16,766	6,265	740	219	69,004
	99.3%	99.2%	91.5%	29.6%	14.1%	1.7%	66.8%
Curtailment	4,947	13,294	7,990	10,987	2,472	3,352	43,042
	37.0%	41.6%	43.6%	51.8%	47.1%	25.4%	41.7%
30-Day Delinquency	6,496	12,498	5,213	5,908	1,243	4,442	35,800
	48.5%	39.1%	28.4%	27.9%	23.7%	33.6%	34.6%
60-Day Delinquency	733	935	274	229	40	2,171	4,382
	5.5%	2.9%	1.5%	1.1%	0.8%	16.4%	4.2%
90-Day Delinquency	178	251	78	65	3	1,126	1,701
	1.3%	0.8%	0.4%	0.3%	0.1%	8.5%	1.7%
120-Day Delinquency	100	152	51	34	3	104	444
	0.8%	0.5%	0.3%	0.2%	0.1%	0.8%	0.4%
Bankruptcy	31	60	24	21	6	4	146
	0.2%	0.2%	0.1%	0.1%	0.1%	0.03%	0.1%
Foreclosure	69	125	36	24		2	256
	0.5%	0.4%	0.2%	0.1%		0.02%	0.3%
REO	19	29	5	2			55
	0.1%	0.1%	0.03%	0.01%			0.1%
Default *	97	140	38	7	1		283
	0.7%	0.4%	0.2%	0.03%	0.02%		0.3%

* Default is defined as a loan that is liquidated or prepays when the status is 90-day delinquency or below.

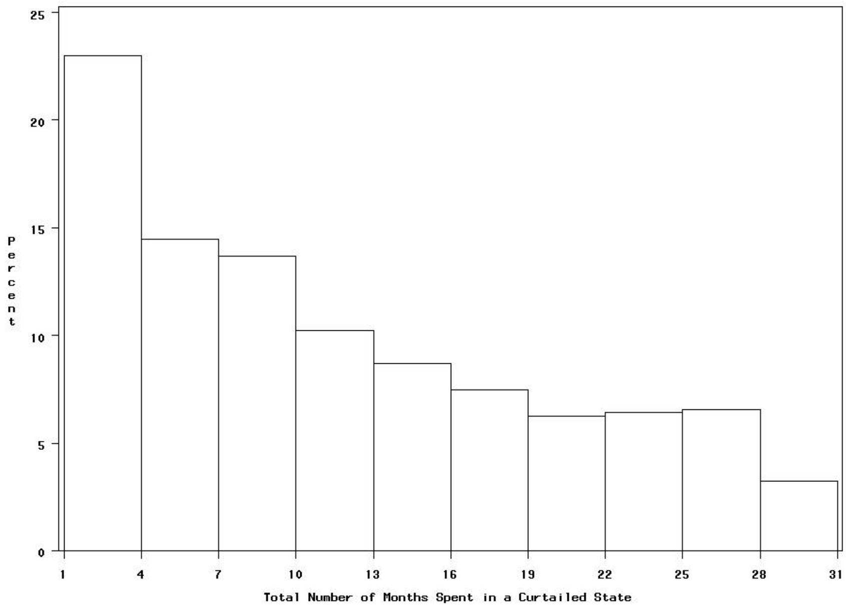
This table shows the number of loans that enter the listed event for at least one period. Percentages are based on the total number of *loans* issued in the year displayed at the top. 67% of all loans prepay prior to the end of the five-year observation period. This percentage is above 90% for loans issued prior to 2003. While this suggests very short mortgage lives, it is not surprising since the lowest point for interest rates was in 2003. There is also substantial curtailment and delinquency activity. Curtailment is most common for loans issued in 2003 and 2004 the years with the lowest delinquency behavior. 35% of the loans enter 30-day delinquency at some point over the observation period. However, this event is not linked to further delinquent behavior as only 4% of the loans enter 60-day delinquency, and only 0.3% terminate in a delinquent state.

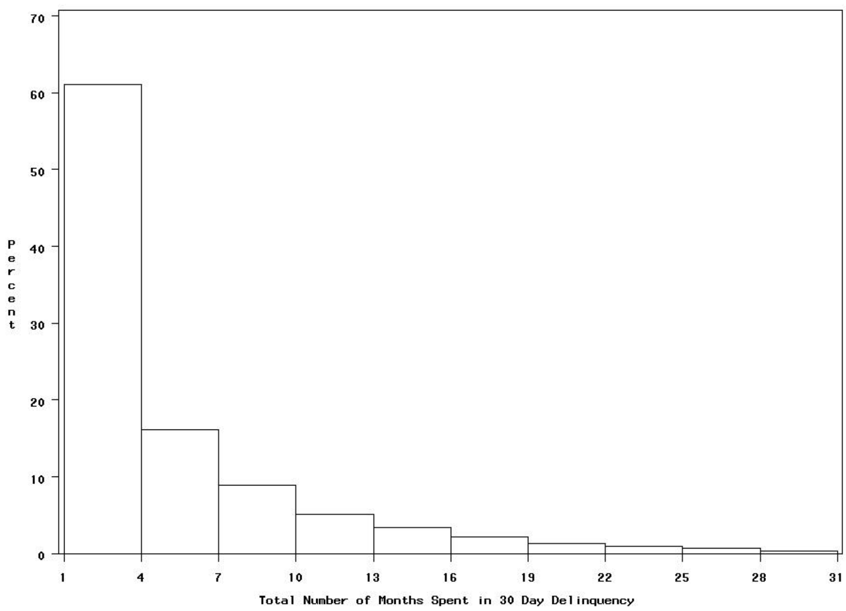
Table 2.2: Average Spell Length and Count

	Spell Length (Months)		Spell Count		% Recover
	Avg.	Max.	Avg.	Max	
Curtailment	11	47	1.3	12	-
30-Day Delinquency	3	38	2	14	69
60-Day Delinquency	1	25	1.3	11	43
90-Day Delinquency	1	23	1.2	7	24
120-Day Delinquency	5	35	1.2	7	33
Bankruptcy	6	30	1.1	3	-
Foreclosure	3	18	1.2	6	36
REO	5	17	1	1	0

This table shows the average and maximum length of spells experienced by the borrowers. It also shows the average and maximum *number* of spells. Spells for curtailment and 30-day delinquency are longer than one month, on average. For all events, other than REO, the average count is greater than one. This indicates that borrowers regularly move in and out of a *current* state. Recovery is defined as a loan that achieves a current, curtailed, or 30-day delinquency status at some point after the first instance of a serious delinquency event. A prepayment from a delinquent state is not considered a recovery. For 30-day delinquency, only a transition to a current or curtailed state is counted as a recovery rate. These calculations do not take into account a potential relapse into delinquency after the recovery event. Recovery from the first instance of a delinquent event is high, even for foreclosure. In contrast with subprime mortgages, it appears that borrowers of jumbo mortgages have the facility to avoid default.

Figure 2.2: Total Number of Months Loans are Curtailed or in 30-Day Delinquency





These histograms graph the total number of months a loan spends in a curtailed state or in 30-day delinquency. Only loans with spells of at least one month are considered. Spells of curtailment are naturally long. A curtailed state is intended to be a permanent state since returning to the amortization schedule after a spell of delinquency eliminates the benefit obtained in previous periods. On the other hand 30-day delinquency is an event with shorter spells. The borrower will want to return to a current status as soon as he is able to make up the missed payments.

Table 2.3: Average Estimated Derivative of Transition Probabilities from Current

	Multinomial Logit Model				Ordered Logit Model			
	Prepay	Curtail	Current	30-Day	Prepay	Curtail	Current	30-Day
Age (Months)	0.09	-0.15	0.03	0.03	0.00†	0.00†	0.00†	0.00†
LTV at Origination	0.08	0.07	-0.12	-0.03	0.06	0.03	-0.02	-0.07
Property Appraisal (per $100,000)	0.09	-0.03	0.00	-0.05	0.04	0.02	-0.02	-0.05
Coupon Gap (1-month lag)	4.12	-0.17†	-4.85	0.89	1.43	0.74	-0.70	-1.48
Coupon Gap (3-month lag)	0.33	0.41	-0.82	0.08†	0.42	0.22	-0.16	-0.47
Slope of Yield Curve (1-month lag)	2.26	0.11	-2.15	-0.22†	0.92	0.48	-0.35	-1.04
Current LTV	-0.09	-0.09	0.16	0.03	-0.07	-0.03	0.03	0.07
State Productivity	0.01	0.00†	-0.01	0.00	0.00†	0.00†	0.00†	0.00†
Mortgage Term < 30 years	1.24	0.45	-1.45	-0.23	0.82	0.43	-0.32	-0.93
Baseline FICO: 850 - 750								
FICO: 750 - 720	-0.31	0.05	-1.69	1.94	-0.59	-0.31	0.18	0.73
FICO: 720 - 660	-0.83	-0.03†	-2.90	3.74	-1.53	-0.81	0.32	2.02
FICO: 660 - 620	-1.47	-0.13†	-4.01	5.60	-2.46	-1.34	-1.21	5.01
FICO: 620 - 300	-2.26	0.04†	-4.57	6.78	-3.11	-1.73	-4.30	9.13
Baseline: Jan - Mar								
Apr - Jun	0.70	0.14	0.43	-1.27	0.87	0.45	-0.41	-0.90
Jul - Sep	0.79	0.13	0.15	-1.07	0.84	0.44	-0.40	-0.88
Oct - Dec	0.09†	0.04	-0.09	-0.04†	0.19	0.10	-0.08	-0.22
Baseline: Standard Refinancing								
Equity Refinancing	-0.88	-0.01†	0.06	0.84	-0.58	-0.31	0.18	0.72
Purchase	-0.56	0.61	0.06	-0.11†	0.11	0.06	-0.04	-0.12
Relocation	-1.75	0.35	-3.92	-1.81	0.07†	0.04†	-0.03†	-0.08†

† Coefficient is not statistically significant.

The pairing of prepayment with curtailment versus staying current and entering 30-day delinquency is a natural result of the ordered logit model. For nearly all of the covariates, the derivatives of the prepayment probability are consistent across the two models, although the derivatives are higher under the multinomial logit model. However, for several key factors the probability derivatives for the other events are substantially different from those predicted by the ordered logit model. This difference is the result of the IIA property.

Table 2.4: Average Estimated Derivative of Transition Probabilities from Curtailed

	Multinomial Logit Model				***Ordered Logit Model***			
	Prepay	Curtail	Current	30-Day	Prepay	Curtail	Current	30-Day
Age (Months)	0.13	-0.11	-0.07	0.04	0.11	-0.02	-0.04	-0.05
LTV at Origination	0.03	0.10	-0.08	-0.06	0.06	-0.01	-0.02	-0.03
Property Appraisal (per $100,000)	0.09	-0.13	0.04	-0.01	0.02	0.00	-0.01	-0.01
Coupon Gap (1-month lag)	3.84	-3.74	0.01	-0.10†	2.50	-0.48	-0.82	-1.20
Coupon Gap (3-month lag)	0.10	0.03	-0.13	0.00	0.29	-0.06	-0.09	-0.14
Slope of Yield Curve (1-month lag)	2.40	-2.06	-0.30	-0.05	1.55	-0.31	-0.50	-0.74
Current LTV	-0.04	-0.13	0.09	0.08	-0.09	0.02	0.03	0.04
State Productivity	0.00	0.00†	0.00	0.00†	0.00	0.00	0.00	0.00
Mortgage Term < 30 years	0.88	-1.06	0.11	0.09†	0.60	-0.12	-0.20	-0.29
Baseline FICO: 850 - 750								
FICO: 750 - 720	-0.21†	-0.87	-0.11	1.18	-0.44	0.07	0.15	0.22
FICO: 720 - 660	-0.63†	-1.36	-0.23	2.21	-1.25	0.14	0.45	0.66
FICO: 660 - 620	-1.37†	-1.15	-0.50	3.00	-2.11	-0.49	1.03	1.56
FICO: 620 - 300	-2.09†	-0.99	-0.59	3.66	-3.06	-2.43	2.13	3.35
Baseline: Jan - Mar								
Apr - Jun	0.58	0.21	-0.27	-0.51†	0.84	-0.22	-0.25	-0.37
Jul - Sep	0.61	-0.16	-0.14	-0.31†	0.68	-0.17	-0.21	-0.30
Oct - Dec	0.30	0.11	-0.34	-0.07	0.58	-0.14	-0.18	-0.26
Baseline: Standard Refinancing								
Equity Refinancing	-0.53	-0.55	0.38	0.70	-0.78	0.09	0.28	0.41
Purchase	-0.31	-0.54	0.79	0.06	-0.54	0.08	0.19	0.27
Relocation	-1.29	0.54	-2.23	-0.34	-1.21	0.02	0.48	0.71

† Coefficient is not statistically significant.

For loans that are curtailed at the beginning of a period, the probability tradeoffs for the ordered logit model is centered around curtailment. The substitution patterns under the multinomial logit model do not match those observed under the ordered logit model. For many of the key variables, an increase in the probability of prepayment is offset primarily by a decrease in the probability of maintaining a curtailed status. A puzzling result is that under the multinomial logit model the FICO score has no effect on the probability of prepayment.

Table 2.5: Average Estimated Derivatives of Transitions Probabilities from 30-Day Delinquency (MNL Model)

	Prepay	Curtail	Current	30-Day	60-Day	Bankruptcy
Age (Months)	0.13	-0.02	-0.45	0.34	0.00	0.00
LTV at Origination	0.12†	0.24	0.27	-0.46	-0.17	0.00†
Property Appraisal (per $100,000)	0.21†	0.02	0.53	-0.84	0.09†	0.00
Coupon Gap (1-month lag)	3.60	-3.07†	-9.78	13.36	-4.10	-0.01†
Coupon Gap (3-month lag)	0.30†	-0.37	0.68	-3.61	3.00	0.00†
Slope of Yield Curve (1-month lag)	1.80	1.60	2.65	-3.87	-2.20	0.01†
Current LTV	-0.12†	-0.31	-0.47	0.66	0.24	0.00
State Productivity	0.01	0.00†	0.00	-0.01	0.00	0.00†
Mortgage Term < 30 years	0.81	1.21	-2.11	3.74	-3.63	-0.02†
Baseline FICO: 850 - 750						
FICO: 750 - 720	0.04†	0.70	-0.31	-0.30†	-0.22†	0.09†
FICO: 720 - 660	-1.09	1.54	-0.12	-0.87†	0.45	0.10†
FICO: 660 - 620	-2.37	1.54	0.29	-1.44	1.85	0.12†
FICO: 620 - 300	-3.57	2.28	-0.35	-1.77†	3.31	0.11†
Baseline: Jan - Mar						
Apr - Jun	0.54†	0.50	2.73	-2.09	-1.67	-0.01†
Jul - Sep	0.55†	0.43†	2.40	-3.18	-0.20	-0.01†
Oct - Dec	0.88	0.66	4.38	-8.60	2.67	0.01†
Baseline: Standard Refinancing						
Equity Refinancing	-0.85	-0.42	0.30	-0.13†	1.11	-0.01†
Purchase	-0.50	-1.28†	-3.46	4.11	1.14	0.00†
Relocation	-1.96†	1.59†	-14.67	-11.26†	-0.13	0.00†

† Coefficient is not statistically significant.

For loans in 30-day delinquency, the negative consequences of the IIA property are particularly strong. A decrease in the FICO score leads to an increase in the probability of entering 60-day delinquency *and* an increase in the probability of curtailing the loan. It seems doubtful that a borrower with poor credit would be equally likely to miss an additional payment as to receive sufficient extra income to not only recover but actually curtail their mortgage.

Table 2.6: Average Estimated Derivative of Transition Probabilities from 60-Day Delinquency (MNL Model)

	Prepay	Curtail	Current	30-Day	60-Day	90-Day	Bankruptcy	Foreclosure
Age (Months)	0.28	-0.08	-0.53	0.01	0.39	-0.10	0.01	0.01
LTV at Origination	0.00†	0.15	-0.13	-0.06†	0.32	-0.28†	0.01†	-0.01†
Property Appraisal (per $100,000)	0.02†	0.26†	0.15	-0.69	-0.36†	0.56	0.04	0.02†
Coupon Gap (1-month lag)	1.00	-0.08†	3.22	4.59	5.24	-14.01	0.10†	-0.07†
Coupon Gap (3-month lag)	-1.22†	-0.89†	-3.54	-1.64	-4.26	11.31	-0.09†	0.33
Slope of Yield Curve (1-month lag)	2.38†	1.89†	3.59	1.47	1.64	-11.24	0.13†	0.13†
Current LTV	-0.06†	-0.16	0.04	-0.21†	-0.12†	0.48†	0.01†	0.02†
State Productivity	0.01†	0.00†	0.01	-0.03	0.02†	-0.01	0.00	0.00†
Mortgage Term < 30 years	-2.76	2.45†	2.46	5.01†	2.49	-10.06	0.24	0.17†
Baseline FICO: 850 - 750								
FICO: 750 - 720	0.85†	1.48†	2.12	1.43†	-2.89	-2.97	-0.09†	0.09†
FICO: 720 - 660	-1.09	0.86†	1.42	1.88†	-2.60	-1.93	1.45†	0.00†
FICO: 660 - 620	-1.77†	-0.38†	-0.19†	2.38†	-1.77†	0.19†	1.66†	-0.12†
FICO: 620 - 300	-1.48†	-0.16†	0.85†	-1.84†	-3.19†	4.49†	1.44†	-0.12†
Baseline: Jan - Mar								
Apr - Jun	-1.06	0.46†	2.75	5.09†	0.86	-4.22	-0.10†	-3.78†
Jul - Sep	-1.26	-0.38†	1.84	-1.48	0.38†	1.42†	-0.16†	-0.35
Oct - Dec	1.39	-2.00†	-2.52	-4.29†	-1.83	9.61	-0.01†	-0.36†
Baseline: Standard Refinancing								
Equity Refinancing	0.78†	-1.16†	0.57†	-2.70†	1.02†	1.58†	0.09†	-0.18†
Purchase	0.48†	-0.79†	0.70	2.24†	-0.99†	0.44†	-1.71†	-0.37
Relocation	-0.30	1.00	6.44	-3.37	2.99	-2.27	-0.08†	-3.57†

† Coefficient is not statistically significant.

A serious drawback of the multinomial logit model is that it cannot handle small event samples. In this analysis, the multinomial logit model was not able to estimate coefficients for any state below 60-day delinquency. Delinquency and default are not frequent events, but it is crucial to accurately forecast the likelihood of these events.

Table 2.7: Average Estimated Derivative of Transition Probabilities from 30-Day Delinquency (OLG Model)

	Prepay	Curtail	Current	30-Day	60-Day	Bankruptcy
Age (Months)	-0.04	-0.05	-0.08	0.15	0.02	0.00
LTV at Origination	0.19	0.21	0.34	-0.64	-0.11	0.00
Property Appraisal (per $100,000)	0.10	0.11	0.17	-0.32	-0.05	0.00
Coupon Gap (1-month lag)	-0.63	-0.65	-0.93	1.93	0.28	0.00
Coupon Gap (3-month lag)	-0.57	-0.62	-1.01	1.89	0.31	0.00
Slope of Yield Curve (1-month lag)	2.19	2.40	3.86	-7.24	-1.21	0.00
Current LTV	-0.27	-0.29	-0.47	0.89	0.15	0.00
State Productivity	0.00	0.00	0.00	-0.01	0.00	0.00
Mortgage Term < 30 years	0.74	0.81	1.31	-2.45	-0.41	0.00
Baseline FICO: 850 - 750						
FICO: 750 - 720	0.25	0.27	0.43	-0.82	-0.13	0.00
FICO: 720 - 660	0.16	0.17	0.28	-0.52	-0.09	0.00
FICO: 660 - 620	-0.29	-0.32	-0.53	0.98	0.17	0.00
FICO: 620 - 300	-0.99	-1.12	-1.93	3.40	0.63	0.00
Baseline: Jan - Mar						
Apr - Jun	1.11	1.19	1.84	-3.57	-0.56	0.00
Jul - Sep	0.68	0.74	1.16	-2.22	-0.36	0.00
Oct - Dec	0.33	0.35	0.56	-1.07	-0.17	0.00
Baseline: Standard Refinancing						
Equity Refinancing	-0.51	-0.56	-0.92	1.69	0.29	0.00
Purchase	-1.35	-1.52	-2.56	4.64	0.78	0.00
Relocation	2.52	2.59	3.61	-7.68	-1.04	0.00

† Coefficient is not statistically significant.

The dominant propensity is for a borrower to miss a payment in one month and then resume his payment schedule in the following month. Under the ordered logit model, an increase in the coupon gap leads to a *decrease* in the probability of prepayment from 30-day delinquency. This suggests that prepayments observed from 30-day delinquency may not be the result of refinancing but rather property sale.

Table 2.8: Average Estimated Derivative of Transition Probabilities from 60-Day Delinquency (OLG Model)

	Prepay	Curtail	Current	30-Day	60-Day	90-Day	Bankruptcy	Foreclosure
Age (Months)†	-0.02	-0.01	-0.01	-0.02	0.01	0.05	0.00	0.00
LTV at Origination†	0.03	0.02	0.03	0.05	-0.02	-0.10	0.00	0.00
Property Appraisal (per $100,000)†	-0.05	-0.04	-0.05	-0.08	0.04	0.17	0.00	0.00
Coupon Gap (1-month lag)	3.24	2.48	2.86	5.08	-1.25	-11.95	-0.14	-0.33
Coupon Gap (3-month lag)	-3.12	-2.38	-2.73	-4.51	2.46	9.99	0.09	0.20
Slope of Yield Curve (1-month lag)	4.01	3.07	3.51	5.80	-3.17	-12.85	-0.11	-0.26
Current LTV	-0.12	-0.09	-0.10	-0.17	0.09	0.38	0.00	0.01
State Productivity†	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Mortgage Term < 30 years	3.37	2.57	2.95	4.87	-2.66	-10.79	-0.09	-0.21
Baseline FICO: 850 - 750								
FICO: 750 - 720	1.82	1.33	1.45	2.06	-1.54	-5.00	-0.04	0.00
FICO: 720 - 660	1.22	0.92	1.04	1.65	-0.98	-3.73	-0.03	0.00
FICO: 660 - 620†	0.66	0.50	0.56	0.90	-0.52	-2.05	-0.02	0.00
FICO: 620 - 300†	-0.13	-0.10	-0.11	-0.19	0.10	0.41	0.00	0.00
Baseline: Jan - Mar								
Apr - Jun	1.53	1.14	1.28	1.95	-1.26	-4.53	-0.03	0.00
Jul - Sep†	-0.35	-0.27	-0.31	-0.52	0.27	1.15	0.01	0.00
Oct - Dec	-2.31	-1.85	-2.22	-4.26	1.48	8.95	0.06	0.00
Baseline: Standard Refinancing								
Equity Refinancing†	-0.50	-0.39	-0.45	-0.77	0.39	1.68	0.01	0.00
Purchase†	0.22	0.17	0.20	0.32	-0.18	-0.72	-0.01	0.00
Relocation	1.48	1.08	1.19	1.74	-1.21	-4.17	-0.03	0.00

† Coefficient is not statistically significant.

A borrower that is 60 days delinquent faces the critical choice of whether to pay off the past due amount or enter 90-day delinquency and be threatened with foreclosure. For loans that are in 60-day delinquency only age, the coupon gap, the mortgage term, the quarterly dummy variables, and the indicator for relocation are relevant to the probability of prepayment. The significant results that are observed appear to be a result of the model's substitution pattern, rather than true effects.

Table 2.9: Average Estimated Derivative of Transition Probabilities from 90-Day Delinquency (OLG Model)

	Prepay	Curtail	Current	30-Day	60-Day	90-Day	120-Day	Bankruptcy	Foreclosure
Age (Months)†	0.01	0.01	0.02	0.02	0.02	0.00	-0.05	0.00	-0.03
LTV at Origination†	0.03	0.03	0.06	0.09	0.08	0.01	-0.17	0.00	-0.12
Property Appraisal (per $100,000)†	-0.01	-0.01	-0.03	-0.04	-0.04	-0.01	0.08	0.00	0.06
Coupon Gap (1-month lag)	0.29	0.27	0.62	0.93	0.96	0.62	-1.50	-0.07	-2.13
Coupon Gap (3-month lag)†	-0.36	-0.34	-0.79	-1.13	-1.04	-0.16	2.22	0.06	1.54
Slope of Yield Curve (1-month lag)	0.60	0.56	1.30	1.86	1.71	0.26	-3.65	-0.10	-2.53
Current LTV	-0.05	-0.04	-0.10	-0.15	-0.13	-0.02	0.29	0.01	0.20
State Productivity†	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00
Mortgage Term $<$ 30 years	1.10	1.03	2.40	3.41	3.15	0.47	-6.71	-0.19	-4.65
Baseline FICO: 850 - 750									
FICO: 750 - 720†	-0.31	-0.29	-0.66	-0.98	-0.95	-0.24	1.94	0.06	1.42
FICO: 720 - 660†	0.30	0.28	0.66	0.92	0.83	0.09	-1.81	-0.05	-1.23
FICO: 660 - 620†	0.14	0.13	0.30	0.42	0.38	0.05	-0.82	-0.02	-0.56
FICO: 620 - 300†	0.18	0.17	0.39	0.55	0.50	0.06	-1.08	-0.03	-0.74
Baseline: Jan - Mar									
Apr - Jun	0.74	0.68	1.62	2.20	1.90	0.02	-4.30	-0.11	-2.74
Jul - Sep†	0.20	0.18	0.43	0.60	0.55	0.06	-1.18	-0.03	-0.80
Oct - Dec†	-0.01	-0.01	-0.01	-0.02	-0.02	0.00	0.04	0.00	0.02
Baseline: Standard Refinancing									
Equity Refinancing†	-0.12	-0.12	-0.27	-0.39	-0.36	-0.07	0.77	0.02	0.54
Purchase†	0.24	0.23	0.53	0.75	0.68	0.09	-1.47	-0.04	-1.01
Relocation†	0.20	0.19	0.44	0.61	0.55	0.05	-1.20	-0.03	-0.81

† Coefficient is not statistically significant.

Tables 2.9 - 2.12 report the probability derivatives for transition from a state below 60-day delinquency under the ordered logit model. For loans in 90-day delinquency, the relevant factors are the one-month lag of the coupon gap, the slope of the yield curve, the contemporaneous LTV ratio, and the term of the mortgage. At 120-day delinquency the transition probabilities are influenced by loan age, the property appraisal at origination, and the one-month lag of the coupon gap. Loans in bankruptcy are more likely to prepay as the coupon gap *falls* and the slope of the yield curve rises. For loans in foreclosure, the only significant factors are the slope of the yield curve and the contemporaneous LTV ratio. The significant factors are likely to proxy for economic shocks that are not directly accounted for.

Table 2.10: Average Estimated Derivative of Transition Probabilities From 120-Day Delinquency (OLG Model)

	Prepay	Curtail	Current	30-Day	60-Day
Age (Months)	0.17	0.04	0.05	0.08	0.06
LTV at Origination†	-0.12	-0.03	-0.03	-0.06	-0.04
Property Appraisal (per $100,000)	-0.92	-0.23	-0.26	-0.45	-0.33
Coupon Gap (1-month lag)	2.18	0.53	0.61	1.05	0.76
Coupon Gap (3-month lag)†	-2.14	-0.53	-0.61	-1.06	-0.77
Slope of Yield Curve (1-month lag)†	-1.20	-0.30	-0.34	-0.59	-0.43
Current LTV†	-0.08	-0.02	-0.02	-0.04	-0.03
State Productivity†	0.03	0.01	0.01	0.02	0.01
Mortgage Term < 30 years†	0.39	0.10	0.11	0.19	0.14
	90-Day	120-Day	Bankruptcy	Foreclosure	
Age (Months)	0.12	-0.04	-0.02	-0.46	
LTV at Origination†	-0.09	0.03	0.01	0.34	
Property Appraisal (per $100,000)	-0.65	0.22	0.10	2.52	
Coupon Gap (1-month lag)	1.45	-0.80	-0.22	-5.57	
Coupon Gap (3-month lag)†	-1.51	0.51	0.23	5.87	
Slope of Yield Curve (1-month lag)†	-0.85	0.29	0.13	3.29	
Current LTV†	-0.05	0.02	0.01	0.21	
State Productivity†	0.02	0.00	0.00	0.00	
Mortgage Term < 30 years†	0.28	-0.09	-0.04	-1.07	

† Coefficient is not statistically significant.

Table 2.11: Average Estimated Derivative of Transition Probabilities From Bankruptcy (OLG Model)

	Prepay	Curtail	Current	30-Day	60-Day
Age (Months)	0.25	0.08	0.07	0.02	0.01
LTV at Origination†	0.00	0.00	0.00	0.00	0.00
Property Appraisal (per $100,000)†	-0.31	-0.09	-0.09	-0.02	-0.01
Coupon Gap (1-month lag)	-1.90	-0.58	-0.57	-0.13	-0.09
Coupon Gap (3-month lag)†	-1.34	-0.40	-0.39	-0.08	-0.06
Slope of Yield Curve (1-month lag)	4.86	1.46	1.40	0.31	0.23
Current LTV†	-0.13	-0.04	-0.04	-0.01	-0.01
State Productivity	0.03	0.01	0.01	0.00	0.00
Mortgage Term $<$ 30 years†	-0.35	-0.10	-0.10	-0.02	-0.02
	90-Day	120-Day	Bankruptcy	Foreclosure	
Age (Months)	0.00	0.02	-0.33	-0.13	
LTV at Origination†	0.00	0.00	0.00	0.00	
Property Appraisal (per $100,000)†	0.00	-0.02	0.40	0.16	
Coupon Gap (1-month lag)	-0.03	-0.15	1.70	1.76	
Coupon Gap (3-month lag)†	-0.02	-0.10	1.71	0.68	
Slope of Yield Curve (1-month lag)	0.08	0.37	-6.23	-2.47	
Current LTV†	-0.13	-0.01	0.16	0.06	
State Productivity	0.03	0.00	-0.04	0.00	
Mortgage Term $<$ 30 years†	-0.01	-0.03	0.45	0.18	

† Coefficient is not statistically significant.

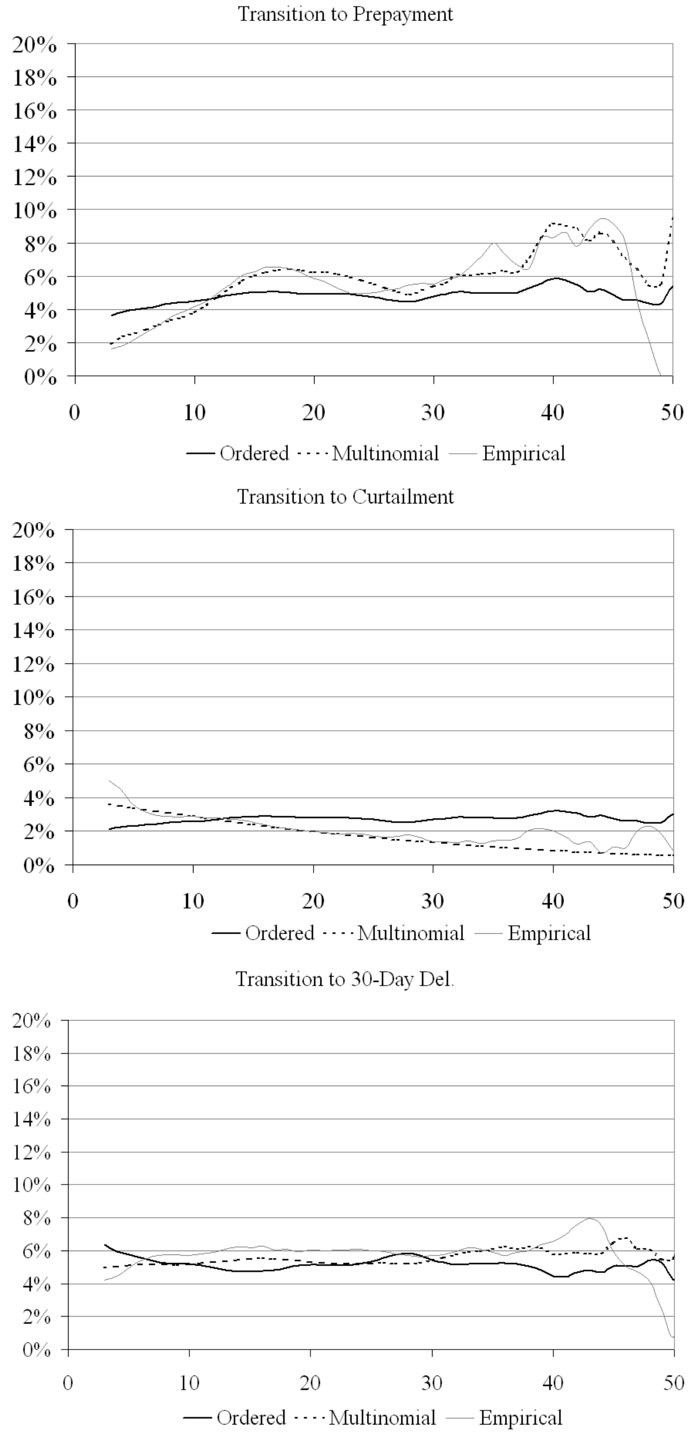
Table 2.12: Average Estimated Derivative of Transition Probabilities from Foreclosure (OLG Model)

	Prepay	Curtail	Current	30-Day	60-Day	90-Day
Age (Months)†	-0.07	-0.02	-0.02	-0.03	0.00	-0.01
LTV at Origination†	-0.01	0.00	0.00	0.00	0.00	0.00
Property Appraisal (per $100,000)†	-0.39	-0.09	-0.09	-0.17	-0.02	-0.05
Coupon Gap (1-month lag)†	0.80	0.19	0.19	0.38	0.04	0.11
Coupon Gap (3-month lag)†	-1.55	-0.36	-0.35	-0.68	-0.07	-0.19
Slope of Yield Curve (1-month lag)	2.43	0.56	0.54	1.06	0.11	0.29
Current LTV	-0.30	-0.07	-0.07	-0.13	-0.01	-0.04
State Productivity†	0.01	0.00	0.00	0.01	0.00	0.00
Mortgage Term $<$ 30 years†	1.78	0.41	0.40	0.78	0.08	0.21
	120-Day	Bankruptcy	Foreclosure	REO	Liquidation	
Age (Months)†	-0.03	-0.02	0.13	0.05	0.01	
LTV at Origination†	0.00	0.00	0.02	0.01	0.00	
Property Appraisal (per $100,000)†	-0.15	-0.11	0.73	0.26	0.07	
Coupon Gap (1-month lag)†	0.34	0.27	-1.60	-0.56	-0.16	
Coupon Gap (3-month lag)†	-0.60	-0.46	2.93	1.04	0.29	
Slope of Yield Curve (1-month lag)	0.94	0.72	-4.58	-1.63	-0.45	
Current LTV	-0.12	-0.09	0.57	0.20	0.06	
State Productivity†	0.01	0.00	-0.02	-0.01	0.00	
Mortgage Term $<$ 30 years†	0.69	0.53	-3.35	-1.19	-0.33	

† Coefficient is not statistically significant.

Figure 2.3: Transition Probabilities by Age(months) - From Current

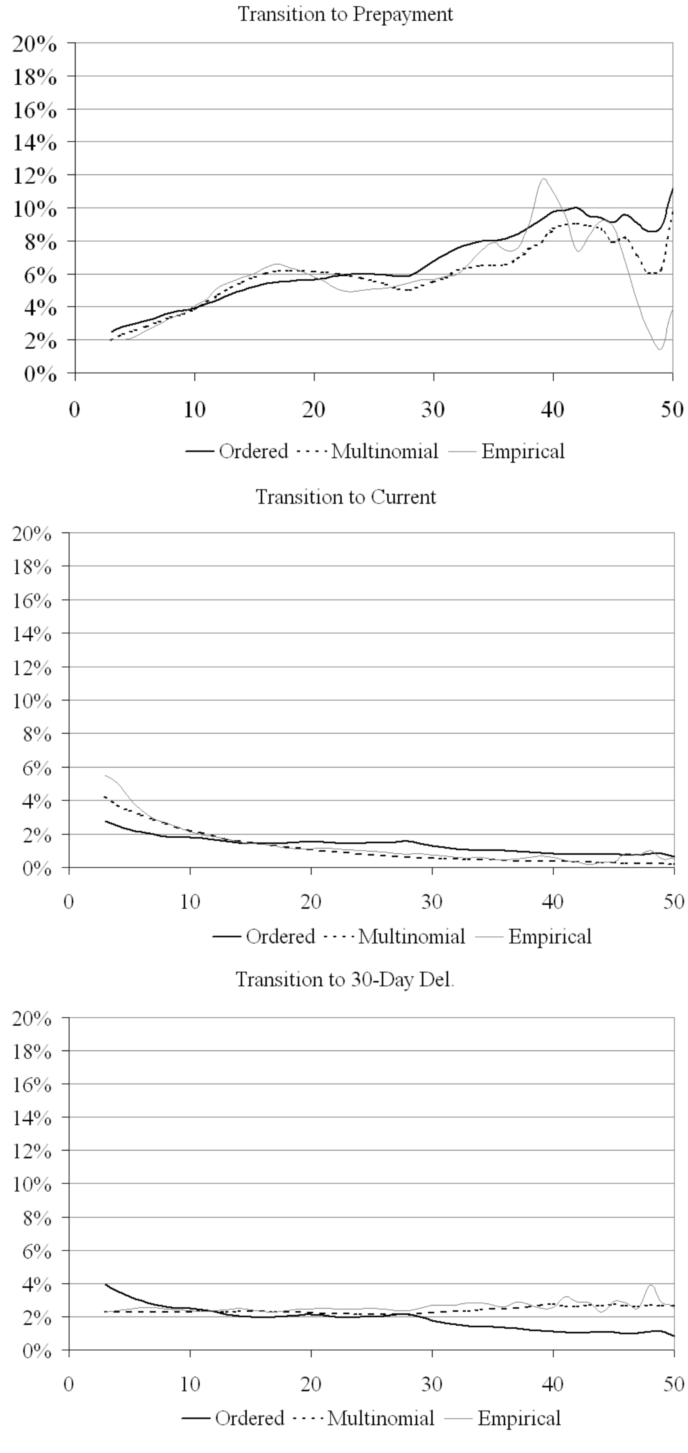
The probability of maintaining a current status is approximately 90%



One of the benefits of considering non-termination events is the ability to forecast transition matrices. Under the Markov property the transition matrices for an individual can be multiplied together to generate the probability of a borrower's choice stream. The curves in Figures 2.3-2.8 are averages over individuals for a given loan age.

Figure 2.4: Transition Probabilities by Age(months) - From Curtailed

The probability of maintaining a curtailed status is approximately 90%



Although multinomial logit models tend to match in-sample empirical probabilities very closely, the evidence given by the probability derivatives suggests that the ordered logit model will be more stable out-of-sample. While the multinomial logit model may capture the substitution patterns that are present in this dataset, the decision process is inherently misspecified.

Figure 2.5: Transition Probabilities by Age(months) - From 30-Day Delinquency to Recovery

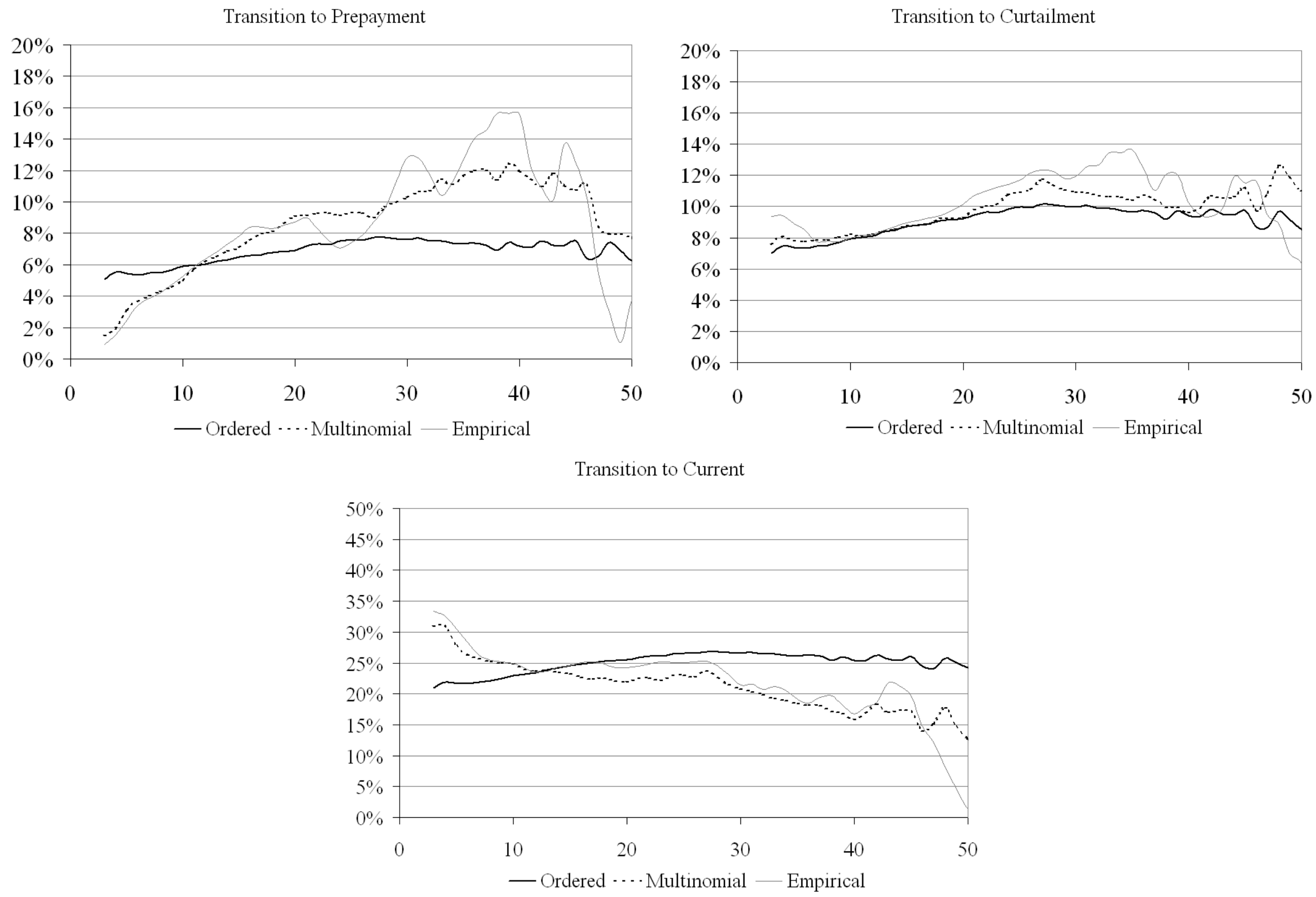


Figure 2.6: Transition Probabilities by Age(months) - From 30-Day Delinquency to Delinquency

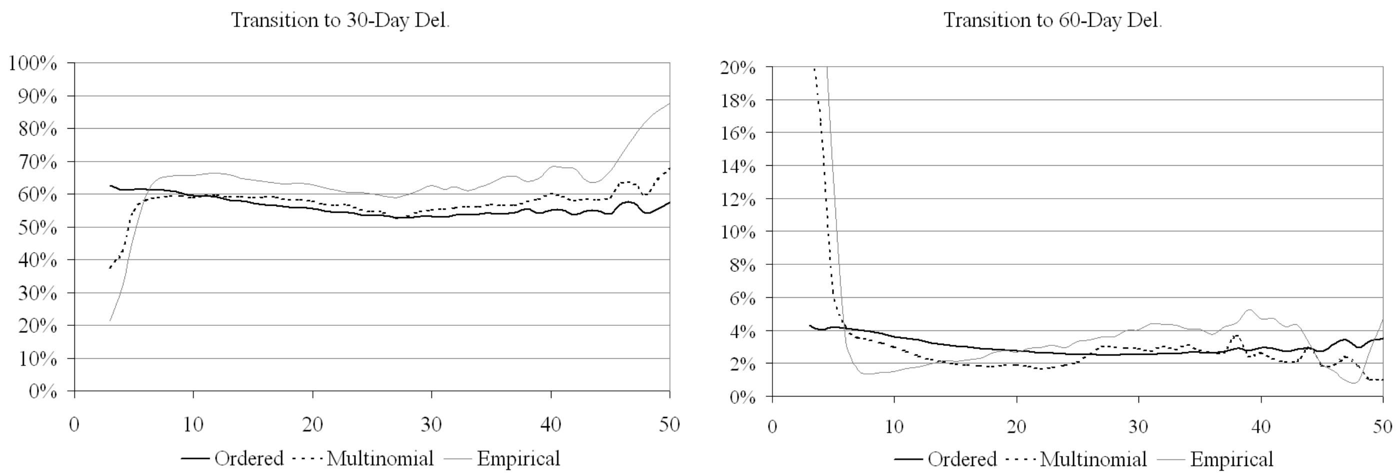


Figure 2.7: Transition Probabilities by Age(months) - From 60-Day Delinquency to Recovery

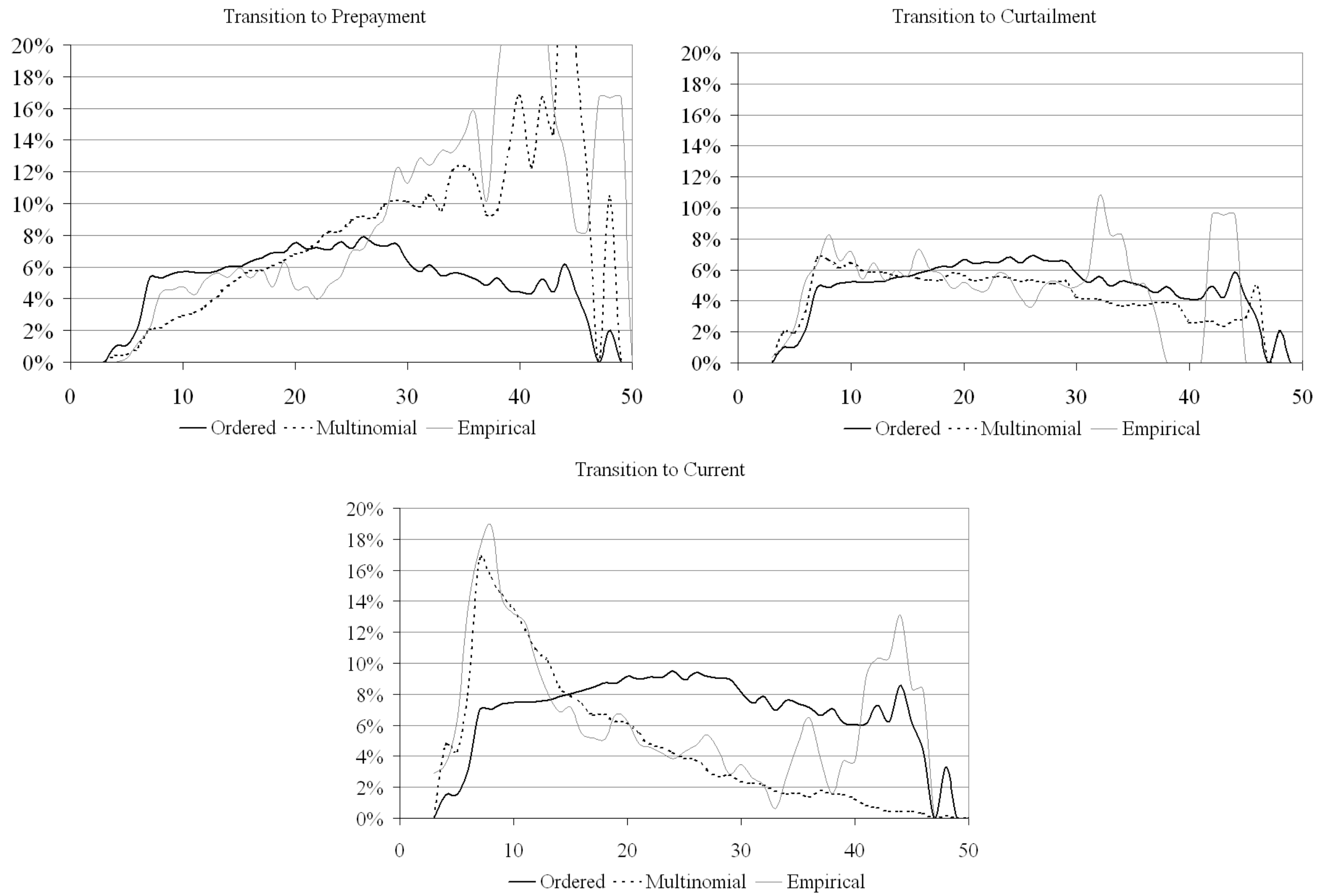
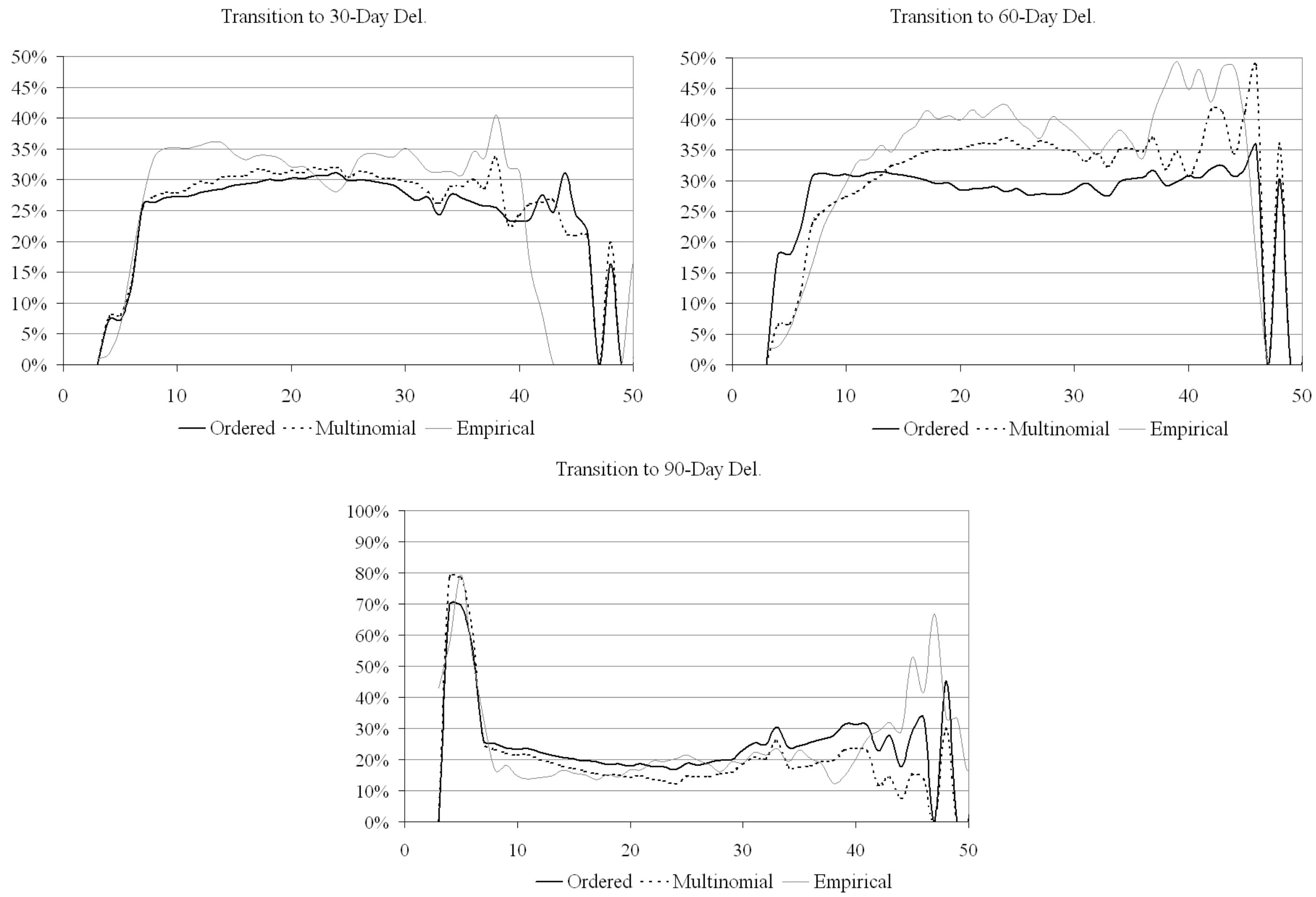


Figure 2.8: Transition Probabilities by Age(months) - From 60-Day Delinquency to Delinquency



Appendix - Coefficient Estimates

Coefficient estimates are reported in this appendix because the parameter set is quite large, and because interpretation of the coefficients is not intuitive. Instead, the focus in the text is on the probability derivatives that are calculated from these coefficient estimates. The estimation results from both the multinomial logit model and the ordered logit model provide evidence to support the claim that it is important to model non-termination events, and to condition the choice set and the utility space on the state of the mortgage at the beginning of the payment period. The significance and magnitude of individual coefficients differs markedly over the initial states. The factors considered in the analysis are similar to those found in most hazard rate models for prepayment. For the states most likely to be represented in those studies (curtailed, current, and 30 days delinquent), most of the factors are significant. As the states enter further into delinquency, however, fewer and fewer of the explanatory variables remain significant. The key difference between the multinomial logit model and the ordered logit model is that for the multinomial logit model the coefficients are specific to an initial state *and* a destination state. For the ordered logit model, all of the choices available from a given initial state are linked by a single set of coefficients.

Table 2.13: Coefficient Estimates I - OLG Model

	Curtail		Current		30-Day		60-Day	
Observations	451915		708733		144174		5457	
Log-likelihood	-167813		-352133		-163051		-7631	
LR Statistic	14737.5	***	12921.1	***	9964.6	***	1866.5	***
Age (Months)	-0.025	***	0.000		0.008	***	0.004	
	(0.001)		(0.001)		(0.001)		(0.005)	
LTV at	-0.014	***	-0.013	***	-0.034	***	-0.007	
Origination	(0.001)		(0.001)		(0.001)		(0.006)	
Property	-4.87E-08	***	-9.66E-08	***	-1.69E-07	***	1.12E-07	
Appraisal	(1.4E-08)		(9.2E-09)		(1.7E-08)		(9.7E-08)	
Coupon Gap	-0.552	***	-0.229	***	0.179	***	-1.307	***
(1-month lag)	(0.020)		(0.014)		(0.021)		(0.100)	
Coup. Gap	0.001		-0.006	***	-0.004	***	0.022	***
Cubed (1-month lag)	(0.001)		(0.001)		(0.001)		(0.003)	
Coupon Gap	-0.063	***	-0.097	***	0.099	***	0.655	***
(3-month lag)	(0.017)		(0.012)		(0.016)		(0.081)	
Slope of Yield	-0.337	***	-0.212	***	-0.378	***	-0.843	***
Curve (1-month lag)	(0.009)		(0.006)		(0.008)		(0.047)	
Current LTV	0.019	***	0.015	***	0.046	***	0.025	***
	(0.001)		(0.001)		(0.001)		(0.006)	
State Productivity	-1.40E-04	***	-6.00E-05		-3.00E-04	***	2.30E-04	
	(6.0E-05)		(7.0E-05)		(4.2E-05)		(2.0E-04)	
Mortgage Term	-0.131	***	-0.190	***	-0.128	***	-0.708	***
< 30 years	(0.014)		(0.010)		(0.015)		(0.090)	
Baseline FICO: 850 - 750								
FICO: 750 - 720	0.098	***	0.142	***	-0.043	***	-0.342	***
	(0.014)		(0.010)		(0.016)		(0.099)	
FICO: 720 - 660	0.285	***	0.381	***	-0.027	**	-0.247	***
	(0.013)		(0.009)		(0.013)		(0.082)	
FICO: 660 - 620	0.560	***	0.757	***	0.052	***	-0.135	
	(0.023)		(0.016)		(0.018)		(0.092)	
FICO: 620 - 300	0.968	***	1.140	***	0.184	***	0.027	
	(0.038)		(0.027)		(0.025)		(0.100)	
Baseline: January - March								
April -	-0.175	***	-0.192	***	-0.184	***	-0.301	***
June	(0.016)		(0.011)		(0.015)		(0.079)	
July -	-0.143	***	-0.186	***	-0.115	***	0.075	
September	(0.016)		(0.011)		(0.015)		(0.080)	
October -	-0.122	***	-0.044	***	-0.055	***	0.539	***
December	(0.015)		(0.011)		(0.015)		(0.076)	
Baseline: Standard Refinancing								
Equity Refinancing	0.177	***	0.140	***	0.089	***	0.109	
	(0.014)		(0.010)		(0.014)		(0.078)	
Purchase	0.121	***	-0.025	***	0.243	***	-0.047	
	(0.015)		(0.010)		(0.014)		(0.067)	
Relocation	0.288	***	-0.016		-0.380	***	-0.281	***
	(0.020)		(0.013)		(0.022)		(0.107)	

Table 2.14: Coefficient Estimates II - OLG Model

	90-Day		120-Day		Bankruptcy		Foreclosure	
Observations	1066		543		753		886	
Log-likelihood	-1885		-771		-488		-1098	
LR Statistic	98.7	***	44.5	***	61.9	***	62.2	***
Age (Months)	-0.003		-0.030	**	-0.045	***	0.011	
	(0.009)		(0.014)		(0.017)		(0.014)	
LTV at	-0.013		0.022		0.000		0.001	
Origination	(0.012)		(0.016)		(0.019)		(0.016)	
Property	6.52E-08		1.65E-06	***	5.48E-07		5.98E-07	
Appraisal	(2.6E-07)		(4.5E-07)		(5.3E-07)		(4.0E-07)	
Coupon Gap	-0.920	***	-0.840	**	-0.427		0.161	
(1-month lag)	(0.232)		(0.403)		(0.425)		(0.330)	
Coup. Gap	0.026	***	0.014		0.028	***	-0.008	
Cubed (1-month lag)	(0.006)		(0.009)		(0.011)		(0.008)	
Coupon Gap	0.170		0.384		0.235		0.239	
(3-month lag)	(0.165)		(0.246)		(0.292)		(0.218)	
Slope of Yield	-0.281	***	0.215		-0.853	***	-0.374	**
Curve (1-month lag)	(0.099)		(0.174)		(0.217)		(0.175)	
Current LTV	0.022	*	0.014		0.022		0.047	***
	(0.012)		(0.016)		(0.014)		(0.014)	
State Productivity	4.74E-04		-5.64E-03		-5.00E-03	*	-2.01E-03	
	(1.2E-03)		(4.0E-03)		(2.9E-03)		(2.8E-03)	
Mortgage Term	-0.516	***	-0.070		0.061		-0.274	
< 30 years	(0.198)		(0.343)		(0.408)		(0.272)	
Baseline FICO: 850 - 750								
FICO: 750 - 720	0.151		-0.253		0.878		-0.080	
	(0.277)		(0.473)		(0.603)		(0.391)	
FICO: 720 - 660	-0.139		-0.020		0.474		0.200	
	(0.224)		(0.344)		(0.520)		(0.320)	
FICO: 660 - 620	-0.063		-0.236		0.328		-0.108	
	(0.239)		(0.353)		(0.550)		(0.314)	
FICO: 620 - 300	-0.083		-0.115		0.308		-0.429	
	(0.241)		(0.347)		(0.630)		(0.331)	
Baseline: January - March								
April -	-0.324	**	-0.191		-0.261		-0.410	**
June	(0.166)		(0.245)		(0.342)		(0.205)	
July -	-0.090		0.463	*	-0.538		0.091	
September	(0.171)		(0.247)		(0.328)		(0.224)	
October -	0.003		0.039		-1.209	***	-0.313	
December	(0.163)		(0.243)		(0.317)		(0.206)	
Baseline: Standard Refinancing								
Equity Refinancing	0.059		0.043		-0.099		-0.034	
	(0.165)		(0.290)		(0.308)		(0.230)	
Purchase	-0.113		0.123		-0.114		-0.335	
	(0.146)		(0.230)		(0.281)		(0.208)	
Relocation	-0.092		-0.020		-0.187		-0.104	
	(0.231)		(0.336)		(0.396)		(0.310)	

Table 2.15: Coefficient Estimates (Transition From Curtailed) - MNL Model

	Prepay		Curtail		Current	30-Day		Bankruptcy	
Observations	451915								
Log-likelihood	-159185								
LR Statistic	31993.58	***							
Intercept	-5.086	***	3.442	***	-	-1.645	***	-8.846	***
	(0.133)		(0.109)			(0.148)		(2.504)	
Age (Months)	0.070	***	0.041	***	-	0.062	***	0.097	***
	(0.003)		(0.002)			(0.003)		(0.030)	
LTV at	0.055	***	0.049	***	-	0.023	***	0.088	***
Origination	(0.003)		(0.003)			(0.003)		(0.037)	
Property	-7.23E-08	**	-2.88E-07	***	-	-2.98E-07	***	-3.42E-06	***
Appraisal	(3.1E-08)		(2.5E-08)			(4.0E-08)		(1.3E-06)	
Coupon Gap	1.992	***	-0.081	*	-	-0.074		-0.640	
(1-month lag)	(0.059)		(0.048)			(0.062)		(0.737)	
Coup. Gap	-0.047	***	0.003		-	0.002		-0.001	
Cubed (1-month lag)	(0.003)		(0.002)			(0.003)		(0.024)	
Coupon Gap	0.104	**	0.082	**	-	0.084	*	0.449	
(3-month lag)	(0.043)		(0.038)			(0.049)		(0.518)	
Slope of Yield	0.704	***	0.163	***	-	0.169	***	0.163	
Curve (1-month lag)	(0.022)		(0.017)			(0.024)		(0.305)	
Current LTV	-0.065	***	-0.057	***	-	-0.019	***	-0.055	*
	(0.003)		(0.003)			(0.003)		(0.031)	
State Productivity	1.04E-03	***	-8.64E-06		-	-1.30E-04		4.90E-04	
	(3.4E-04)		(2.0E-04)			(2.1E-04)		(3.1E-03)	
Mortgage Term	-0.077	**	0.125	***	-	-0.025		-2.233	***
< 30 years	(0.033)		(0.038)			(0.043)		(0.830)	
Baseline FICO: 850 - 750									
FICO: 750 - 720	0.024		0.057	*	-	0.594	***	0.142	
	(0.036)		(0.031)			(0.044)		(0.916)	
FICO: 720 - 660	0.008		0.126	***	-	1.125	***	1.712	***
	(0.034)		(0.029)			(0.039)		(0.645)	
FICO: 660 - 620	0.015		0.293	***	-	1.643	***	2.394	***
	(0.065)		(0.059)			(0.068)		(0.717)	
FICO: 620 - 300	-0.084		0.349	***	-	1.989	***	2.529	***
	(0.112)		(0.103)			(0.113)		(0.856)	
Baseline: January - March									
April -	0.292	***	0.170	***	-	-0.061		-0.538	
June	(0.039)		(0.033)			(0.044)		(0.736)	
July -	0.220	***	0.086	***	-	-0.051		0.765	
September	(0.039)		(0.033)			(0.044)		(0.541)	
October -	0.270	***	0.207	***	-	0.176	***	0.548	
December	(0.039)		(0.034)			(0.044)		(0.577)	
Baseline: Standard Refinancing									
Equity Refinancing	-0.344	***	-0.236	***	-	0.082	**	0.554	
	(0.039)		(0.034)			(0.042)		(0.499)	
Purchase	-0.550	***	-0.488	***	-	-0.456	***	-0.253	
	(0.039)		(0.034)			(0.045)		(0.499)	
Relocation	-0.986	***	-0.700	***	-	-0.859	***	-10.053	
	(0.048)		(0.041)			(0.057)		(40.869)	

Table 2.16: Coefficient Estimates (Transition From Current) - MNL Model

	Prepay		Curtail		Current	30-Day		Bankruptcy	
Observations	708733								
Log-likelihood	-334536								
LR Statistic	48116.43	***							
Intercept	-8.975	***	-3.049	***	-	-3.551	***	-13.680	***
	(0.070)		(0.076)			(0.056)		(2.366)	
Age (Months)	0.022	***	-0.060	***	-	0.006	***	0.076	***
	(0.001)		(0.002)			(0.001)		(0.028)	
LTV at	0.021	***	0.030	***	-	-0.004	***	-0.006	
Origination	(0.002)		(0.002)			(0.001)		(0.041)	
Property	2.17E-07	***	-1.28E-07	***	-	-1.03E-07	***	-1.90E-06	*
Appraisal	(1.5E-08)		(2.1E-08)			(1.6E-08)		(1.1E-06)	
Coupon Gap	2.216	***	-0.015		-	0.222	***	-0.618	
(1-month lag)	(0.031)		(0.031)			(0.022)		(0.663)	
Coup. Gap	-0.051	***	0.000		-	0.001		0.007	
Cubed (1-month lag)	(0.001)		(0.002)			(0.001)		(0.022)	
Coupon Gap	0.087	***	0.166	***	-	0.026		0.736	
(3-month lag)	(0.018)		(0.023)			(0.018)		(0.501)	
Slope of Yield	0.557	***	0.068	***	-	-0.013		-0.104	
Curve (1-month lag)	(0.012)		(0.012)			(0.008)		(0.275)	
Current LTV	-0.024	***	-0.038	***	-	0.004	***	0.049	
	(0.002)		(0.002)			(0.001)		(0.038)	
State Productivity	2.40E-03	***	1.16E-04		-	3.09E-04	**	4.99E-03	
	(2.4E-04)		(1.9E-04)			(1.5E-04)		(6.4E-03)	
Mortgage Term	0.309	***	0.189	***	-	-0.025	*	-0.711	
< 30 years	(0.016)		(0.021)			(0.015)		(0.628)	
Baseline FICO: 850 - 750									
FICO: 750 - 720	-0.048	***	0.038	*	-	0.395	***	0.322	
	(0.016)		(0.020)			(0.016)		(0.765)	
FICO: 720 - 660	-0.152	***	0.024		-	0.756	***	1.757	***
	(0.015)		(0.019)			(0.014)		(0.556)	
FICO: 660 - 620	-0.285	***	-0.005		-	1.126	***	2.016	***
	(0.025)		(0.034)			(0.019)		(0.641)	
FICO: 620 - 300	-0.463	***	0.070		-	1.360	***	1.026	
	(0.043)		(0.060)			(0.029)		(1.132)	
Baseline: January - March									
April -	0.156	***	0.048	**	-	-0.248	***	0.782	
June	(0.017)		(0.022)			(0.016)		(0.541)	
July -	0.182	***	0.047	**	-	-0.206	***	0.598	
September	(0.018)		(0.022)			(0.016)		(0.565)	
October -	0.023		0.018		-	-0.007		0.314	
December	(0.017)		(0.022)			(0.015)		(0.600)	
Baseline: Standard Refinancing									
Equity Refinancing	-0.207	***	-0.004		-	0.159	***	-0.325	
	(0.017)		(0.021)			(0.014)		(0.498)	
Purchase	-0.133	***	0.237	***	-	-0.024		-0.602	
	(0.016)		(0.021)			(0.016)		(0.454)	
Relocation	-0.454	***	0.099	***	-	-0.391	***	-1.419	**
	(0.020)		(0.027)			(0.020)		(0.617)	

Table 2.17: Coefficient Estimates I (From 30-Day Delinquency) - MNL Model

	Prepay		Curtail		Current	30-Day	
Observations	144174						
Log-likelihood	-151400						
LR Statistics	33265.05	***					
Intercept	-8.210	***	-1.429	***	-	-1.539	***
	(0.162)		(0.106)			(0.074)	
Age (Months)	0.044	***	0.016	***	-	0.028	***
	(0.002)		(0.002)			(0.001)	
LTV at	0.004		0.016	***	-	-0.022	***
Origination	(0.003)		(0.002)			(0.002)	
Property	5.97E-08		-1.99E-07	***	-	-4.05E-07	***
Appraisal	(4.1E-08)		(3.2E-08)			(2.4E-08)	
Coupon Gap	2.566	***	0.060		-	1.148	***
(1-month lag)	(0.065)		(0.040)			(0.028)	
Coup. Gap	-0.055	***	0.000		-	-0.020	***
Cubed (1-month lag)	(0.002)		(0.002)			(0.001)	
Coupon Gap	-0.006		-0.076	**	-	-0.106	***
(3-month lag)	(0.037)		(0.032)			(0.021)	
Slope of Yield	0.161	***	0.077	***	-	-0.196	***
Curve (1-month lag)	(0.024)		(0.017)			(0.011)	
Current LTV	0.005		-0.017	***	-	0.035	***
	(0.003)		(0.003)			(0.002)	
State Productivity	1.73E-03	***	-1.70E-04		-	-3.70E-04	***
	(4.0E-04)		(1.1E-04)			(8.0E-05)	
Mortgage Term	0.257	***	0.233	***	-	0.176	***
< 30 years	(0.036)		(0.028)			(0.019)	
Baseline FICO: 850 - 750							
FICO: 750 - 720	0.017		0.093	***	-	0.006	
	(0.036)		(0.032)			(0.020)	
FICO: 720 - 660	-0.188	***	0.181	***	-	-0.024	
	(0.031)		(0.027)			(0.018)	
FICO: 660 - 620	-0.426	***	0.161	***	-	-0.063	***
	(0.043)		(0.037)			(0.024)	
FICO: 620 - 300	-0.607	***	0.270	***	-	-0.050	
	(0.061)		(0.052)			(0.034)	
Baseline: January - March							
April -	-0.042		-0.054	*	-	-0.166	***
June	(0.036)		(0.031)			(0.020)	
July -	-0.036		-0.051		-	-0.174	***
September	(0.036)		(0.031)			(0.020)	
October -	-0.106	***	-0.111	***	-	-0.379	***
December	(0.035)		(0.031)			(0.020)	
Baseline: Standard Refinancing							
Equity Refinancing	-0.159	***	-0.062	**	-	-0.020	
	(0.035)		(0.027)			(0.018)	
Purchase	0.096	***	-0.004		-	0.244	***
	(0.032)		(0.030)			(0.019)	
Relocation	-0.927		-0.309		-	-0.786	
	(0.053)		(0.042)			(0.027)	

Table 2.18: Coefficient Estimates II (From 30-Day Delinquency) - MNL Model

	60-Day		Bankruptcy	
Observations	144174			
Log-likelihood	-151400			
LR Statistics	33265.05	***		
Intercept	-1.703	***	-16.881	
	(0.181)		(33.620)	
Age (Months)	0.019	***	0.118	***
	(0.003)		(0.034)	
LTV at	-0.073	***	-0.057	
Origination	(0.005)		(0.058)	
Property	6.14E-08		-3.58E-06	*
Appraisal	(5.3E-08)		(1.9E-06)	
Coupon Gap	-1.793	***	-0.804	
(1-month lag)	(0.062)		(0.990)	
Coup. Gap	0.042	***	0.021	
Cubed (1-month lag)	(0.002)		(0.029)	
Coupon Gap	1.002	***	-0.230	
(3-month lag)	(0.056)		(0.713)	
Slope of Yield	-0.886	***	0.619	
Curve (1-month lag)	(0.023)		(0.473)	
Current LTV	0.103	***	0.101	*
	(0.005)		(0.053)	
State Productivity	-4.10E-04	***	8.65E-03	
	(1.1E-04)		(1.1E-02)	
Mortgage Term	-1.159	***	-1.743	
< 30 years	(0.063)		(1.172)	
Baseline FICO: 850 - 750				
FICO: 750 - 720	-0.066		7.289	
	(0.053)		(33.470)	
FICO: 720 - 660	0.155	***	8.294	
	(0.044)		(33.457)	
FICO: 660 - 620	0.620	***	9.598	
	(0.056)		(33.457)	
FICO: 620 - 300	1.150	***	8.740	
	(0.067)		(33.463)	
Baseline: January - March				
April -	-0.700	***	-1.135	
June	(0.059)		(0.910)	
July -	-0.178	***	-0.536	
September	(0.056)		(0.770)	
October -	0.715	***	0.668	
December	(0.048)		(0.619)	
Baseline: Standard Refinancing				
Equity Refinancing	0.371	***	-0.515	
	(0.047)		(0.836)	
Purchase	0.553	***	-0.170	
	(0.045)		(0.649)	
Relocation	-0.576	***	-0.175	
	(0.076)		(0.676)	

Table 2.19: Coefficient Estimates I (From 60-Day Delinquency) - MNL Model

	Prepay		Curtail		Current	30-Day	
Observations	5456						
Log-likelihood	-7109						
LR Statistic	2892.07	***					
Intercept	-5.022	***	-1.703	*	-	-0.191	
	(1.078)		(0.898)			(0.674)	
Age (Months)	0.147	***	0.065	***	-	0.086	***
	(0.017)		(0.018)			(0.014)	
LTV at	0.022		0.052	**	-	0.018	
Origination	(0.021)		(0.022)			(0.017)	
Property	-2.57E-07		3.22E-07		-	-5.36E-07	**
Appraisal	(3.4E-07)		(2.5E-07)			(2.1E-07)	
Coupon Gap	1.146	***	-0.414		-	0.249	
(1-month lag)	(0.390)		(0.309)			(0.238)	
Coup. Gap	-0.049	***	-0.005		-	-0.023	***
Cubed (1-month lag)	(0.011)		(0.009)			(0.007)	
Coupon Gap	0.280		0.344		-	0.490	**
(3-month lag)	(0.260)		(0.257)			(0.196)	
Slope of Yield	-0.053		-0.137		-	-0.507	***
Curve (1-month lag)	(0.180)		(0.150)			(0.113)	
Current LTV	-0.019		-0.040	*	-	-0.014	
	(0.021)		(0.023)			(0.018)	
State Productivity	-4.00E-05		-1.99E-03		-	-3.49E-03	**
	(2.9E-03)		(2.4E-03)			(1.8E-03)	
Mortgage Term	-0.954	***	0.145		-	-0.207	
$<$ 30 years	(0.317)		(0.240)			(0.186)	
Baseline FICO: 850 - 750							
FICO: 750 - 720	-0.173		-0.009		-	-0.288	
	(0.328)		(0.316)			(0.240)	
FICO: 720 - 660	-0.469	*	-0.036		-	-0.167	
	(0.281)		(0.271)			(0.203)	
FICO: 660 - 620	-0.347		-0.055		-	0.115	
	(0.308)		(0.309)			(0.228)	
FICO: 620 - 300	-0.477		-0.170		-	-0.225	
	(0.325)		(0.332)			(0.244)	
Baseline: January - March							
April -	-0.688	***	-0.338		-	-0.253	
June	(0.256)		(0.233)			(0.176)	
July -	-0.571	**	-0.368		-	-0.358	**
September	(0.251)		(0.242)			(0.182)	
October -	0.674	***	-0.044		-	0.237	
December	(0.240)		(0.253)			(0.189)	
Baseline: Standard Refinancing							
Equity Refinancing	0.073		-0.333		-	-0.191	
	(0.249)		(0.237)			(0.175)	
Purchase	-0.011		-0.278		-	-0.014	
	(0.224)		(0.221)			(0.165)	
Relocation	-1.003	***	-0.659	**	-	-1.050	***
	(0.332)		(0.311)			(0.234)	

Table 2.20: Coefficient Estimates II (From 60-Day Delinquency) - MNL Model

	60-Day		90-Day		Bankruptcy		Foreclosure	
Observations	5456							
Log-likelihood	-7109							
LR Statistic	2892.07	***						
Intercept	-2.004	***	1.121		-28.020		-11.009	***
	(0.695)		(0.694)		(195.500)		(4.141)	
Age (Months)	0.102	***	0.078	***	0.131	**	0.137	***
	(0.014)		(0.015)		(0.061)		(0.047)	
LTV at	0.033	*	0.003		0.056		-0.014	
Origination	(0.017)		(0.018)		(0.079)		(0.063)	
Property	-4.02E-07	*	9.66E-08		2.82E-06	**	2.37E-07	
Appraisal	(2.2E-07)		(2.1E-07)		(1.2E-06)		(9.6E-07)	
Coupon Gap	0.545	**	-1.628	***	0.306		0.683	
(1-month lag)	(0.241)		(0.244)		(1.600)		(1.468)	
Coup. Gap	-0.033	***	0.011		-0.006		-0.044	
Cubed (1-month lag)	(0.007)		(0.007)		(0.042)		(0.035)	
Coupon Gap	0.392	**	1.270	***	-0.122		1.594	**
(3-month lag)	(0.195)		(0.209)		(1.244)		(0.737)	
Slope of Yield	-0.501	***	-1.280	***	0.372		-0.104	
Curve (1-month lag)	(0.114)		(0.115)		(0.861)		(0.740)	
Current LTV	-0.010		0.025		0.082		0.048	
	(0.018)		(0.018)		(0.073)		(0.063)	
State Productivity	-1.27E-03		-2.95E-03	*	-4.03E-03	*	7.03E-03	
	(1.8E-03)		(1.8E-03)		(2.3E-03)		(1.3E-02)	
Mortgage Term	-0.316	*	-1.033	***	1.309		0.167	
$<$ 30 years	(0.190)		(0.215)		(1.125)		(0.852)	
Baseline FICO: 850 - 750								
FICO: 750 - 720	-0.456	*	-0.533	**	-0.990		-0.060	
	(0.241)		(0.245)		(327.200)		(1.264)	
FICO: 720 - 660	-0.341	*	-0.359	*	9.867		-0.211	
	(0.203)		(0.207)		(195.400)		(1.121)	
FICO: 660 - 620	-0.047		0.037		11.547		-0.345	
	(0.227)		(0.237)		(195.400)		(1.189)	
FICO: 620 - 300	-0.275		0.138		9.901		-0.525	
	(0.242)		(0.251)		(195.400)		(1.234)	
Baseline: January - March								
April -	-0.429	**	-0.706	***	-1.232		-12.811	
June	(0.178)		(0.198)		(1.192)		(160.000)	
July -	-0.288		-0.202		-1.434		-1.466	**
September	(0.182)		(0.198)		(1.273)		(0.698)	
October -	0.341	*	1.011	***	0.300		-0.772	
December	(0.189)		(0.198)		(0.897)		(0.697)	
Baseline: Standard Refinancing								
Equity Refinancing	-0.043		0.015		0.525		-0.686	
	(0.176)		(0.185)		(0.855)		(0.707)	
Purchase	-0.143		-0.074		-11.999		-1.356	**
	(0.166)		(0.174)		(119.000)		(0.631)	
Relocation	-0.808	***	-1.041	***	-1.540		-12.616	
	(0.226)		(0.249)		(1.297)		(213.700)	

Chapter 3

Focus on Unobserved Heterogeneity and Correlation Over Time

3.1 Introduction

A thorough understanding of mortgage borrower behavior is essential for investors and originators of residential mortgages, as well as policy makers looking to enact market regulations. Accurate models of borrower behavior are rare, however, because of the complicated dynamics that incorporate both financial and behavioral elements. Reduced-form mortgage valuation models, in particular, commonly make the assumption that borrower behavior is independent in each payment period. This assumption is made to facilitate computation, but it is not an accurate representation of the borrower's decision process. In this chapter, it is shown that the use of random coefficients improves the performance of an ordered logit model for predicting borrower behavior, by accounting for unobserved heterogeneity and correlation in behavior over time.

The previous chapters argue that a random utility model is a useful framework for analyzing borrower behavior. A discrete choice panel data model is a straightforward method for incorporating time-varying covariates and correlated competing risks, and the latent utility process establishes a structural relationship between the borrower's decision and key factors. A random utility model also accounts for past behavior by conditioning the choice set and the utility space on the state of the mortgage at the beginning of the payment period. An ordered logit model is particularly well suited for modeling borrower behavior. Since this model is based on a single utility function, an ordered logit model is able to circumvent the independence from irrelevant alternatives (IIA) property, a weakness of the multinomial logit models commonly used with respect to subprime mortgages, without increasing the dimensionality of the model. This becomes a crucial advantage when random coefficients are used.[1]

The decision made by a borrower with respect to his mortgage is influenced by past events and attitudes, as well as the historical path of key factors, such as interest rates. This causes correlation in the borrower's behavior over time, a feature that is ignored by many models. Much of the relevant correlation is due to unobserved factors — preferences, family structure, financial savvy — that account for a high degree of heterogeneity in borrower behavior. Random coefficients can be used to model unobserved heterogeneity by specifying a component of the error term in the utility function that is randomly distributed over individ-

[1]Since there are several mortgage events and a substantial number of covariates, the simulation methods required to estimate a model with random coefficients can be quite complicated for other models that circumvent the IIA property, such as a multinomial probit model, where the decision process relies on a multivariate distribution.

uals. Since this distribution is constant over time, the coefficients also serve to induce correlation over a borrower's choice stream.

The use of a random intercept is prevalent in the literature on modeling unobserved heterogeneity in discrete choice models, and has been used in several studies with respect to mortgage borrower behavior. Han and Hausman (1990) add a heterogeneity term, ω, to the log-hazard rate specification in their proportional hazard rate competing risks model.

$$\delta_t = log \int_0^t \lambda_0(\tau) d\tau = \beta v + \varepsilon + \omega$$

This term is assumed to follow a discrete mass point distribution to represent different "types" of agents. Deng, Quigley, and Van Order (2000) apply this model to analyze the prepayment and default behavior of Ginnie Mae mortgage borrowers. They find that a model that accounts for unobserved heterogeneity yields more accurate probability estimates than simpler proportional hazard rate models lacking this feature.

Unobserved heterogeneity can also be modeled through transaction costs, both monetary and intangible. Intangible costs, such as the time spent searching for a new home or access to refinancing opportunities, are believed to account for much of the heterogeneity among borrowers. Downing, Stanton, and Wallace (2005), for example, model unobserved heterogeneity through transaction costs that are proportional to the remaining balance on the mortgage. In their specification, unobserved heterogeneity does not enter as a random intercept, but instead is associated with one of the covariates of the model. Transaction costs could also vary with variables other than the remaining balance, such as the contract rate on the mortgage or credit quality. While heterogeneity that enters through transaction costs is unobserved, it may influence the latent process through observed factors. Therefore, it is natural to extend the prevailing methods for modeling unobserved heterogeneity by including random coefficients, as well as a random intercept.

Random coefficients logit models, also known as mixed logit models, are common in studies of demand estimation and transportation science because of the important role that agents' tastes and preferences play in these areas. The implementation of these models has been facilitated by improvements in computer power and advances in simulation-based estimators. Early random coefficients models, such as Ben-Akiva and Bolduc (1996), used random coefficients in cross-sectional studies to allow for variation in the effects of a few key factors. Revelt and Train (1998) and Bhat (1999, 2000) explore random coefficients models with repeated choice-making by the individuals under analysis.[2] The use of random coefficients in samples with repeated choices allows the researcher to capture heterogeneity that arises from behavioral factors such as variation in preferences.

Most studies that use random coefficients to account for unobserved heterogeneity and taste variation find that controlling for these unobserved effects is essential to accurately predict agents' behavior. While a random coefficients model has not been previously used with respect to mortgages, the behavior of mortgage borrowers shares many of the fundamental aspects of areas in which random coefficients have been applied successfully. Previous studies have found that variation

[2]Bhat (1999) proposes an ordered logit model with random coefficients.

in mortgage borrower behavior cannot be explained entirely by financial dynamics. Since mortgage payment is a repeated event that is made at regular intervals, the borrower's decision is likely to be influenced to a large degree by unobserved factors.

In this chapter, a comparison is made between an ordered logit model with fixed coefficients and an ordered logit model with random coefficients. Two versions of the random coefficients model are estimated: one with the standard specification of independent coefficients, and one where the coefficients are correlated to account for the full path of borrower behavior, including changes in the status of the mortgage. Section 3.2 outlines the details of the models. To illustrate the features of each model a subsample is used. Section 3.3 shows that while the subsample represents only a small portion of the total sample, the origination and payment trends are consistent with those discussed in the previous chapters. The results in section 3.4 indicate that random coefficients are necessary for accurately modeling borrower behavior, and that correlation between coefficients needs to be considered.

3.2 Models

An ordered logit model is well suited for the analysis of mortgage borrower behavior because the choices faced by a borrower can be set on a continuum defined relative to the amortization schedule.[3] All mortgages events, including

[3]This term is used throughout to convey fluidity in a borrower's choice stream. It is not used in the literal sense of a continuous scale since the choice set is discrete.

non-termination events such as curtailment and delinquency, are ordered either by the payment required to achieve that state or by a natural time dependence. This ordering is shown in Figure 2.1 (Chapter 2). The goal is to model the behavior of the borrower as he moves along the continuum. When the ends of the continuum are reached, the mortgage terminates, through either prepayment or default.

It is assumed that in each payment period the borrower makes a choice related to his mortgage. Each choice, or transition, is defined by an initial state and a destination state. Given an initial state, i, a transition to an event available to the borrower over that period is driven by a single latent utility process, $U_{i,nt}$. This process can be thought of as a propensity for mortgage payment. The researcher observes a response variable, $y_{i,nt}$, that indicates which state on the continuum is chosen in that period. A choice towards the left of the continuum corresponds to a higher value of the latent process. For J possible choices, the latent process is divided into J contiguous segments separated by $J-1$ unobserved thresholds.

$$y_{i,nt} = j \quad if \quad k_{ij+1} < U_{i,nt} < k_{ij}, \quad j = 0...J-1, \quad k_{i0} = \infty, k_{iJ} = -\infty$$

Utility is specified as a linear function of observed and unobserved characteristics.

$$U_{i,nt} = \beta_i v_{nt} + \varepsilon_{i,nt}$$

For a logit model, the error term is assumed to follow an independent extreme value distribution. Under the assumption of fixed coefficients, the probability of moving from state i to an interior state j in a single payment period is given by the difference of two logistic functions. Similar formulas can be derived for the tail events.

$$
\begin{aligned}
P_{ij,nt} &= Prob(k_{ij+1} < U_{i,nt} < k_{ij}) \\
&= Prob(k_{ij+1} - \beta_i v_{nt} < \varepsilon_{i,nt} < k_{ij} - \beta_i v_{nt}) \\
&= \frac{\exp\left(k_{ij} - \beta_i v_{nt}\right)}{1 + \exp\left(k_{ij} - \beta_i v_{nt}\right)} - \frac{\exp\left(k_{ij+1} - \beta_i v_{nt}\right)}{1 + \exp\left(k_{ij+1} - \beta_i v_{nt}\right)}
\end{aligned}
$$

3.2.1 Independent Random Coefficients

The assumption underlying a fixed coefficients model is that all previous behavior is captured by the choice in the preceding period, and correlation over time is induced only through time-invariant covariates. Random coefficients are used to capture correlation in behavior that is a result of unobserved heterogeneity, taste variation, and path dependence. In a random coefficients model, the coefficients that enter the utility function are borrower-specific random variables drawn from a multivariate normal distribution.

$$U_{i,nt} = \alpha_{ni} + \beta_{ni} v_{nt} + \varepsilon_{i,nt}$$

$$\{\alpha_{ni}, \beta_{ni}\} \sim N(\mu, \Omega).$$

μ is a vector representing the mean of the coefficients (and intercepts) corresponding to *all* initial states. For K explanatory variables and I initial states, μ has length $I(K+1)$. Ω is a $I(K+1) \times I(K+1)$ matrix of covariances. In an ordered logit model, the threshold parameters encompass the random intercept that represents unobserved heterogeneity.

To understand the role of random coefficients, it is useful to separate the coefficients into their deterministic and stochastic components. Since the coefficients are drawn from a multivariate normal distribution, the vector of coefficients corresponding to a given individual, n, can be written as:[4]

$$\vec{\beta_n} = \mu + \Pi \vec{z_n},$$

where $\vec{z_n}$ is a vector of independent draws from a standard normal distribution. The length of $\vec{z_n}$, and thus the size of Π, can be varied to induce the required correlation pattern between coefficients. In many applications, the standard assumption is to assume that the coefficients are random but independent. In that case, $\vec{z_n}$ is a $I(K+1)$ vector, Ω is a diagonal matrix, and Π is the Cholesky factor ("square root") of Ω.

Even in models with fixed coefficients, there is a dependence on past behavior, albeit limited. The benefit of a random coefficients model is that the stochastic component of a coefficient is specific to an individual but constant over time, so that there is correlation throughout the life of the loan. Under the assumption of *independent* random coefficients, however, since coefficients are specific to an initial state, the borrower's behavior is correlated only while he remains in the same state. Once the borrower transits to a new state it is as if he were a different individual. Despite this restriction, independent random coefficients are an improvement over a fixed parameters model since, for example, a distinction will be

[4]The coefficients for all initial states are drawn by a borrower at time $t = 0$.

drawn between a borrower that has been in 30-day delinquency for several months and one that has just entered that state.

3.2.2 Correlated Random Coefficients

Despite the clear benefits of a random coefficients model with independent coefficients, it is possible to induce correlation over the full sequence of a borrower's choices by allowing the coefficients to be *correlated*. The most general method for specifying correlation between coefficients is to let the covariance matrix, Ω, be non-diagonal. The drawback of this approach is that it drastically increases the number of parameters in the model.

An alternative is to treat $\vec{z_n}$ as a common shock that affects all the coefficients. In the simplest case, z_n is a random scalar and Π is a vector of length $I(K+1)$. The covariance matrix can be calculated by taking the outer square product of Π. The implication of this specification is that all coefficients are perfectly correlated. While this is clearly too restrictive, by allowing for just two random shocks, i.e. $\vec{z_n}$ is a 2×1 vector, it is possible to obtain pair-specific correlations. The model approaches the most general case as the number of independent random shocks increases. In the following section two correlated models will be compared against a model with independent random coefficients: one with two random shocks (Model I), and one with three random shocks (Model II).

3.2.3 Estimation

The probability of the choice stream *conditional* on the coefficients is given as in a fixed coefficients model.

$$Prob(\vec{y}_n|\beta) = P_{ni_1j_1}P_{nj_1j_2}...P_{nj_{T-1}j_T}$$

To derive the likelihood function the unconditional probability is calculated by integrating over the coefficients.

$$\begin{aligned} L(\beta|v) &= \prod_{n=1}^{N} Prob(\vec{y}_n) \\ &= \prod_{n=1}^{N} \int Prob(\vec{y}_n|\beta)\phi(\beta)d\beta \end{aligned}$$

Since this integral does not have a closed-form solution, maximum simulated likelihood is used for estimation.[5] For a given borrower repeated draws are made from the parameter distribution, and the probability of his choice stream is calculated for each set of draws. The average over all draws is the simulated probability that enters into the likelihood function.

$$SL(\beta|v) = \prod_{n=1}^{N} \frac{1}{R} \sum_{r=1}^{R} P_{ni_1j_1,r}P_{nj_1j_2,r}...P_{nj_{T-1}j_T,r}$$

3.3 Data

Since the parameter space in a random coefficients model is quite large, estimation can be computationally intensive. To illustrate the features of this model

[5]See Gourieroux and Monfort (1993) and Hajivassiliou and Ruud (1994).

a subset of the dataset discussed in the previous chapters is used. The full sample is made up of approximately 100,000 fixed-rate, residential, first-lien, non-conforming mortgages that serve as collateral for securities issued by Wells Fargo Mortgage Backed Securities Trust.[6] The loans were issued between 2000 and 2005. On average, the loans in the dataset follow the characteristics of jumbo loans. The average balance is $454,000, 95% of the loans have a loan to value (LTV) ratio below 80%, 65% have a FICO score above 720,[7] and 42% of the loans were contracted on property in California. Most of the mortgages have a term of 15 or 30 years. The subsample consists of 2,500 loans selected at random. There is no stratification in the sampling. Although this represents a small proportion of the total sample, the origination statistics, loan characteristics, and performance trends are comparable to those reported in the previous chapters for the full sample.

Table 3.1 reports origination trends for the full sample and the subsample. In both samples the number of loans issued in 2001 makes up the largest share of the sample, at 30%, while loans issued in 2004 represents the smallest share, at approximately 5%. All other years have loan issuance between 10% and 20%. Both samples also have similar frequencies for mortgage purpose and region of the property. 40% of the loans were contracted for a standard refinancing, 30% for property purchase, 20% for a cash-out refinancing, and 10% for an employer-sponsored relocation. The states with the highest representation are California

[6] The data was obtained from www.ctslink.com.

[7] FICO scores measure credit quality on a scale of 300 to 850.

(40%), the New York metropolitan area (12%), and the Great Lakes region (11%).

We would expect a representative subsample to have the same distribution for loan characteristics as the full sample. In Table 3.2 it is shown that the mean and standard deviation for key loan characteristics at origination — coupon rate, property appraisal, LTV ratio, and FICO score — are nearly the same for the subsample and the full sample. The average coupon rate reaches its lowest point in 2003, and then rises to 5.7% in 2005. Property appraisal rises steadily from 2000 to 2004, with a dip in 2005. LTV ratios are consistently below 80%. Average FICO scores are high, at above 720, in all years for both samples.

In addition to similar origination trends and loan characteristics, the two samples also display similar performance trends. Table 3.3 shows that in both samples 67% of the loans have prepaid before the end of the observation period, with over 90% of the loans issued prior to 2003 experiencing prepayment. 40% of the loans exhibit at least one month of curtailment, and 35% experience at least one month of 30-day delinquency. For both samples, the incidence of 60% day delinquency is low, at 4%. While the percent of the sample that experiences more severe delinquency is the same between the two samples, estimation of events below 60-day delinquency is difficult with the subsample since the number of observed events is very low. As a result, the continuum is condensed so that 90-day delinquency is the event furthest to the right.[8] Only observations with an initial state to the left of 90-day delinquency on the continuum are considered.

[8]The assumption underlying this simplification is that 90-day delinquency invariably leads to default, which is consistent with the definition of default employed by many previous studies.

3.4 Results

The estimation of a random coefficients model comprises both the mean and variance of the distribution for each coefficient, or more generally the mean and covariance matrix for the joint distribution of all coefficients in the model. A fixed coefficients model can be written as a special case of the random coefficients model, where the covariance matrix is assumed to be a zero matrix and the fixed coefficients estimates make up the mean vector of the "distribution". In the first subsection, a comparison is made between a fixed coefficients model and an independent random coefficients model. In the second subsection, the independent random coefficients model is compared to two models with correlated random coefficients.

3.4.1 Fixed vs. Independent Random Coefficients

Tables 3.4, 3.6, 3.8, and 3.10 report the parameter estimates for ordered logit models with the assumptions of fixed coefficients and independent random coefficients. The parameters of the model are estimated separately for each initial state since the coefficients are independent in both models. As mentioned in the previous section only observations with an initial state of curtailed, current, 30 days delinquent, or 60 days delinquent are considered.

A key concern for investors and originators of residential mortgages is how changes in economic conditions affect expected mortgage payments. The time-varying factors of interest are the age of the loan, the coupon gap, measured as the difference between the contract rate and the prevailing yield on the ten-year Treasury bond, the slope of the yield curve, calculated as the difference between the yields on the ten-year and two-year Treasury bonds, the contemporaneous LTV ratio, which measures the home equity the borrower has accumulated, and the borrower's FICO score. A seasonal dummy variable is included to control for differences in the housing market during the spring and summer months versus the fall and winter months. The LTV ratio at origination and the purpose for which the mortgage was contracted (refinancing vs. property purchase) are included to proxy for characteristics of the borrower at origination.

The tables show that for all initial states, the random coefficients model has a better fit than the fixed coefficients model. Since the random coefficients model has a larger parameter space than the fixed coefficients model, the adjusted log-likelihood ratio index of Ben-Akiva and Lerman (1985),[9] which includes a penalty for the number of parameters, is used to measure goodness of fit. This result is also confirmed by examination of the Akaike Information Criterion (AIC),[10] which is lower for the random coefficients model in all initial states.

Examination of Standard Deviation Estimates

The threshold parameters are a distinguishing feature of an ordered logit model. In order for each mortgage event to be well defined, the thresholds must

[9]The formula for the Ben-Akiva and Lerman adjusted log-likelihood ratio index is $1 - \frac{LL-K}{LL_0}$, where where K is the number of parameters, LL is the log-likelihood value of the full model, and LL_0 is the log-likelihood value for a model with only an intercept.

[10]The formula for the AIC is $2K - 2LL$, where K is the number of parameters and LL is the log-likelihood value.

maintain the ordering established in a fixed coefficients model. In the random coefficients model, the first intercept, representing the threshold for prepayment, is assumed to follow a normal distribution, like the coefficients on the covariates. The differences between each subsequent threshold are assumed to follow a squared normal distribution to ensure that the distances remain positive. Taken together, the thresholds make up the intercept of the utility function. The non-zero standard deviations estimated under the random coefficients model for the thresholds, which is observed in all initial states, indicates the presence of unobserved heterogeneity, and confirms the importance of modeling a borrower's behavior with a heterogeneity term.

Tables 3.4, 3.6, and 3.8 show that non-zero standard deviations are estimated for several of the coefficients, as well. For loans in a curtailed state, Table 3.4 shows that the coefficient on the slope of the yield curve is estimated to have a standard deviation that is significantly different from zero. This suggests that borrowers that have curtailed their mortgage exhibit heterogeneous responses to changes in interest rate expectations. For borrowers in a current state, on the other hand, Table 3.6 shows that the standard deviation on the coefficient of the slope of the yield curve is not statistically different from zero. Instead, heterogeneity in borrower behavior is captured through non-zero standard deviations for the coefficients on the cube of the coupon gap and the lowest FICO score bucket (620-300).

In Table 3.6 the estimated standard deviation for the lowest FICO score bucket has nearly the same magnitude as the estimated mean, indicating a wide distribution of behavior for *current* borrowers with poor credit. Under the fixed coefficients model, the coefficient on the cube of the coupon gap is a small positive number, while under the random coefficients model, the estimated mean for this coefficient is not statistically different from zero but the estimated standard deviation is. Thus, while on average the effect of the coupon gap may be neutral, for *current* loans there is a distribution of behavior, with some borrowers having a positive response and others a negative response. These results signal a potentially serious misspecification under the assumption of fixed coefficients.

To fully understand the impact of the assumption of random coefficients versus fixed coefficients, it is useful to calculate the estimated derivatives of each transition probability with respect to the covariates. These derivatives measure how much the probability of a given transition event is expected to change with a one unit change in a covariate.[11] The numerical derivative is calculated for each observation under both models. For the random coefficients model, the derivative is calculated with respect to the simulated probability. The values reported in Tables 3.5, 3.7, and 3.9 are averages over all observations.

The differences in the parameter estimates between the fixed and random coefficients models translate to these derivatives. For example, even though Table 3.6 shows that for *current* mortgages the estimated standard deviation of the level of the coupon gap is not significantly different from zero, Table 3.7 reports larger

[11]This value is similar to the elasticities that are commonly calculated in regression analysis. Since the focus in this analysis is on transition probabilities, the derivative is a sufficient measure.

derivatives (in absolute value) for the coupon gap under the random coefficients model, due to the distribution in the cube of coupon gap. Heterogeneity in behavior among *current* borrowers with respect to their FICO score leads to a lower probability of entering 30-day delinquency under the random coefficients model, offset by a higher probability of remaining current. It is interesting to note that even though a non-zero standard deviation is estimated for the coefficient on the slope of the yield curve in Table 3.4 for curtailed loans, in Table 3.5 there is not a significant difference between the derivatives with respect to this factor calculated under the fixed and random coefficients models.

The impact of a random coefficients model is most notable with respect to mortgages in 30-day delinquency. Table 3.8 shows that for these loans the coefficients on the lowest FICO score bucket and the mortgage purpose have an estimated standard deviation that is significantly different from zero. Table 3.9 shows that, under both the fixed and random coefficients models, borrowers in 30-day delinquency with good credit (FICO scores between 620 and 720) are more likely to prepay, curtail, or recover to current than delinquent borrowers with excellent credit (FICO scores above 720). Distinct differences between the two models, however, are observed with respect to borrowers with poor credit (FICO scores below 620). Under the fixed coefficients model, these borrowers are more likely to remain in 30-day delinquency or enter 60-day delinquency than borrowers with higher FICO scores. Under the random coefficients model, borrowers in 30-day delinquency with poor FICO scores are also more likely to *prepay* than other borrowers. The random coefficients model also predicts a smaller derivative for staying in 30-day delinquency and a higher derivative for entering 60-day delinquency. These results suggest that borrowers in distress are equally likely to resolve their liability through prepayment as through default, an important result that has not been previously reported in the literature on delinquent loans.

There are also differences between the fixed and random coefficients models with respect to mortgage purpose. Under the assumption of fixed coefficients, loans contracted for property purchase are more likely to prepay, curtail, or recover to current than loans contracted for a refinancing. Under the random coefficients model, even though the mean of the distribution for the coefficient on property purchase is not significantly different from zero, the estimated derivatives imply a lower probability of prepayment, curtailing, or recovering to current for these loans. This effect is also observed in Tables 3.5 and 3.7 with respect to current and curtailed mortgages.

Examination of Mean Estimates

While the focus up to this point has been on the factors with a significant non-degenerate distribution under the random coefficients models, the more flexible specification of the random coefficients model also leads to differences between the coefficient estimated under a fixed coefficients model and the mean of the distribution estimated under a random coefficients model. The differences in these parameters have important economic implications, even though the coefficients may not have a significant non-degenerate distribution. The most notable example is the effect of the coupon gap on loans in 30-day delinquency.

Normally, the probability of prepayment increases with the coupon gap. A higher coupon gap is associated with prevailing mortgage rates that are lower than the rate contracted at the time of origination, indicating a higher benefit from refinancing. However, Table 3.9 shows that for loans in 30-day delinquency, under the fixed coefficients model an increase in the coupon gap is associated with a lower probability of prepayment. This is puzzling because prepayment is just as common from 30-day delinquency as it is for current or curtailed loans. A possible explanation is that prepayment of delinquent loans is due to property sale rather than refinancing, i.e. the borrower prevents default by selling the property. This is not a satisfactory explanation, however, because 30-day delinquency is not a severe enough credit event to substantially restrict a borrower's refinancing opportunities.

The random coefficients model predicts an entirely different result, and suggests that estimates obtained under the assumption of fixed coefficients may be incorrect. The derivatives with respect to the coupon gap reported in Table 3.9 for the random coefficients model have the opposite sign of those calculated for the fixed coefficients model. These results conform to the expectation that an increase in the coupon gap will even facilitate the prepayment of loans in 30-day delinquency. This has been observed anecdotally over the last five years, as the availability of adjustable-rate and negative-amortization mortgages expanded the refinancing opportunities available to borrowers with credit constraints.

Differences between the fixed coefficient and the mean estimate of the coefficient distribution also play a role among current and curtailed loans. For curtailed mortgages, the results of the fixed coefficients model imply that an increase in the coupon gap only has a positive effect on the probability of prepayment. Under the random coefficients model, an increase in the coupon gap also leads to a higher probability of remaining in a curtailed state. In this way, under the assumption of random coefficients, curtailed loans behave similarly to current loans with respect to the coupon gap. Other differences are larger derivatives with respect to the slope of the yield curve for current loans, and a stronger spring and summer effect for all initial states under the random coefficients model.

Examination of Estimates from 60-Day Delinquency

The results for mortgages in 60-day delinquency illustrate a problem that is often encountered in the estimation of random coefficients models. In the subsample there are very few observations in 60-day delinquency. This does not hinder estimation of an ordered logit model with fixed coefficients, as it would a multinomial logit model.[12] Table 3.10 shows that under the fixed coefficients model an increase in the slope of the yield curve, property purchase as the mortgage purpose, and the spring and summer factor all contribute to an increase in the probability of prepayment, and a corresponding decrease in the probability of falling further into delinquency, i.e. default. Age and an increase in the contemporaneous LTV ratio lead to a higher probability of default.

[12]See the previous chapter for a more detailed discussion.

At first glance it appears that, in contrast, the random coefficients model suggests that only the threshold parameters are significant factors for loans in 60-day delinquency. Closer examination, however, reveals that both the parameter estimates and their standard errors are much larger than those estimated under the fixed coefficients model. As the number of simulations used for the maximum simulated likelihood estimator increases, the parameter estimates explode and, ultimately, the model fails. This is a problem that has been frequently observed in the estimation of random coefficient models, and appears to be exacerbated by the low number of observations in 60-day delinquency. A detailed examination as to the cause of this outcome, and strategies for potential solutions is an area of current research.[13]

3.4.2 Independent vs. Correlated Random Coefficients

The results in the previous subsection indicate that random coefficients are necessary for accurately modeling a borrower's behavior. A random coefficients model is able to capture unobserved heterogeneity and correlation in behavior over time that is not available in a fixed coefficients model. To allow the borrower's behavior to be correlated through the full life of the loan, including changes in the status of the mortgage, it is necessary for all of the coefficients in the model to be correlated.

The correlated random coefficients model, presented in the previous section,

[13]Ben-Akiva, Bolduc, and Walker (2001) discuss identification issues in mixed logit models, and address several special cases including random coefficients.

provides a general framework for estimating correlated coefficients. The correlation between coefficients is driven by the number of independent random shocks specified, i.e. the size of $\vec{z_n}$. The drawback is that in order to maintain a manageable number of parameters, i.e. a small number of independent random shocks, restrictions are implicitly imposed on the correlation structure of the coefficients. In particular, a lower number of independent shocks implies a higher degree of correlation between coefficients. Therefore, to successfully estimate a model with correlated coefficients the researcher must have a thorough understanding of the model dynamics.

To illustrate the differences between independent and correlated random coefficients, a streamlined specification is used in this subsection. Only loans that are current, curtailed, or 30 days delinquent are considered. 60-day delinquency is the event furthest to the right on the continuum. The analysis is focused on the effect of age, the LTV ratio at origination, and the level of the coupon gap.[14]

Table 3.11 reports the parameter estimates of the independent random coefficients model reestimated under the streamlined specification. There are 19 independent random shocks, one for each coefficient, so that Π is restricted to be a 19×19 diagonal matrix. Since the covariance matrix of the joint distribution is obtained by calculating the outer square product of Π, the diagonal of Π, which is reported in the table, contains the standard deviation for each coefficient. Table

[14]Several important factors are missing from this streamlined analysis, namely the slope of the yield curve, the contemporaneous LTV ratio, and the borrower's FICO scores. Not only will this misspecification bias the estimation results, but it will overemphasize the importance of unobserved heterogeneity.

3.12 shows the results of a correlated model with two independent random shocks (Model I), including all of the elements of Π. The results of a correlated model with three independent random shocks (Model II) are in Table 3.13.

The first question to ask when comparing these three models is: does the correlation between coefficients matter? Table 3.14 shows the covariance matrix estimated under correlated model II. For age and the LTV ratio at origination, both the variance and covariance terms are quite small, consistent with the results in the previous subsection. For the threshold parameters and the coupon gap, the variance and covariance terms are significant, indicating correlation between coefficients that should not be ignored. This result is confirmed by the adjusted likelihood ratio indices calculated for each model, which are slightly higher for the correlated models.

Since Π takes on a different form in each model, direct comparison of the parameter estimates is difficult. To understand the effect of modeling borrower behavior with correlated random coefficients, Table 3.15 reports the average estimated derivatives for the independent random coefficients model and the two correlated random coefficients models. In general, the derivatives calculated under correlated model I and correlated model II are very similar. The exception is the derivatives with respect to the coupon gap calculated for loans in 30-day delinquency. For loans in 30-day delinquency, correlated model II predicts that a change in the coupon gap has a larger effect (in absolute value) on the probability of prepayment and entering 60-day delinquency than correlated model I. Under correlated model II, an increase in the coupon gap is also associated with a lower probability of recovering to current, while in the independent model and correlated model I the probability of recovering to current is not affected by changes in the coupon gap.

There are clear differences between the independent random coefficients model and the correlated random coefficients models, which suggests that it is important to take into account correlation between a borrower's responses in different mortgage states. For example, for loans in a curtailed state, under the correlated models an increase in the coupon gap is associated with a lower probability of remaining in a curtailed state, while under the independent model an increase in the coupon gap implies a higher probability of remaining in a curtailed state. Also with respect to the coupon gap, for loans in a current state the correlated models predict a larger positive derivative on the probability of prepayment, offset by a larger negative derivative on the probability of remaining in a current state. A change in the coupon gap is also associated with a neutral effect on the probability of falling into 30-day delinquency under the correlated models, while the independent model predicts a strong negative effect. For loans in 30-day delinquency, the derivatives calculated under the correlated models are, in general, smaller in magnitude than those calculated under the independent model.

3.5 Conclusion

In this chapter, it has been shown that random coefficients are essential for accurately modeling the payment behavior of residential mortgage borrowers. The decision made by the borrower with respect to his mortgage is correlated in each payment period due to unobserved factors and consistency in preferences. Random coefficients are used to capture the effects of unobserved heterogeneity and correlation over time by specifying a factor of the error component that is specific to a borrower but constant over time.

The results in this chapter show that predicted borrower behavior varies substantially with the model specification. In the comparison of a fixed coefficients ordered logit model and an independent random coefficients ordered logit model, it is found that the random coefficients model has a better fit, and predicts behavior that is in line with economic intuition. This is particularly true for delinquent loans, and it is an important result because these loans have not previously been studied in depth.

To fully capture the correlation in a borrower's behavior over time it is necessary for the coefficients corresponding to different initial states to be correlated. This chapter presents a generalized framework for modeling correlated coefficients while maintaining a manageable number of parameters. For the covariates most affected by random coefficients, such as the threshold parameters and the coupon gap, there is a significant difference in the response predicted under an independent random coefficients model and a correlated random coefficients model. The model estimates also indicate that there is correlation between coefficients that should be taken into account.

Random coefficients have not previously been used with respect to mortgage borrower behavior. However, this is a natural application because a borrower's decision is highly influenced by behavioral factors that cannot be captured in a model driven solely by financial dynamics. The results in this chapter suggest that, since predicted behavior is highly dependent on the model specification, models that do not consider random coefficients will lead to mispricing of mortgage instruments. A random coefficients model provides a further benefit with respect to pricing. Since a *distribution* of behavior is estimated, simulation through this distribution can be used to calculate empirical distributions for probabilities and prices. This enhances the information available to investors and originators, and allows them to make statistical inference with respect to mortgage investments.

Table 3.1: Comparison of Origination Trends

	Subsample	Full Sample
No. of Loans	2,500	103,347
	Percent of Sample	
Year of Origination		
2000	13.9	13.0
2001	29.8	30.9
2002	17.9	17.7
2003	19.6	20.5
2004	5.6	5.1
2005	13.3	12.8
Mortgage Term		
15-Year	27.2	27.4
30-Year	71.4	71.3
Other	1.4	1.3
Mortgage Purpose		
Standard Refinancing	40.8	42.2
Cash-Out Refinancing	19.9	20.7
Property Purchase	30.2	28.2
Employer Relocation	9.2	8.9
Region		
CA-HI	42.3	41.2
NY-NJ-CT	12.2	12.5
Midatlantic	8.4	9.2
Texas	3.5	4.1
Florida	3.1	2.8
South	5.2	5.1
Lakes	10.8	11.1
Mountain	5.7	5.9
Northwest	2.6	2.7
Plains/Midwest	2.6	2.2
New England	3.4	3.4

Since estimation of a random coefficients model requires the use of a simulated likelihood function, a subsample of the data is used in this chapter to simplify computation. Although only a small proportion of the total sample is used, the subsample is representative of the origination trends present in the full sample.

Table 3.2: Comparison of Loan Characteristics by Issue Year

	Subsample		Full Sample	
No. of Loans		2,500		103,347
	Mean	*Std. Dev.*	*Mean*	*Std. Dev.*
	Coupon Rate at Origination			
2000	8.11	0.41	8.08	0.45
2001	7.21	0.47	7.17	0.44
2002	6.57	0.41	6.51	0.42
2003	5.48	0.43	5.43	0.45
2004	5.44	0.38	5.46	0.39
2005	5.78	0.27	5.71	0.32
All	6.60	1.01	6.47	1.03
	Property Appraisal at Origination			
2000	525,000	240,000	521,000	268,000
2001	650,000	326,000	653,000	360,000
2002	814,000	607,000	766,000	486,000
2003	955,000	604,000	936,000	592,000
2004	968,000	881,000	938,000	652,000
2005	818,000	481,000	845,000	531,000
All	762,000	522,000	769,000	505,000
	Loan To Value Ratio at Origination			
2000	75	13	75	13
2001	71	14	70	14
2002	65	15	65	16
2003	59	16	59	16
2004	60	17	61	16
2005	68	13	67	15
All	67	15	66	16
	FICO Score at Origination			
2000	725	47	724	49
2001	728	48	726	49
2002	736	48	732	50
2003	736	45	737	44
2004	735	49	736	43
2005	741	45	743	43
All	733	47	732	48

This table reports the mean and standard deviation of key loan characteristics by issue year. For the coupon rate, the LTV ratio, and the FICO score both the mean and standard deviation are similar across the subsample and the full sample. For the property appraisal at origination, the standard deviation of the subsample is slightly larger, but in general the subsample displays the same distribution of loan characteristics as the full sample.

Table 3.3: Comparison of Loan Performance by Issue Year

No. of Loans		Subsample (SS): 2,500			Full Sample (FS): 103,347				
		Percent of Sample Entering Event at Least One Time							*No. Obs.*
		2000	2001	2002	2003	2004	2005	All	
Prepayment	SS	99.4	99.3	92.8	29.9	12.9	1.8	66.8	1,670
	FS	99.3	99.2	91.5	29.6	14.1	1.7	66.8	66,408
Curtailment	SS	37.1	44	39.2	52.6	48.6	25.9	41.7	1,042
	FS	37	41.6	43.6	51.8	47.1	25.4	41.7	41,248
30-Day Delinquency	SS	51.2	39.1	28.3	26.6	30	34.9	35.3	883
	FS	48.5	39.1	28.4	27.9	23.7	33.6	34.6	34,354
60-Day Delinquency	SS	6	3.5	2	1.4	0.7	15.7	4.6	116
	FS	5.5	2.9	1.5	1.1	0.8	16.4	4.2	4,123
90-Day Delinquency	SS	1.2	0.9	0.7	0.2	-	8.4	1.7	43
	FS	1.3	0.8	0.4	0.3	0.1	8.5	1.7	1,586
120-Day Delinquency	SS	0.6	0.7	0.5	-	-	1.2	0.5	13
	FS	0.8	0.5	0.3	0.2	0.1	0.8	0.4	426
Bankruptcy	SS	0.3	0.1	0.2	-	-	-	0.1	3
	FS	0.2	0.2	0.1	0.1	0.1	0.03	0.1	136
Foreclosure	SS	0.6	0.7	0.5	-	-	-	0.4	9
	FS	0.5	0.4	0.2	0.1	-	0.02	0.3	248
REO	SS	-	-	-	-	-	-	-	0
	FS	0.1	0.1	0.03	0.01	-	-	0.1	51
Default*	SS	0.6	0.5	0.5	-	-	-	0.3	8
	FS	0.7	0.4	0.2	0.03	0.02	-	0.3	274

* Default is defined as a loan that is liquidated or prepays when the status is 90-day delinquency or below.

This table shows the percentage of borrowers in each sample that enter the given event at least once. It is striking to note that the similarities between the two samples also hold in terms of the trends with respect to origination year. For example, in both samples over 90% of the loans issued prior to 2003 have prepaid.

Table 3.4: Parameter Estimates Corresponding to a Curtailed State

	Fixed		**Random**			
Observations	11,227		11,227			
Log-likelihood	-4266		-3777			
Adj. LL Ratio Index	0.05		0.16			
AIC	8558		7606			
	Coefficient		Mean		Std. Dev.	
$Intercept1$	4.81	***	6.66	***	0.14	
	(0.173)		(0.337)		(0.145)	
$\sqrt{Int1 - Int2}$	2.55	***	2.96	***	0.71	***
	(0.012)		(0.045)		(0.042)	
$\sqrt{Int2 - Int3}$	0.74	***	0.82	***	1.30	***
	(0.019)		(0.079)		(0.097)	
Age	0.034	***	0.042	***	0.001	
(months)	(0.005)		(0.007)		(0.007)	
LTV at	0.006		0.005		0.001	
Origination	(0.004)		(0.007)		(0.002)	
Coupon Gap	0.566	***	1.420	***	0.019	
(1-month lag)	(0.072)		(0.154)		(0.067)	
Coupon Gap	0.003		-0.024	***	0.007	
Cubed (1-month lag)	(0.003)		(0.008)		(0.005)	
Slope of Yield	0.392	***	0.412	***	0.132	**
Curve (1-month lag)	(0.049)		(0.071)		(0.068)	
Current LTV	-0.019	***	-0.017	***	0.001	
	(0.004)		(0.007)		(0.002)	
Baseline FICO: 850 - 720						
FICO: 720 - 620	-0.25	***	-0.13		0.02	
	(0.054)		(0.108)		(0.273)	
FICO: 620 - 300	-1.41	***	-1.36	***	0.88	
	(0.104)		(0.455)		(0.717)	
Baseline: Oct - Mar						
Apr - Sep	0.15	**	0.23	***	0.14	
	(0.069)		(0.073)		(0.161)	
Baseline: Refinancing						
Purchase /	-0.05		0.06		0.17	
Relocation	(0.060)		(0.107)		(0.195)	

Standard errors are reported in parenthesis:
*** 1% Significance; ** 5% Significance; * 10% Significance
The adjusted log-likelihood ratio index follows Ben-Akiva and Lerman (1985).
AIC is the Akaike Information Criterion.

The distribution for the intercepts, or threshold parameters, estimated under the random coefficients model indicates that unobserved heterogeneity is a factor for loans in a curtailed state. This table shows that there is also heterogeneity in the response of borrowers to changes in the slope of the yield curve.

Table 3.5: Average Simulated Derivatives of Transition Probabilities From a Curtailed State

	Fixed Coefficients				**Random Coefficients**			
Transition to:	Prepay	Curtail	Current	30-Day	Prepay	Curtail	Current	30-Day
Age (months)	0.16	-0.03	-0.05	-0.08	0.20	-0.01	-0.08	-0.11
Org. LTV†	0.03	-0.004	-0.01	-0.01	0.02	-0.001	-0.01	-0.01
Coupon Gap	2.95	-0.61	-0.91	-1.42	4.22	1.13	-2.24	-3.11
Yield Curve Slope	1.84	-0.30	-0.60	-0.94	2.09	-0.23	-0.80	-1.06
Current LTV	-0.09	0.01	0.03	0.04	-0.08	0.003	0.03	0.04
Baseline FICO: 850 - 720								
FICO: 720 - 620	-1.14	0.11	0.40	0.63	-0.63†	0.01†	0.27†	0.36†
FICO: 620 - 300	-3.90	-6.25	3.63	6.52	-3.42	-4.54	3.04	4.92
Baseline: Oct - Mar								
Apr - Sep	0.72	-0.12	-0.24	-0.37	1.13	-0.10	-0.44	-0.59
Baseline: Refinancing								
Purchase / Relocation†	-0.24	0.04	0.08	0.12	0.32	-0.10	-0.10	-0.13

† Coefficient is not statistically significant.

The derivatives in this table are averages over all observations with a curtailed state. Even though the previous table indicates that there is heterogeneity in borrowers' behavior with respect to the slope of the yield curve, the derivatives corresponding to that factor are not significantly different between the fixed coefficients model and the random coefficients model. There is a significant difference, however, in the response to the coupon gap. This is because the mean of the distribution for the coefficient on the coupon gap is not similar to the fixed coefficient.

Table 3.6: Parameter Estimates Corresponding to a Current State

	Fixed		**Random**			
Observations	17,397					
Log-likelihood	-8747		-8038			
Adj. LL Ratio Index	0.02		0.09			
AIC	17520		16128			
	Coefficient		Mean		Std. Dev.	
$Intercept1$	3.49	***	4.91	***	0.12	
	(0.125)		(0.250)		(0.092)	
$\sqrt{Int1 - Int2}$	0.72	***	0.52	***	0.64	***
	(0.016)		(0.048)		(0.039)	
$\sqrt{Int2 - Int3}$	2.36	***	2.71	***	0.57	***
	(0.007)		(0.027)		(0.028)	
Age	-0.015	***	0.002		3.0E-04	
(months)	(0.003)		(0.007)		(0.007)	
LTV at	0.019	***	0.017	**	0.002	
Origination	(0.004)		(0.008)		(0.001)	
Coupon Gap	0.251	***	0.739	***	0.042	
(1-month lag)	(0.051)		(0.114)		(0.044)	
Coupon Gap	0.010	***	0.005		0.016	***
Cubed (1-month lag)	(0.002)		(0.006)		(0.003)	
Slope of Yield	0.188	***	0.326	***	0.045	
Curve (1-month lag)	(0.034)		(0.051)		(0.053)	
Current LTV	-0.028	***	-0.028	***	0.001	
	(0.004)		(0.008)		(0.002)	
Baseline FICO: 850 - 720						
FICO: 720 - 620	-0.32	***	-0.22	***	0.06	
	(0.037)		(0.078)		(0.157)	
FICO: 620 - 300	-1.70	***	-1.40	***	1.07	**
	(0.097)		(0.398)		(0.469)	
Baseline: Oct - Mar						
Apr - Sep	0.18	***	0.31	***	0.02	
	(0.051)		(0.054)		(0.140)	
Baseline: Refinancing						
Purchase /	0.03		-0.04		0.13	
Relocation	(0.042)		(0.086)		(0.165)	

Standard errors are reported in parenthesis:
*** 1% Significance; ** 5% Significance; * 10% Significance
The adjusted log-likelihood ratio index follows Ben-Akiva and Lerman (1985).
AIC is the Akaike Information Criterion.

For loans that are current, significant heterogeneity in behavior is observed with respect to the coupon gap and the lowest FICO score bucket. It is important to model the heterogeneity among borrowers with poor credit because the magnitude of the standard deviation estimate is close to that of the mean estimate. This suggests a serious misspecification under the fixed coefficients model.

Table 3.7: Average Simulated Derivatives of Transition Probabilities From a Current State

	Fixed Coefficients				**Random Coefficients**			
Transition to:	Prepay	Curtail	Current	30-Day	Prepay	Curtail	Current	30-Day
Age (months)	-0.06	-0.04	0.03	0.07	0.01†	0.005†	-0.005†	-0.01†
Org. LTV	0.08	0.05	-0.04	-0.09	0.08	0.06	-0.06	-0.08
Coupon Gap	1.84	1.04	-1.06	-1.82	4.26	2.73	-3.40	-3.59
Yield Curve Slope	0.81	0.47	-0.36	-0.92	1.47	1.09	-1.04	-1.52
Current LTV	-0.12	-0.07	0.05	0.14	-0.13	-0.09	0.09	0.13
Baseline FICO: 850 - 720								
FICO: 720-620	-1.31	-0.75	0.46	1.60	-0.95	-0.72	0.62	1.05
FICO: 620-300	-3.76	-2.31	-10.87	16.94	-2.88	-3.07	-3.57	9.52
Baseline: Oct - Mar								
Apr-Sep	0.78	0.45	-0.35	-0.88	1.37	1.02	-0.95	-1.44
Baseline: Refinancing								
Purchase / Relocation†	0.11	0.06	-0.05	-0.12	-0.16	-0.13	0.07	0.22

† Coefficient is not statistically significant.

For loans in a current state, the impact of a random coefficients model is felt through stronger predicted effects for the coupon gap and weaker effects for loans with poor credit relative to borrowers with higher FICO scores.

Table 3.8: Parameter Estimates Corresponding to 30-Day Delinquency

	Fixed		**Random**			
Observations	3,540					
Log-likelihood	-4107		-3433			
Adj. LL Ratio Index	0.03		0.18			
AIC	8242		6922			

	Coefficient		Mean		Std. Dev.	
$Intercept1$	1.99	***	4.48	***	0.11	
	(0.194)		(0.495)		(0.127)	
$\sqrt{Int1 - Int2}$	1.01	***	0.75	***	0.80	***
	(0.023)		(0.057)		(0.058)	
$\sqrt{Int2 - Int3}$	1.16	***	1.16	***	1.15	***
	(0.012)		(0.046)		(0.065)	
$\sqrt{Int3 - Int4}$	1.95	***	2.47	***	0.88	***
	(0.018)		(0.088)		(0.101)	
Age	0.012	***	0.020	***	0.002	
(months)	(0.003)		(0.008)		(0.009)	
LTV at	0.016	***	0.026	**	0.003	
Origination	(0.004)		(0.012)		(0.002)	
Coupon Gap	-0.399	***	0.477	**	0.044	
(1-month lag)	(0.072)		(0.201)		(0.048)	
Coupon Gap	0.005	*	-0.001		0.003	
Cubed (1-month lag)	(0.003)		(0.008)		(0.003)	
Slope of Yield	0.440	***	0.407	***	0.072	
Curve (1-month lag)	(0.040)		(0.074)		(0.071)	
Current LTV	-0.032	***	-0.038	***	4.0E-04	
	(0.004)		(0.011)		(0.002)	
Baseline FICO: 850 - 720						
FICO: 720 - 620	0.12	***	0.14		0.06	
	(0.044)		(0.137)		(0.143)	
FICO: 620 - 300	-0.36	***	-0.64		2.01	***
	(0.075)		(0.452)		(0.480)	
Baseline: Oct - Mar						
Apr - Sep	0.08		0.40	***	0.23	
	(0.072)		(0.096)		(0.161)	
Baseline: Refinancing						
Purchase /	0.06		-0.12		0.36	**
Relocation	(0.047)		(0.145)		(0.176)	

Standard errors are reported in parenthesis:
*** 1% Significance; ** 5% Significance; * 10% Significance
The adjusted log-likelihood ratio index follows Ben-Akiva and Lerman (1985).
AIC is the Akaike Information Criterion.

For loans in 30-day delinquency, the coefficients on the lowest FICO score bucket and the mortgage purpose have an estimated standard deviation that is significantly different from zero. Since these are both dummy variables, the coefficients will ultimately affect the intercept of the utility function.

Table 3.9: Average Simulated Derivatives From 30-Day Delinquency

Transition to:	Prepay	Curtail	Current	30-Day	60-Day
		Fixed Coefficients			
Age (months)	0.07	0.08	0.11	-0.22	-0.04
Org. LTV	0.09	0.11	0.15	-0.30	-0.06
Coupon Gap	-1.80	-2.03	-2.74	5.57	1.00
Yield Curve Slope	2.59	3.00	4.31	-8.28	-1.61
Current LTV	-0.19	-0.22	-0.31	0.60	0.12
Baseline FICO: 850 - 720					
FICO: 720-620	0.73	0.85	1.21	-2.34	-0.45
FICO: 620-300	-1.87	-2.29	-3.75	6.38	1.53
Baseline: Oct - Mar					
Apr-Sep†	0.44	0.51	0.74	-1.42	-0.28
Baseline: Refinancing					
Purchase / Relocation†	0.33	0.38	0.54	-1.05	-0.21
		Random Coefficients			
Age (months)	0.13	0.10	0.01	-0.17	-0.06
Org. LTV	0.17	0.13	0.01	-0.22	-0.08
Coupon Gap	3.08	2.30	0.13	-3.99	-1.52
Yield Curve Slope	2.71	2.03	0.11	-3.53	-1.32
Current LTV	-0.25	-0.19	-0.01	0.33	0.12
Baseline FICO: 850 - 720					
FICO: 720-620†	0.90	0.67	0.03	-1.16	-0.44
FICO: 620-300	3.53	-3.72	-5.20	1.12	4.27
Baseline: Oct - Mar					
Apr-Sep	2.75	1.98	0.05	-3.51	-1.27
Baseline: Refinancing					
Purchase / Relocation	-0.47	-0.63	-0.25	0.92	0.43

† Coefficient is not statistically significant.

Under the random coefficients model, borrowers in 30-day delinquency with poor FICO scores are more likely to *prepay* than other borrowers. This suggests that borrowers in distress are equally likely to resolve their liability through prepayment as through default. Normally, the probability of prepayment increases with the coupon gap. However, under the fixed coefficients model an increase in the coupon gap is associated with a lower probability of prepayment. The random coefficients model predicts an entirely different result, and suggests that the fixed coefficient estimates may be incorrect.

Table 3.10: Parameter Estimates Corresponding to 60-Day Delinquency

	Fixed		**Random**			
Observations	133					
Log-likelihood	-175		-150			
Adj. LL Ratio Index	0.06		0.11			
AIC	380		360			
	Coefficient		Mean		Std. Dev.	
$Intercept1$	6.11	***	52.24		0.46	
	(1.540)		(40.175)		(2.491)	
$\sqrt{Int1 - Int2}$	1.06	***	1.90	**	1.54	**
	(0.187)		(0.831)		(0.846)	
$\sqrt{Int2 - Int3}$	0.68	***	1.64	***	0.99	
	(0.140)		(0.634)		(1.586)	
$\sqrt{Int3 - Int4}$	1.41	***	3.75	***	1.26	
	(0.096)		(1.306)		(0.783)	
$\sqrt{Int4 - Int5}$	1.01	***	2.12	***	0.32	
	(0.098)		(0.789)		(0.810)	
Age	-0.071	**	-0.672		0.216	
(months)	(0.037)		(0.718)		(0.254)	
LTV at	0.078	**	0.812		0.051	
Origination	(0.040)		(0.981)		(0.079)	
Coupon Gap	0.636		14.616		1.109	
(1-month lag)	(0.517)		(11.536)		(2.129)	
Coupon Gap	-0.010		-0.163		0.228	
Cubed (1-month lag)	(0.018)		(0.315)		(0.168)	
Slope of Yield	1.174	***	8.139		5.729	
Curve (1-month lag)	(0.307)		(6.810)		(3.481)	
Current LTV	-0.092	**	-0.978		0.041	
	(0.040)		(1.017)		(0.053)	

Standard errors are reported in parenthesis:

*** 1% Significance; ** 5% Significance; * 10% Significance

The adjusted log-likelihood ratio index follows Ben-Akiva and Lerman (1985).

AIC is the Akaike Information Criterion.

continued on next page

Table 3.10: Parameter Estimates Corresponding to 60-Day Delinquency *(continued)*

	Coefficient		Mean	Std. Dev.
Baseline FICO: 850 - 720				
FICO: 720 - 620	-0.29		-7.74	13.96
	(0.443)		(6.163)	(9.229)
FICO: 620 - 300	0.30		-0.57	18.37
	(0.605)		(9.30)	(21.40)
Baseline: Oct - Mar				
Apr - Sep	0.73	**	10.17	5.64
	(0.358)		(8.471)	(5.284)
Baseline: Refinancing				
Purchase /	0.83	*	-1.19	1.94
Relocation	(0.437)		(6.457)	(4.741)

Standard errors are reported in parenthesis:
*** 1% Significance; ** 5% Significance; * 10% Significance

At first glance it appears that the random coefficients model suggests that only the threshold parameters are significant for loans in 60-day delinquency. Closer examination reveals that both the parameter estimates and their standard errors are much larger than those estimated under the fixed coefficients model. As the number of simulations used for the maximum simulated likelihood estimator increases, the parameter estimates explode and, the model fails.

Table 3.11: Parameter Estimates - Independent Random Coefficients Model

Observations	32,273			
Log-likelihood	-15893			
Adj. LL Ratio Index	0.10			
AIC	31862			
	Mean (μ)		**Diagonal of Π**	
Parameters Corresponding to a Curtailed State				
$Intercept1$	4.790	***	0.006	
	(0.191)		(0.044)	
$\sqrt{Int1 - Int2}$	2.652	***	0.498	***
	(0.021)		(0.018)	
$\sqrt{Int2 - Int3}$	0.746	***	0.657	***
	(0.028)		(0.031)	
Age	0.022	***	-0.005	*
(months)	(0.005)		(0.003)	
LTV at	-0.007	***	0.001	
Origination	(0.003)		(0.001)	
Coupon Gap	0.871	***	0.094	***
(1-month lag)	(0.041)		(0.018)	
Parameters Corresponding to a Current State				
$Intercept1$	3.936	***	-0.061	
	(0.134)		(0.040)	
$\sqrt{Int1 - Int2}$	0.635	***	0.307	***
	(0.019)		(0.016)	
$\sqrt{Int2 - Int3}$	2.479	***	-0.456	***
	(0.015)		(0.014)	
Age	0.006	*	-0.008	***
(months)	(0.004)		(0.002)	
LTV at	-0.008	***	0.000	
Origination	(0.002)		(0.001)	
Coupon Gap	0.619	***	-0.019	
(1-month lag)	(0.034)		(0.016)	

Standard errors are reported in parenthesis:
*** 1% Significance; ** 5% Significance; * 10% Significance
The adjusted log-likelihood ratio index follows Ben-Akiva and Lerman (1985). AIC is the Akaike Information Criterion.

continued on next page

Table 3.11: Parameter Estimates - Independent Random Coefficients Model *(continued)*

	Mean (μ)		**Diagonal of Π**	
Parameters Corresponding to 30-Day Delinquency				
$Intercept1$	2.575	***	-0.139	***
	(0.225)		(0.046)	
$\sqrt{Int1 - Int2}$	0.931	***	0.536	***
	(0.028)		(0.023)	
$\sqrt{Int2 - Int3}$	1.293	***	0.614	***
	(0.020)		(0.019)	
$\sqrt{Int3 - Int4}$	2.509	***	0.889	***
	(0.054)		(0.079)	
Age	0.041	***	-0.031	***
(months)	(0.006)		(0.003)	
LTV at	-0.013	***	0.003	***
Origination	(0.002)		(0.001)	
Coupon Gap	0.103	**	0.095	***
(1-month lag)	(0.043)		(0.017)	

Standard errors are reported in parenthesis:

*** 1% Significance; ** 5% Significance; * 10% Significance

Table 3.12: Parameter Estimates - Correlated Random Coefficients Model I

Observations	32,273					
Log-likelihood	-15455					
Adj. LL Ratio Index	0.12					
AIC	31024					
	Mean (μ)		Π			
Parameters Corresponding to a Curtailed State						
$Intercept1$	5.190	***	0.288		-1.152	***
	(0.262)		(0.299)		(0.295)	
$\sqrt{Int1 - Int2}$	2.919	***	0.137	***	-0.629	***
	(0.032)		(0.030)		(0.031)	
$\sqrt{Int2 - Int3}$	1.886	***	0.979	***	-0.821	***
	(0.123)		(0.045)		(0.087)	
Age	0.020	***	0.000		-0.003	
(months)	(0.006)		(0.007)		(0.007)	
LTV at	-0.006	*	0.001		0.000	
Origination	(0.003)		(0.004)		(0.004)	
Coupon Gap	1.008	***	0.165	**	-0.347	***
(1-month lag)	(0.064)		(0.066)		(0.068)	
Parameters Corresponding to a Current State						
$Intercept1$	4.108	***	-0.551	***	-1.550	***
	(0.152)		(0.200)		(0.201)	
$\sqrt{Int1 - Int2}$	0.634	***	0.011		0.393	***
	(0.023)		(0.029)		(0.040)	
$\sqrt{Int2 - Int3}$	2.566	***	-0.375	***	-0.429	***
	(0.018)		(0.018)		(0.020)	
Age	0.019	***	-0.015	***	-0.007	
(months)	(0.004)		(0.005)		(0.006)	
LTV at	-0.011	***	0.008	***	-0.004	*
Origination	(0.002)		(0.003)		(0.002)	
Coupon Gap	0.687	***	-0.436	***	-0.504	***
(1-month lag)	(0.042)		(0.049)		(0.048)	

Standard errors are reported in parenthesis:

*** 1% Significance; ** 5% Significance; * 10% Significance

The adjusted log-likelihood ratio index follows Ben-Akiva and Lerman (1985).

AIC is the Akaike Information Criterion.

continued on next page

Table 3.12: Parameter Estimates - Correlated Random Coefficients Model I *(continued)*

	Mean (μ)		Π			
Parameters Corresponding to 30-Day Delinquency						
$Intercept1$	2.286	***	-0.522	*	0.221	
	(0.240)		(0.295)		(0.273)	
$\sqrt{Int1 - Int2}$	0.474	***	-0.503	***	0.657	***
	(0.054)		(0.031)		(0.037)	
$\sqrt{Int2 - Int3}$	1.039	***	0.751	***	-0.091	***
	(0.031)		(0.024)		(0.033)	
$\sqrt{Int3 - Int4}$	2.207	***	0.135	**	-0.483	***
	(0.042)		(0.056)		(0.052)	
Age	0.023	***	-0.003		0.030	***
(months)	(0.006)		(0.007)		(0.007)	
LTV at	-0.012	***	-0.006	*	-0.006	**
Origination	(0.003)		(0.004)		(0.003)	
Coupon Gap	0.038		-0.078		0.111	**
(1-month lag)	(0.049)		(0.060)		(0.056)	

Standard errors are reported in parenthesis:

*** 1% Significance; ** 5% Significance; * 10% Significance

Table 3.13: Parameter Estimates - Correlated Random Coefficients Model II

Observations	32,273							
Log-likelihood	-15314							
Adj. LL Ratio Index	0.13							
AIC	30780							
	Mean (μ)		Π					
Parameters Corresponding to a Curtailed State								
$Intercept1$	5.181	***	0.631	*	-0.385		-0.838	***
	(0.259)		(0.312)		(0.409)		(0.314)	
$\sqrt{Int1 - Int2}$	2.933	***	0.301	***	-0.301	***	-0.645	***
	(0.038)		(0.033)		(0.037)		(0.034)	
$\sqrt{Int2 - Int3}$	2.191	***	1.327	***	-0.785	***	-1.036	***
	(0.132)		(0.062)		(0.055)		(0.075)	
Age	0.016	***	-0.011		-0.012		-0.009	
(months)	(0.006)		(0.007)		(0.009)		(0.008)	
LTV at	-0.008	**	0.009	**	-0.001		0.010	**
Origination	(0.004)		(0.004)		(0.004)		(0.004)	
Coupon Gap	1.079	***	0.103		0.017		-0.450	***
(1-month lag)	(0.064)		(0.069)		(0.071)		(0.078)	
Parameters Corresponding to a Current State								
$Intercept1$	4.143	***	-0.651	***	0.578	***	-1.667	***
	(0.159)		(0.188)		(0.225)		(0.212)	
$\sqrt{Int1 - Int2}$	0.650	***	-0.078	***	-0.078	**	0.314	***
	(0.021)		(0.028)		(0.035)		(0.034)	
$\sqrt{Int2 - Int3}$	2.611	***	-0.309	***	-0.114	***	-0.457	***
	(0.022)		(0.018)		(0.022)		(0.020)	
Age	0.016	***	-0.004		-0.004		-0.003	
(months)	(0.005)		(0.005)		(0.006)		(0.005)	
LTV at	-0.009	***	-0.001		0.009	***	-0.004	**
Origination	(0.002)		(0.002)		(0.003)		(0.002)	
Coupon Gap	0.662	***	-0.235	***	0.014		-0.571	***
(1-month lag)	(0.041)		(0.043)		(0.052)		(0.051)	

Standard errors are reported in parenthesis:

*** 1% Significance; ** 5% Significance; * 10% Significance

The adjusted log-likelihood ratio index follows Ben-Akiva and Lerman (1985).

AIC is the Akaike Information Criterion.

continued on next page

Table 3.13: Parameter Estimates - Correlated Random Coefficients Model II *(continued)*

	Mean (μ)		Π					
Parameters Corresponding to 30-Day Delinquency								
$Intercept1$	2.579	***	-0.723	**	1.045	***	-0.309	
	(0.278)		(0.313)		(0.352)		(0.294)	
$\sqrt{Int1 - Int2}$	0.814	***	-0.630	***	0.310	***	0.227	***
	(0.049)		(0.031)		(0.035)		(0.034)	
$\sqrt{Int2 - Int3}$	0.943	***	0.615	***	0.362	***	-0.044	
	(0.037)		(0.029)		(0.030)		(0.028)	
$\sqrt{Int3 - Int4}$	2.090	***	0.516	***	-0.311	***	-0.239	***
	(0.037)		(0.057)		(0.049)		(0.040)	
Age	0.036	***	-0.016	**	0.012	*	-0.006	
(months)	(0.007)		(0.007)		(0.007)		(0.007)	
LTV at	-0.011	***	-0.004		0.022	***	-0.014	***
Origination	(0.004)		(0.004)		(0.005)		(0.004)	
Coupon Gap	0.045		-0.172	***	-0.362	***	0.407	***
(1-month lag)	(0.060)		(0.062)		(0.072)		(0.062)	

Standard errors are reported in parenthesis:

*** 1% Significance; ** 5% Significance; * 10% Significance

Table 3.14: Covariance Matrix for the Distribution of Coefficients - Correlated Random Coefficients Model II

1.25	0.85	2.01	0.01	0.00	0.44	0.76	-0.28	0.23	0.00	0.00	-0.33	0.60	-0.71	0.29	0.65	-0.01	0.00	-0.31
0.85	0.60	1.30	0.01	0.00	0.32	0.70	-0.20	0.24	0.00	0.00	0.29	-0.33	-0.43	0.11	0.40	0.00	0.00	-0.21
2.01	1.30	3.45	0.00	0.00	0.59	0.41	-0.37	0.15	0.00	-0.01	0.27	-1.46	-1.32	0.58	1.18	-0.02	-0.01	-0.37
0.01	0.01	0.00	0.00	0.00	0.00	0.02	0.00	0.01	0.00	0.00	0.01	0.00	0.00	-0.01	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	-0.02	0.00	-0.01	0.00	0.00	-0.01	-0.01	0.00	0.01	0.00	0.00	0.00	0.00
0.44	0.32	0.59	0.00	0.00	0.21	0.69	-0.15	0.17	0.00	0.00	0.23	0.08	-0.16	0.09	0.16	0.00	0.01	-0.21
0.76	0.70	0.41	0.02	-0.02	0.69	3.54	-0.52	0.90	0.01	0.01	1.11	1.59	0.21	-0.12	-0.12	0.03	0.04	-0.78
-0.28	-0.20	-0.37	0.00	0.00	-0.15	-0.52	0.11	-0.11	0.00	0.00	-0.16	-0.12	0.10	-0.09	-0.09	0.00	-0.01	0.17
0.23	0.24	0.15	0.01	-0.01	0.17	0.90	-0.11	0.32	0.00	0.00	0.33	0.25	0.06	-0.21	-0.02	0.01	0.01	-0.09
0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	-0.01	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	-0.01
0.33	0.29	0.27	0.01	-0.01	0.23	1.11	-0.16	0.33	0.00	0.00	0.38	0.36	0.02	-0.11	0.01	0.01	0.01	-0.20
-0.60	-0.33	-1.46	0.00	-0.01	0.08	1.59	-0.12	0.25	0.00	0.01	0.36	1.71	0.71	-0.05	-0.63	0.03	0.03	-0.38
-0.71	-0.43	-1.32	0.00	0.00	-0.16	0.21	0.10	0.06	0.00	0.00	0.02	0.71	0.54	-0.29	-0.48	0.01	0.01	0.09
0.29	0.11	0.58	-0.01	0.01	0.09	-0.12	-0.09	-0.21	0.00	0.00	-0.11	-0.05	-0.29	0.51	0.22	-0.01	0.01	-0.25
0.65	0.40	1.18	0.00	0.00	0.16	-0.12	-0.09	-0.02	0.00	0.00	0.01	-0.63	-0.48	0.22	0.42	-0.01	-0.01	-0.07
-0.01	0.00	-0.02	0.00	0.00	0.00	0.03	0.00	0.01	0.00	0.00	0.01	0.03	0.01	-0.01	-0.01	0.00	0.00	0.00
0.00	0.00	-0.01	0.00	0.00	0.01	0.04	-0.01	0.01	0.00	0.00	0.01	0.03	0.01	0.01	-0.01	0.00	0.00	-0.01
-0.31	-0.21	-0.37	0.00	0.00	-0.21	-0.78	0.17	-0.09	0.00	-0.01	-0.20	-0.38	0.09	-0.25	-0.07	0.00	-0.01	0.33

In a correlated model, the covariance matrix of the joint distribution can be calculated by taking the outer square product of the Π matrix. This table shows the covariance matrix for the correlated model with three independent random shocks. The results for the correlated model with two independent random shocks is very similar. For several key variables, such as the threshold parameters and the coupon gap, it is clear that there is a significant correlation between the coefficients. This suggests that to accurately model borrower behavior it is necessary to account for correlation over the full life of the mortgage.

Table 3.15: Average Simulated Derivatives of Transition Probabilities - Comparison of Random Coefficients Models

	Independent Model			**Correlated Model I**			**Correlated Model II**		
	Age	Org. LTV	Coup. Gap	Age	Org. LTV	Coup. Gap	Age	Org. LTV	Coup. Gap
From a Curtailed State									
Prepayment	0.11	-0.03	4.28	0.10	-0.03	4.78	0.06	-0.03	5.01
Curtailment	0.02	-0.01	0.93	-0.04	0.00	-2.78	-0.03	0.02	-2.33
Current	-0.04	0.01	-1.65	-0.03	0.01	-0.99	-0.01	0.005	-1.07
30-Day Delinquency	-0.09	0.03	-3.56	-0.04	0.02	-1.02	-0.02	0.01	-1.61
From a Current State									
Prepayment	0.03	-0.03	2.74	0.09	-0.05	3.24	0.07	-0.04	2.97
Curtailment	0.02	-0.02	1.61	0.03	-0.05	0.48	0.04	-0.04	0.56
Current	-0.01	0.005	-0.38	-0.12	0.07	-3.70	-0.08	0.04	-3.16
30-Day Delinquency	-0.04	0.05	-3.96	-0.002	0.03	-0.03	-0.04	0.04	-0.38
From 30-Day Delinquency									
Prepayment	0.36	-0.09	0.84	0.16	-0.07	0.34	0.23	-0.11	1.32
Curtailment	0.25	-0.08	0.66	0.15	-0.05	0.43	0.23	-0.02	0.35
Current	0.0004	-0.03	0.08	0.12	-0.08	0.02	0.14	-0.06	-0.97
30-Day Delinquency	-0.52	0.16	-1.31	-0.32	0.17	-0.46	-0.44	0.16	-0.07
60-Day Delinquency	-0.10	0.04	-0.27	-0.11	0.03	-0.34	-0.16	0.03	-0.62

Under the correlated models, for loans in a curtailed state an increase in the coupon gap is associated with a lower probability of remaining in a curtailed state, while under the independent model an increase in the coupon gap implies a higher probability of remaining in a curtailed state. Also with respect to the coupon gap, for loans in a current state the correlated model predicts a larger positive derivative on the probability of prepayment, offset by a larger negative derivative on the probability of remaining in a current state. For loans in 30-day delinquency, the derivatives calculated under the correlated models are, in general, smaller in magnitude than those calculated under the independent model.

Bibliography

Allison, P. D., 1995, *Survival Analysis Using SAS: A Practical Guide* (SAS Publishing, Cary, NC).

Ben-Akiva, M., and M. Bierlaire, 1999, Discrete choice methods and their applications in short term travel decisions, in R. W. Hall, ed., *Handbook of Transportation Science*, 5–33 (Kluwer Academic Publishers, Dordrecht, The Netherlands).

Ben-Akiva, M., and D. Bolduc, 1996, Multinomial probit with a logit kernel and a general parametric specification of the covariance structure, MIT Working Paper.

Ben-Akiva, M., D. Bolduc, and J. Walker, 2001, Specification, identification, & estimation of the logit kernel (or continuous mixed logit) model, MIT Working Paper.

Ben-Akiva, M., and S. Lerman, 1985, *Discrete Choice Analysis: Theory and Application to Travel Demand* (The MIT Press, Cambridge, MA).

Bhat, C., 1999, An analysis of evening commute stop-making behavior using repeated choice observations from a multi-day survey, *Transportation Research: Part B* 33, 495–510.

Bhat, C., 2000, Incorporating observed and unobserved heterogeneity in urban work travel mode choice modeling, *Transportation Science* 34, 228–238.

Black, F., and M. Scholes, 1973, The pricing of options and corporate liabilities, *Journal of Political Economy* 8, 637–659.

Brennan, M. J., and E. S. Schwartz, 1985, Determinants of gnma mortgage prices, *American Real Estate and Urban Economics Association Journal* 13, 209–228.

Breslow, N. E., 1974, Covariance analysis of censored survival data, *Biometrics* 30, 89–99.

Calem, P. S., and M. LaCour-Little, 2004, Risk-based capital requirements for mortgage loans, *Journal of Banking and Finance* 28, 647–672.

Capozza, D. R., and T. A. Thomson, 2005, Optimal stopping and losses on subprime mortgages, *Journal of Real Estate Finance and Economics* 30, 115–131.

Clapp, J. M., G. M. Goldberg, J. P. Harding, and M. LaCour-Little, 2001, Movers and shuckers: Interdependent prepayment decisions, *Real Estate Economics* 29, 411–450.

Cox, D. R., 1972, Regression model and life-tables, *Journal of the Royal Statistical Society, Series B* 34, 187–220.

Dagsvik, J. K., 1994, Discrete and continuous choice max-stable processes and independence from irrelevant alternatives, *Econometrica* 62, 1179–1205.

Daly, A., and M. Bierlaire, 2006, A general and operational representation of generalised extreme value models, *Transportation Research Part B* 40, 285–305.

DeFranco, R., 2002, *Modeling Residential Mortgage Termination and Severity Using Loan Level Data*, Ph.D. thesis, Department of Economics, University of California, Berkeley.

Deng, Y., J. M. Quigley, and R. V. Order, 2000, Mortgage terminations, heterogeneity, and the exercise of mortgage options, *Econometrica* 68, 275–307.

Downing, C., R. Stanton, and N. Wallace, 2005, An empirical test of a two-factor valuation model: How much do house prices matter?, *Real Estate Economics* 33, 681–710.

Duffie, D., and K. Singleton, 1999, Modeling term structures of defaultable bonds, *Review of Financial Studies* 12, 687–720.

Dunn, K. B., and J. J. McConnell, 1981a, A comparison of alternative models for pricing GNMA mortgage-backed securities, *Journal of Finance* 36, 471–483.

Dunn, K. B., and J. J. McConnell, 1981b, Valuation of mortgage-backed securities, *Journal of Finance* 36, 599–617.

Dunn, K. B., and C. S. Spatt, 2005, The effect of refinancing costs and market imperfections on the optimal call strategy and the pricing of debt contracts, *Real Estate Economics* 33, 595–617.

Efron, B., 1977, The efficiency of cox's likelihood function for censored data, *Journal of the American Statistical Association* 72, 557–565.

Gourieroux, C., and A. Monfort, 1996, *Simulation-Based Econometric Methods* (Oxford University Press, New York, NY).

Green, J., and J. B. Shoven, 1986, The effects of interest rates on mortgage prepayments, *Journal of Money, Credit, and Banking* 18, 41–59.

Green, R. K., and M. LaCour-Little, 1999, Some truths about ostriches: Who doesn't prepay their mortgages and why they don't, *Journal of Housing Economics* 8, 233–48.

Greene, W. H., and D. A. Hensher, 2003a, A latent class model for discrete choice analysis: Contrasts with mixed logit, *Transportation Research: Part B* 37, 681–698.

Greene, W. H., and D. A. Hensher, 2003b, The mixed logit model: The state of practice, *Transportation* 30, 133–176.

Hajivassiliou, V. A., and P. A. Ruud, 1994, Classical estimation methods for ldv models using simulation, in R. F. Engle, and D. L. McFadden, eds., *Handbook of Econometrics*, volume IV, 2383–2441 (Elsevier Science, New York, NY).

Han, A., and J. A. Hausman, 1990, Flexible parametric estimation of duration and competing risk models, *Journal of Applied Econometrics* 5, 1–28.

Hayre, L., 2001, *SalomonSmithBarney Guide to Mortgage-Backed and Asset-Backed Securities* (John Wiley & Sons, New York, NY).

Heckman, J., and B. Singer, 1984a, The identifiability of the proportional hazard model, *Review of Economic Studies* LI, 231–241.

Heckman, J., and B. Singer, 1984b, A method for minimizing the impact of distributional assumptions in econometric models for duration data, *Econometrica* 52, 271–320.

Jenkins, S. P., 1995, Easy estimation methods for discrete-time duration models, *Oxford Bulletin of Economics and Statistics* 57, 129–138.

Kaplan, E. L., and P. Meier, 1958, Nonparametric estimation from incomplete observations, *Journal of American Statistical Association* 53, 457–481.

Kau, J. B., D. C. Keenan, W. J. Muller, and J. F. Epperson, 1992, A generalized valuation model for fixed-rate residential mortgages, *Journal of Money, Credit and Banking* 24, 279–299.

Kau, J. B., D. C. Keenan, W. J. Muller, III, and J. F. Epperson, 1995, The valuation at origination of fixed rate mortgages with default and prepayment, *Journal of Real Estate Finance and Economics* 11, 5–39.

Lancaster, T., 1990, *The Econometric Analysis of Transition Data* (Cambridge University Press, New York, NY).

Lin, C.-C., and T. T. Yang, 2005, Curtailment as a mortgage performance indicator, *Journal of Housing Economics* 14, 294–314.

Longstaff, F. A., 2004, Optimal recursive refinancing and the valuation of mortgage-backed securities, NBER Working Papers: 10422.

Longstaff, F. A., and E. S. Schwartz, 1995, A simple approach to valuing risky fixed and floating rate debt, *Journal of Finance* 50, 789–819.

Marschak, J., 1960, Binary choice constraints on random utility indications, in K. Arrow, ed., *Stanford Symposium on Mathematical Methods in the Social Sciences*, 312–329 (Stanford University Press, Stanford, CA).

McFadden, D., 1984, Econometric analysis of qualitative response models, in Z. Griliches, and M. D. Intriligator, eds., *Handbook of Econometrics*, volume II, 1396–1457 (Elsevier Science, New York, NY).

McFadden, D., and P. A. Ruud, 1994, Estimation by simulation, *The Review of Economics and Statistics* 76, 591–608.

McFadden, D., and K. Train, 2000, Mixed mnl models for discrete response, *Journal of Applied Econometrics* 15, 447–470.

Merton, R. C., 1973, Theory of rational option pricing, *Bell Journal of Economics and Management Science* 4, 141–183.

Pavlov, A. D., 2001, Competing risks of mortgage termination: Who refinances, who moves, and who defaults?, *Journal of Real Estate Finance and Economics* 23, 185–211.

Pennington-Cross, A., 2006, The duration of foreclosures in the subprime mortgage market: a competing risks model with mixing, Federal Reserve Bank of St. Louis, Working Papers: 2006-027.

Ramlau-Hansen, H., 1983, Smoothing counting process intensities by means of kernel functions, *The Annals of Statistics* 11, 453–466.

Revelt, D., and K. Train, 1998, Mixed logit with repeated choices: Households' choices of appliance efficiency level, *Review of Economics and Statistics* 80, 647–657.

Richard, S. F., and R. Roll, 1989, Prepayments on fixed rate mortgage-backed securities, *Journal of Portfolio Management* 15, 73–82.

Ruud, P. A., 2000, *An Introduction to Classical Econometric Theory* (Oxford University Press, New York, NY).

Schwartz, E. S., and W. N. Torous, 1989, Prepayment and the valuation of mortgage-backed securities, *Journal of Finance* 44, 375–392.

Small, K. A., 1987, A discrete choice model for ordered alternatives, *Econometrica* 55, 409–424.

Stanton, R., 1995, Rational prepayment and the value of mortgage-backed securities, *The Review of Financial Studies* 8, 677–708.

Stern, S., 1997, Simulation-based estimation, *Journal of Economic Literature* 35, 2006–2039.

Thurstone, L., 1927, A law of comparative judgement, *Psychological Review* 34, 273–286.

Train, K. E., 2003, *Discrete Choice Methods with Simulation* (Cambridge University Press, New York, NY).

Walker, J., 2001, *Extended Discrete Choice Models: Integrated Framework, Flexible Error Structures, and Latent Variables*, Ph.D. thesis, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology.

Wang, X., and K. M. Kockelman, 2005, Occupant injury severity using a heteroscedastic ordered logit model, *Transportation Research Record* 195–204.