THREE ESSAYS ON MORTGAGE DEFAULT AND PREPAYMENT

by

XIAOWEI LI

(Under the Direction of Donald C. Keenan)

ABSTRACT

Investigating the residential mortgage defaults and prepayments has been the subject of research for the past three decades. The literature on the probability of the mortgage default and prepayment is often used to inform credit risk policy and asset pricing strategy. This literature has evolved from the use of logistic regressions to the use of survival and frailty models that control for unobserved heterogeneity.

I apply a shared-frailty survival model to analyze the mortgage termination risks. In particular, I investigate whether mortgage originated in the same Metropolitan Statistical Area (MSA) share common unobserved factors and how these factors affect the mortgage termination risks. A similar approach is applied to examine the group-level frailty effect for mortgages with the same origination year.

INDEX WORDS: Mortgage, Shared Frailty, Survival Model, MSA, Redlining

THREE ESSAYS ON MORTGAGE DEFAULT AND PREPAYMENT

by

XIAOWEI LI

M.A., University of South Carolina, 2004

LL.B., Peking University, P.R. China, 2002

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2008

THREE ESSAYS ON MORTGAGE DEFAULT AND PREPAYMENT

by

XIAOWEI LI

Major Professor: Donald C. Keenan

Committee: Scott E. Atkinson
James B. Kau
Jason Seligman

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
August 2008

DEDICATION

For my family, who offered me unconditional love and support throughout the course of this dissertation.

ACKNOWLEDGEMENTS

I would never have been able to finish my dissertation without the guidance of my committee members, help from my friends, and support from my family.

I would especially like to thank my major advisor, Dr. Donald Keenan, for his excellent guidance, caring, and efforts to help me finish my dissertation. I am also grateful to my dissertation committee members, Dr. Atkinson, Dr. Kau and Dr. Seligman, for their constructive comments and valuable advice.

I extend many thanks to my friends, Yue Fu, Bing Xu, Sharri Byron, Yi Yang, Xia Liu, Jinping Guan, for their encouragement and support.

Finally, I'd like to thank my family. My parents, Hengzhang Li and Shuying Jin, are a constant source of support. I am especially thankful to my husband, Bin Lu, for his patience and for being with me through the ups and downs. His knowledge of programming and computers was also extremely helpful.

The research was partially funded by the University of Georgia Graduate School Dean's Award. I extend my thanks to the Graduate School for their financial support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION

Investigating the residential mortgage terminations, via either default or prepayment, has been the subject of research for long time. Most of the research in the studies of mortgage termination can be divided into two groups: the theoretical approaches in modeling mortgage terminations (via either default or prepayment), and the structural option-theoretic approach and the reduced-form approach. The option-theoretic method models default (i.e. terminate mortgage payments in return to giving up the possession of the house) as a put option and prepayment (i.e. pay off the loan to get the ownership of the house) as a call option (Dunn and McConnell, 1981; Foster and Van Order, 1984; Titman and Torous, 1989; Kau et al. 1992, 1995). Kau and Keenan (1995) provide a complete survey of option-theoretical models of mortgage pricing. Several of the empirical estimation studies on mortgage termination risks have been conducted in the framework of survival models (Green and Shoven, 1986; Quigley and Van Order, 1991). First introduced by Green and Shoven (1986), Cox's proportional hazards model (PHM) has been widely used in the literature of mortgage termination risks and demonstrated to be effective.

The major assumption that survival models were based on is that individuals' survival times are independent, conditional on the included observed covariates. This assumption, however, may cause the estimates to be biased. Some literature, like the economics duration literature has documented a "mass-point" approach to take account of the unobserved heterogeneity (Deng et al, 2000; Ciochetti et al, 2002; Pennington-Cross, 2003). Another more natural way to model the within-group correlation between individual survival times is a shared-frailty survival model

(Lancaster, 1990; Klein, 1992; Sastry, 1997). This dissertation contributes to the mortgage termination studies by applying the shared-frailty survival models to examine whether survival times of mortgages originated in the same group (in either same region or same year) are correlated with each other. The magnitude of the correlation is also estimated.

Another empirical research this dissertation has conducted is to investigate how the mortgage contract rates, default rates, and prepayment rates are influenced by the socioeconomic characteristics of neighborhoods. Redlining studies in regarding to studies of the discrimination against the low-income and minority neighborhoods have been drawn great attention for a while. Most of these studies find little evidence of differential treatment of race (Avery and Buynak, 1981; Gabriel and Rosenthal, 1991; Schill and Wachter, 1993; Holmes and Horvitz, 1994). However, even though loan denial rates may not vary by neighborhood characteristics, redlining may occur in a more subtle form, such as a variation in loan pricing by neighborhood. Using a rich dataset with loan-characteristics and external socioeconomic information, I examine whether mortgage rates on fixed-rate loans vary by the income and racial composition of the neighborhood.

This dissertation is organized as follows: chapter 2 is the study of the shared-frailty survival model for those mortgages originated in the same Metropolitan Statistical Area (MSA); chapter 3 presents whether mortgages originated in the same year are correlated in their survival times; chapter 4 is the neighborhood studies of how mortgage contract rates, and terminations rates are connected with the demographic and economic characteristics; chapter 5 concludes.

CHAPTER 2

AN ANALYSIS OF MORTGAGE TERMINATION RISKS: A SHARED FRAILTY APPROACH WITH MSA-LEVEL RANDOM EFFECTS

2.1. Introduction

Investigating the residential mortgage defaults and prepayments has been the subject of research for the past three decades. Much of the research stems from the importance of credit risk management of mortgage lending institutions and secondary market agencies. In addition, the modeling of the default and prepayment behavior of the underlying borrowers has contributed to the valuation and hedging of the mortgage securities (Schwartz and Torous, 1989; Kau et al., 1990; Ambrose and Sanders, 2003). There are two major theoretical approaches in modeling mortgage terminations (via either default or prepayment): the structural option-theoretic approach and the reduced-form approach. The option-theoretic method models default (i.e. terminate mortgage payments in return to giving up the possession of the house) as a put option and prepayment (i.e. pay off the loan to get the ownership of the house) as a call option (Dunn and McConnell, 1981; Foster and Van Order, 1984; Titman and Torous, 1989; Kau et al. 1992, 1995). Kau and Keenan (1995) provide a complete survey of option-theoretical models of mortgage pricing. The more recently developed reduced-form approach (Kau et al. 2006) assumes that the mortgage terminations process is a doubly-stochastic process[1] and these stochastic processes determine the price of the credit risk. Several of the empirical estimation studies on mortgage termination risks have been conducted in the framework of survival models (Green and Shoven, 1986; Quigley and Van Order, 1991). First introduced by Green and Shoven

[1] The doubly-stochastic property assumes that both the arrival of the termination and the hazard rate of the termination risk are random.

(1986), Cox's proportional hazards model (PHM) has been widely used in the literature of mortgage termination risks and demonstrated to be effective.

The first argument for the inclusion of the unobserved heterogeneity when estimating the risk factors in determining mortgage terminations is that it is almost impossible to account for all the variables necessary for estimation. The potential effect of the unobservable variables on the probability of duration has been studied and the proportional hazards model with the unobservable heterogeneity, or frailty, has been introduced. An additional motivation to use the frailty model is to relax the assumption of the "conditional independence" of the survival times of Cox model. A common assumption made in modeling the effects of potential risk factors on survival is that event times of the members of the population, conditional on the observed covariates, are statistically significant. In practice, it may be the case that the event times of members in same subgroups of the population are associated since members of these groups share a common unobserved trait. In mortgage termination studies, one mortgage may have an association between termination times among mortgages originated in the same region. If an association like this is ignored, the estimates of covariate effects are suspect. For these two arguments mentioned above, in this essay, I propose a shared-frailty model which incorporates the unobserved heterogeneity and the within-group associations into analysis.

The economics duration literature has documented a "mass-point" approach to take account of unobserved heterogeneity. This approach assumes that the unobserved heterogeneity is discretely distributed and observations are divided into finite unknown groups (Heckman and Singer, 1984). Initially applied in the unemployment duration studies (McCall, 1996), the empirical literature has adopted this approach in mortgage termination analysis recently (Deng et al, 2000; Ciochetti et al, 2002; Pennington-Cross, 2003). Their approach models individual mortgage borrowers as coming from two or more distinct groups with unobserved characteristics. An alternative modeling the frailty assumes a continuous distribution for the frailty variable (for example Lancaster (1990) presents a gamma distribution for the frailty variable). In addition to the different assumption of the frailty distribution, the second continuous approach assumes that frailties are not observation-specific but instead are shared across groups of observations, causing observations within the same group to be correlated. This approach serves as an extension of the standard Cox's model. Thus, the continuous-frailty approach is a natural way to model the correlation among survival times of mortgages. The second method, or the shared-frailty approach, has been widely applied in the biomedical and sociology research in determining how individuals' survivals in subgroups are associated (Klein, 1992; Sastry, 1997). Few studies of mortgage termination have applied the second approach. Among the few studies, Follain et al (1997) introduces a semi-parametric proportional hazards model with gamma-distributed investors' heterogeneity to study mortgage prepayments. Their findings suggest that allowing for the unobserved heterogeneity does improve the results substantially.

Within the framework of the shared-frailty, this essay uses an extensive and geographically diverse sample of single family fixed-rate mortgages to address the issue of possible correlation of mortgage survival times, controlling for the observed covariates. Along with other risk factors, the unobserved group-level factor could have an impact on the relative survival probabilities of mortgages because it changes the relative weights placed on different observed covariates. The primary objective of this essay is to examine how unobserved characteristics at the metropolitan statistical area (MSA) level affect the survival times of mortgages. The remainder of the essay is organized as follows: section 2 specifies the model; section 3 explains data and variables; section 4 displays the empirical results, and section 5 concludes.

2.2. The Survival Model with Shared Frailty

A hazard function gives the probability of mortgage termination during a particular time period, conditional on the mortgage not having been previously terminated. By expressing this conditional probability of default or prepayment as a function of various explanatory variables and frailty, one can assess the statistical significance of these variables in influencing mortgage terminations. Now let the continuous random variable T represent the time till the mortgage is terminated via either default or prepayment. The hazard function is defined by

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \le T < t + \Delta t \mid T \ge t)}{\Delta t} \qquad (1)$$

The probability of survival at time t can be given in terms of the hazard function:

$$S(t) = P(T \ge t) = \exp(-\int_0^t h(u)du \qquad (2)$$

Default and prepayment risks are assumed to be conditionally independent --- default and prepayment regressions are estimated separately treating the other type of the termination risk as censored. The independence assumption is stated to be "untestable" but the derivation of the partial likelihood function for the joint survival function was proved to produce the same results as the cause-specific hazard function (Kalbfleisch and Prentice, 2002). In this essay, a variable indicating the type of failure is specified as K, with the values ,$1,...,k$,in addition to the random duration variable T. Consider the existence of k random variables, $T^{(1)}, T^{(2)},...,T^{(k)}$, one for each destination, interpreted as latent duration. Only the smallest latent time period is observed, namely, $T = \min[T^{(1)},...,T^{(k)}]$. For each individual, only one $T^{(k)}$ is observed and others are considered censored. Thus, the cause-specific hazard rate, representing the instantaneous risk of termination of cause k can be expressed as:

$$h_k(t) = \lim_{\Delta t \to 0} \frac{P(t \le T < t + \Delta t, K = k \mid T \ge t)}{\Delta t} \tag{3}$$

By law of total probability, the overall hazard is:

$$h(t) = \sum_{k=1}^{k} h_k(t) \tag{4}$$

The mortgage termination event is denoted as $k, k = p, d$ referring to default and prepayment respectively. The sample data contains mortgages originated in I groups. Let n_i be the total number of mortgages in the ith group, and $n = \sum_{i=1}^{I} n_i$ be the total number of mortgages in the sample. Let T_{ij} denote the observed or censored lifetime of the jth mortgage in the ith group, and let $N_{ij}^{k}, k = p, d$ be the censoring indicator, where $N_{ij}^{k} = 1$ indicates the events of default or prepayment and $N_{ij}^{k} = 0$ indicates the censoring time. The total number of defaults (prepayments) in group i is given by $N_i = \sum_{j=1}^{n_i} N_{ij}^{d}$ ($N_i = \sum_{j=1}^{n_i} N_{ij}^{p}$). Let $X_{ij}(t)$ be a $1 \times K$ vector of covariates at time t. The vector $X_{ij}(t)$ usually includes both loan-specific variables and macroeconomic variables. Let $h_{kij}(t), k = p, d$ be the hazard function for the jth mortgage in the ith group. Mortgages within the same group may have dependent survival times due to some unobserved covariate information summarized in a frailty, v_{ki}. If , for example, one groups together mortgages originated from the same MSA, then the frailty may reflect the common environment or policy effect on survivals of all the mortgages. Note that if $g = n$, the size of the subgroup becomes 1 and in such a case the individual mortgage is affected by its own frailty. Suppose that the censoring is independent with the termination of mortgages. The hazard function for this individual mortgage is given by

$$h_{kij}(t, X_{ij}, \beta, \theta) = h_{k0}(t)\exp(X_{ij}'\beta_k)v_{ki}, k = p, d \tag{5}$$

where $h_{k0}(t)$ is a non-parametric baseline hazard function and β_k is the unknown coefficients. Conditional on the unobserved v_{ki}, the lifetime of mortgages in the ith group are independent. When the unknown v_{ki} is integrated out, the lifetimes become dependent; the dependence is induced by the common value of v_{ki}.

The shared frailty model specified in (5) is a natural approach for modeling dependence and taking into account of unobservable heterogeneity. The frailty has an assumed prior distribution which is updated as the default and prepayment information set evolves over time. The frailties are assumed to follow a certain distribution but the location of each group-specific frailty is not known. There is a range of choices for the distribution of the frailties --- the most popular is the gamma distribution $\Gamma(\alpha, \gamma)$, partly due to its inherent flexibility of the distribution with respect to the parameters of α and γ.[2] With gamma frailties, the scale parameter needs to be restricted for identification reasons, and the standard restriction is $\gamma = 1/\alpha$ which implies a mean value of one for the frailty variable. The specification of model (2) is completed by assuming that the group-level frailties v_{ki} are independent and identically distributed with a gamma distribution $\Gamma(\frac{1}{\theta}, \theta)$, with $\theta > 0$.

The likelihood function for jth mortgage in the ith subgroup is defined by

$$L_{ij}(t, X, \beta, \theta) = \{[h_{dij}(t, X, \beta, \theta)v_{di}]^{N_{ij}^d} S_{dij}(t, X, \beta, \theta)\}\{[h_{pij}(t, X, \beta, \theta)]^{N_{ij}^p} S_{pij}(t, X, \beta, \theta)v_{pi}\}$$

---

[2] The gamma density function of a random variable x is given by $f(x;\alpha,\gamma) = x^{\alpha-1}\dfrac{e^{-x/\gamma}}{\gamma^{\alpha}\Gamma(\alpha)}$ for $x > 0, \alpha > 0, \gamma > 0$. The expected value of x is $E(x) = \alpha\gamma$ and variance is $Var(x) = \alpha\gamma^2$. The parameter α is referred as the scale parameter and the γ as the shape parameter.

(6)

The overall likelihood function is:

$$L(t, X, \beta, \theta) = \sum_{j=1}^{n_i} \sum_{i=1}^{G} L_{ij}(t, X, \beta, \theta) \tag{7}$$

The parameters of this model, θ and β, can be estimated with the expectation-maximization (EM) algorithm (Dempster et al. 1977). To implement the EM algorithm, the expectation of the log likelihood function in equation (6) needs to be derived first. Secondly, this function is maximized with respect to the unknown parameters. The algorithm proceeds iteratively until the parameter estimates converge. The model is estimated using the statistical software STATA version 10[3]. The expected value of the frailty for each group can also be calculated.

2.3. Data and Variables

The essay uses data on 30-year fixed rate single-family residential mortgages from a large financial service institution. The data set contains 1,038,098 observations on individual mortgage loans issued between 1976 and 2004. For each mortgage, the available information includes the year and month of origination and termination (if it has been closed), indicators of termination --- defaulted, prepaid, sold, or censored, and a number of loan-specific characteristics observed at the time of origination. The characteristics available are the original loan amount, the original loan-to-value ratio, the contract rate at origination, the amount of points paid at origination, and the zip code where the property is located. The variable measuring borrowers' credit risk at origination, Fair Isaac Corporation (FICO) credit scores, are available

[3] The estimation of the shared-frailty model in STATA consists of two steps. In the first step, the optimization is in terms of θ only. For fixed θ, the second step consists of fitting a standard Cox model via penalized log likelihood, with v_{ki} introduced as estimable coefficients of dummy variables identifying the group. For more details of estimation, please see Gutierrez (2002) and the Survival Analysis and Epidemiological Tables Reference Manual (2007).

for about 48.7 percent of all the observations in the dataset[4]. The right-censoring issue also exists in the dataset: (1) 13,410 loans were sold to other banks after origination, and (2) 462,699 loans were out of follow-up after Dec 31, 2004. The survival analysis employed in this essay accounts for the right-censoring problem.

The decisions about the choice of variables included in the estimation are based on motivations for default and prepayment. The following is a short discussion of covariate specification. The current essay is not explicitly based on an option-theoretic pricing model, and the choice of covariates is more intuitive.

(1) Loan-to-value (LTV) ratio. Option-pricing theory predicts that the higher LTV is associated with the higher default risk due to a high possibility of negative equity. However, higher LTV ratio at the origination may indicate more strict underwriting standards, or high down payment requirements for risky loans, which in turn has a negative effect on the default risk. Thus, the predictive effect of the LTV ratio on the default risk is determined by these two opposing directions. For prepayments, higher LTV makes refinancing more costly and increases the value of the default. The higher LTV ratio is expected to decrease the prepayment probability. There is also an opposing effect of high LTV on the prepayment option due to the strict underlying restriction and thus the effect of LTV on prepayment risk depends on both effects.

(2) Original loan size and points paid at the origination. It is argued that a large original loan balance provides a large dollar incentive to prepayment and default (Clapp et al., 2001). However, if we view the loan size as a proxy for borrower's total wealth, we should expect wealthier borrowers to be less likely to default. Thus, the effect of loan size on default is mixed. Points paid at the origination, as a percentage of the original loan size, can be combined with the

[4] The FICO score is available after 1998. However, not every mortgage originated after 1998 was documented with its borrower's FICO score. In the dataset, around 73 percent of the mortgages originated after 1998 were attached with borrower's initial FICO scores at the origination.

interest rate to estimate the speed of prepayment. For a given contract rate, loans with low points are expected to be prepaid more rapidly (Stanton and Wallace, 1998). For defaults, higher points may lead to negative default rate since the effective costs of the default are higher.

(3) The interest rate spread. A mortgage can be viewed as an annuity of fixed prepayment. When the market interest rate is different from the contract rate at the origination, the present value of future payments must be discounted at the market rate. Thus, a decrease of market interest rate makes prepayments more valuable and thus, declines the probability of default risk. On the other hand, a low market interest rate makes the borrowers tend to default and move to another residence, given the transaction costs are low. Therefore, the effect of mortgage interest rate on default risk can be characterized as netting out two opposing effects. Although the market interest rate is a key factor in predicting prepayment behavior, the decision to prepay could depend on how long rates have been down and on whether they are expected to drop further. Thus, the recording date of refinancing may be much later than the original rate of the borrower's decision. For this reason, following Schwartz and Torous (1989), I use the two-month lagged value of the average yield on 10-year US Treasury Bonds as a proxy for market interest rate. The relative difference between the current contract rate at the origination and the two-month lagged interest rate is denoted as the variable spread.[5]

(4) Housing price dynamics. The housing price index reflecting the appreciation trend of housing prices plays an important role in observing mortgage termination behavior. In this essay quarterly Housing Price Index (HPI)[6] from Office of Federal Housing Enterprise Oversight (OFHEO) was used to measure the state-level movements of the single family prices. It is found

---

[5] The spread is calculated as: $(r_c - r)/r_c \times 100$, where r_c is the contract rate and r is the proxy of the market interest rate.

[6] The HPI is a weighted, repeated sales-index, meaning that it measures average price changes in repeat sales or refinancing on the same properties. For more information, see the OFHEO website: http://www.ofheo.gov/.

that high default rates on home mortgages strongly tend to follow real estate price declines (Case et al., 1996). According to Mattey and Wallace (2001), differences in house price dynamics across regions are an important source of the heterogeneity of mortgage prepayment rates. I use the price index ratio, which is calculated as house price index at the time of mortgage termination divided by the index at the time of mortgage origination, to reflect the relative change of the housing price changes in the same state. It is expected that both default rates and prepayments rates are negatively related with the price index ratio: as the house price increases, the option of selling a house is showing a profit and this can lower both the risks of defaults and prepayments.

(5) Borrower default and prepayment may also depend on local economic conditions. For example, borrowers may default due to a “trigger event”, such as a loss in income due to an unexpected job loss. To capture this effect, I include the monthly state-unemployment rate at loan termination as an indicator of general economy. It is known that defaults are negatively related with the economy. In bad economy with high unemployment rates, it is expected that the default rate is high. Also it is expected that the unemployment rate is negatively related with the prepayment rate since borrowers who lose their jobs may not be able to refinance.

(6) The FICO score at the origination, and other demographic variables. Since credit history is a key determinant of mortgage loan approval, it clearly should have some bearings on the likelihood of mortgage termination. In our current dataset, FICO scores are available for most of the borrowers with mortgages originated after 1998. Other demographic variables, such as the racial composition and the median household income at the ZIP code level, are available from the Bureau Census 2000 survey data[7]. For a test of the how the FICO score and other

[7] The racial composition and median income for neighborhoods are relatively stable. I merged subsample of mortgage (which were originated after 1998) data with the Bureau Census 2000 survey data. Thus, for each

demographic variables affect mortgage termination risks, I include the FICO score, racial composition and household income, together with other loan-specific information, to analyze the sample data with mortgages originated after 1998. The demographic variables from Bureau Census 2000 survey dataset are assumed to be relatively stable between 1998 and 2004. It is expected that borrowers with poor credit ratings will have difficulty finding refinance options and are more likely to default on a mortgage. On the other hand, borrowers with higher FICO scores are expected to prepay quickly than those with lower FICO scores. The median household income per ZIP code are expected to be negatively associated with default rates and positively related with prepayment risks.

To examine whether mortgages are associated in their survival times, I consider an application of the shared frailty model in the mortgage termination study. Here, individual mortgages were grouped by metropolitan statistical areas (MSAs) where they were originated. An MSA is a geographical area, defined by the federal Office of Management and Budget (OMB) that represents the metropolitan area of a city. In most of the United States, MSAs consist of a county or a group of counties. The real estate and mortgage literature have used MSA as the analysis unit because of its socioeconomic composition (Megbolugbe and Cho, 1993; Capozza et al, 1997). The use of MSA also allows for the control of spatial variation for the local economic risk (Ambrose and Pennington-Cross, 2000). The frailty in this case may be considered as the combined effect of shared demographic and regional policy effect on individual borrowers. In addition, unmeasured socioeconomic factors may also be important at the MSA level. For example, individuals are attracted to move to areas with certain industry clusters. Families who share similar racial and cultural backgrounds tend to cluster. The results

mortgage which was originated after 1998, each one has ZIP code-level information of racial composition and median income characteristics.

presented in this section are empirical results of estimating default and prepayment risks separately, as well as accounting for the shared frailty at the MSA level.

The entire dataset contains 280 MSAs where mortgages were originated between 1976 and 2004. The number of the mortgages originated in each MSA ranges from 3 to 90,669. To avoid the size differences among MSAs while maintaining the sufficient flexibility of a large data set, I chose the top 44 largest MSAs in terms of the number of the originated mortgages. In each selected MSA, there are more than 4,000 originated mortgages across all years. The total sample size is now restricted to be 734,721. The summary statistics for all loans and loans originated after 1998 are reported in Table 1 and Table 2. In order to test the additional demographic variables on the probability of default and prepayment, I added the demographic information to the mortgages originated after 1998 from the Bureau Census 2000 dataset. Table 3 and Table 4 display the number of loans originated for each MSA, along with the percentage of each region that defaulted or were prepaid during the period of observation. Table 5 and Table 6 display the default and prepayment distribution for mortgages originated after 1998. Note in panel (a), the total number of mortgages originated ranges from 4,038 in Oklahoma City, OK to 90,669 in Los Angeles, CA. The average percent of mortgages prepaid is 31%. The highest prepayment rates of mortgages concentrated in the Midwest area, state of Colorado, and state of Washington. The percentage defaulted shows a different pattern across regions. Mortgages originated in Honolulu, HI have the highest level default rate of 3.67% during the observation period (note that Honolulu area has the highest default rates of 1.08% as well, among those areas with mortgages originated after 1998). Other areas with high defaults rates include the Mid-Atlantic region and New York City.

Figure 1 and figure 2 illustrate smoothed non-parametric Kaplan-Meier estimates of both default and prepayment hazard in each month of mortgage life. The hazard of defaults seems to have a tendency to increase monotonically over the first 10 years of time. After the first 10 years, the default hazard rate begins to fall. The default hazard rates stabilize after about 15 years. For the hazard of prepayment, the trend is similar with that of the default hazard --- rising for the first 10 years and then declining after that. The predominantly higher prepayment rates can be explained by the social reasons which are related with family growth, family break up or regional economic developments (Spahr and Sunderman, 1992). One thing that is interesting to note is that, instead of stabilizing, the prepayment hazard rate begins to increase again. This increase can be explained by the fact that the remaining balance of loans becomes increasingly small.

2.4. Empirical results

In this section I discuss the results of estimating the shared-frailty model described in section 2. In the following analysis, the included covariates are the LTV ratio, the log form of the original loan size[8], the points paid at the origination, original contract rate, the interest rate spread, the monthly state unemployment rate and the HPI ratio at the state level.

2.4.1. No frailty

Table 7 and Table 8 report the results obtained from a standard Cox analysis assuming independence of the survival times. Both the results for the default and prepayment risks are presented. For the estimation of default risk using the full sample, all coefficients have the expected signs and are significant, except for the unemployment rate. The estimation for mortgages originated after 1998 shows that the probability of default is negatively associated with the FICO score, the percentage of African-American population per ZIP code, and the

---

[8] The log loan size variable was formed by inflating a loan's size to match the Freddie Mac 2001 Conventional Mortgage Home Price Index (CMHPI) (available at www.freddiemac.com/finance/cmhpi/), before being taken into logs.

median household income per ZIP code. For both pre- and after 1998 estimations, the original LTV ratio is found to have a positive effect on the probability of default, which confirms the hypothesis that higher original LTV serves as an indicator of possible future negative equity.

For the estimation of the prepayment risk, all covariates are significant and have the correct sign in the full sample. Data for mortgages originated after 1998 were used to test the effects of FICO score and the demographic variables: the results[9] show that prepayment probability is positively related with the FICO score, the state-level annual GDP and the percentage of the Hispanic population per ZIP code while the likelihood of prepayment is negatively related with the median household income and the percentage of African-American population per ZIP code. The sign of LTV is negative, which means that borrowers with higher LTV at the origination are motivated to pay off faster.

The estimated coefficients for the adjusted loan balance at origination, original contract rate, and the spread are all positive and significant in both default and prepayment models. In contrast, points, state level unemployment rate and the house price dynamics exert negative effects on both default and prepayment risks.

2.4.2. One frailty per group

The shared frailty models are used to model within-group correlation in that observations within a group are correlated since they share the same frailty. In this case the associations between duration times of mortgages within a MSA are modeled as a random effect term, an unobserved covariate common to default or prepayment events experienced by a MSA. If no covariates are included in the model, frailty in this model captures all factors that influence the risk of mortgage termination that are not included in the baseline function. Since the model can

[9] The subsample estimation of the prepayment risk is not robust in terms that the sign of the coefficient of SPREAD has been changed. However, data for mortgages originated after 1998 can still be viewed as a valid test since the signs of the FICO score and the demographic variables are consistent for all three regressions.

account for observed covariates, the frailty effects represent the total affects of the unobserved factors on survival chances. The idea is that some clusters of mortgages have different frailties, the extent of which is measured by θ. If the null hypothesis $H_0 : \theta = 0$ is rejected, then the within-group correlation is statistically significant. In addition to describing the within-group correlation, the frailty distribution can be used to display the cross-group heterogeneity. Thus, an alternative way to interpret the distributions of the random effects is to construct a risk ratio that compares the expected frailty value for a high-risk group and a low-risk group. A risk ratio of one indicates that observations in different groups are homogenous while a larger risk ratio indicates greater heterogeneity.

As shown in table 9, the estimated variance θ is equal to 0.6426 for the full sample as a random geographical effect on the risk of default and 0.4363 on prepayment. In both cases θ is significantly different from 0 so that there are meaningful associations between mortgages which were originated in the same MSA. It also indicates that there is significant heterogeneity between mortgages originated in different MSAs. Note that the variance value of the default is greater than that of the prepayment risk. It implies that default-related unobserved risks at the MSA are more dispersed than the prepayment-related unobserved factors.

Before comparing coefficients or marginal effects, it is valuable to assess whether the more complex model offers a significant improvements in terms of overall fit. The conventional Cox model is nested within the shared-frailty model (the former being equivalent to the latter subject to the constraints that the variance of the frailties is equal to zero), so one can apply a likelihood-ratio test to assess the contribution of the heterogeneity terms. The Chi-squared values in table 4 confirm that, for the same sets of covariates, a frailty model provides a statistically significant improvement in fit over a model without frailty.

Turning to the coefficients, the estimated coefficients based on the shared frailty model in table 9 are volatile. Follain et al (1997) noted that the estimated coefficients are larger in the frailty model than the non-frailty one. However, the results in the current study show that the estimated coefficients from the shared-frailty model are not necessarily larger than that of the standard model. Examining parameters that are statistically significant in both models at 95 percent level or higher, I find the change in coefficient magnitudes ranges from 3.7 percent (the spread s effect on the prepayment risk) to 139 percent (the logarithmic form of loan size measure's effect on the default hazard).

As a more helpful measure of the impact of the independent variables, I calculated he percent change in the risk-specific hazard implied by a hypothesized change in a variable, holding all others fixed at their sample means. For the most part, there are few large differences between the models. Loan-to-value ratio, for example, is shown to increase the default hazard 6.0 percent in the non-frailty model, and 8.1 percent in the frailty model. In pure percentage terms, the greatest statistically significant shift between specifications is observed in the effect of house price index ratio on the default hazard. The non-frailty model suggests that the risk of failure via default will be increase 0.01 percent for a one unit increase of the house price index ratio, while the frailty model suggests that shift will be 0.04 percent --- a relative change of 99 percent in the magnitude of the effect. Finally, I note that coefficient for "state-level unemployment rate", which changes sign from negative in the non-frailty model to positive in the frailty model via the default hazard. In the non-frailty model, a one percent increase of the "state-level unemployment rate" is shown to lower the default 8.3 percent while in the frailty model it is shown to increase the default risk 11.4 percent, holding all other variables constant. The positive sign of the state-level unemployment rate is consistent with our expectation --- in a bad economy with high state-level unemployment rate, the default rates are always high. This change of the effect of state-level unemployment rate on the default rate can be explained that the frailty factor integrates out the unobserved factors that may bring correlation with the unemployment rate in the non-frailty model.

Table 10 and figure 3 display the expected frailties values for each MSA grouped by the Bureau Census regions (West, Midwest, South, and New England)[10]. It is interesting to note that the expected frailty values for both default and prepayment risks in the same region show similar patterns. For example, a simple calculation of the Pearson correlation coefficient between default and prepayment frailty values demonstrates that they are correlated. The correlation coefficients for the two arrays of frailty values are 0.67 in South and 0.59 in Midwest. The correlation coefficient between default frailties and prepayment frailties in West is -0.285. In the New England region, however, the frailty values for the default and prepayment hazards are barely correlated with a correlation coefficient of 0.035. These correlation coefficients indicate that frailties show different spatial pattern in the nearby MSAs. For most of the MSAs in the same region (South, Midwest, and West), the default frailty and prepayment frailty values are inter-correlated.

2.5. Conclusion

In the analysis of mortgage termination using the hazard model approach, situations where the survival times of mortgages are not independent are often encountered. In particular, for example, mortgages originated from the same area may be similar in terms of duration times. The shared-frailty model provides a method for modeling survival data when the survival times are not independent. The frailty, representing the effects of measurement errors and missing

[10] The definition of the bureau census regions can be found in the website: http://www.census.gov/geo/www/us_regdiv.pdf.

variables, is modeled as a non-negative latent random variable that acts multiplicatively on the hazard function. As an experiment, this essay uses MSA as the group variable. The empirical results suggest that it is important to control for the MSA-level frailty to account for the within-group correlation among individual mortgages. Differences in environment and in socioeconomic setting are likely to have an important influence on mortgage termination risks.

Table 2.1: Summary Statistics for Loans Originated in Selected MSAs.
(n=734,721)

Variables	Mean	Median	Std.	Min.	Max.
Duration Time (month)	51.07	28	52.10	0	347
LTV (%)	75.12	79	18.55	20	125
Original Loan Amount	139,543	120,000	81,229	20,000	452,000
Points(%)	-0.01	0	0.65	-3	8
FICO Score	721.15	724.19	54.65	427	849
Original Contract Rate	7.06	7	1.28	3.75	19
Spread (%)	24.10	24.57	9.62	-222.82	62.36
State Unemployment Rate	5.14	5.3	1.04	2.1	12.3
State House Price Index	324.78	297.12	115.30	97.88	674.32

Table 2.2: Summary Statistics for Loans in Selected MSAs Originated after 1998.
(n=486,419)

Variables	Mean	Median	Std.	Min.	Max.
Duration Time (month)	27.47	20	20.81	0	83
LTV (%)	73.42	77.67	19.04	20	125
Original Loan Amount	158,842	138,320	84,510	20,000	452,000
Points(%)	-0.06	0	0.73	-3	8
FICO Score	728.25	736	54.28	427	842
Original Contract Rate	6.48	6.25	0.94	3.75	13.25
Spread (%)	28.10	28.12	6.90	-37.60	62.2642
State Unemployment Rate	5.25	5.5	0.89	2.2	8.1
State House Price Index	355.10	335.84	111.92	140.14	674.32
State-Level Annual GDP ($1000)	35.47	34.95	7.73	22.49	116.44
Median Household Income per ZIP code	54,574	52,303	18,869	0	200,001
African-American percentage per ZIP code	10.98	4.22	17.01	0	98.20
Hispanic percentage per ZIP code	15.10	8.44	17.46	0	97.22

Table 2.3: Loan Origination and Defaults by MSA
(n=734,721)

MSA Name	# Originated	# Defaulted	% Defaulted
Albuquerque-NM	5,753	14	0.24%
Atlanta-GA	25,854	54	0.21%
Austin-San Marcos-TX	6,938	9	0.13%
Boston-Worcester-Lawrence-MA-NH-ME-CT	16,349	28	0.17%
Buffalo-Niagara Falls-NY	4,849	21	0.43%
Charleston-North Charleston-SC	4,051	6	0.15%
Charlotte-Gastonia-Rock Hill-NC-SC	14,263	19	0.13%
Chicago-Gary-Kenosha-IL-IN-WI	27,557	116	0.42%
Cincinnati-Hamilton-OH-KY-IN	5,514	22	0.40%
Cleveland-Akron-OH	5,478	39	0.71%
Columbia-SC	4,125	5	0.12%
Columbus-OH	4,213	13	0.31%
Dallas-Fort Worth-TX	23,591	111	0.47%
Denver-Boulder-Greeley-CO	21,901	48	0.22%
Detroit-Ann Arbor-Flint-MI	7,664	15	0.20%
Fort Myers-Cape Coral-FL	4,826	98	2.03%
Greensboro-Winston-Salem-High Point-NC	5,513	10	0.18%
Greenville-Spartanburg-Anderson-SC	4,804	7	0.15%
Honolulu-HI	7,253	267	3.68%
Houston-Galveston-Brazoria-TX	23,130	99	0.43%
Jacksonville-FL	6,605	11	0.17%
Kansas City-MO-KS	11,059	33	0.30%
Los Angeles-Riverside-Orange County-CA	90,400	842	0.93%
Miami-Fort Lauderdale-FL	23,263	62	0.27%
Milwaukee-Racine-WI	4,824	14	0.29%
Minneapolis-St. Paul-MN-WI	34,675	248	0.72%
Nashville-TN	5,354	11	0.21%
New York-Northern New Jersey-Long Island-NY	26,834	385	1.43%
Norfolk-Virginia Beach-Newport News-VA-NC	8,112	108	1.33%
Oklahoma City-OK	4,037	10	0.25%
Orlando-FL	9,339	32	0.34%
Philadelphia-Wilmington-Atlantic City-PA-NJ	15,102	319	2.11%
Phoenix-Mesa-AZ	28,298	139	0.49%

Table 2.3: Loan Origination and Defaults by MSA (continued)
(n=734,721)

Portland-Salem-OR-WA	12,337	43	0.35%
Raleigh-Durham-Chapel Hill-NC	5,871	6	0.10%
Richmond-Petersburg-VA	7,707	58	0.75%
Sacramento-Yolo-CA	15,553	42	0.27%
San Antonio-TX	4,977	19	0.38%
San Diego-CA	18,930	28	0.15%
San Francisco-Oakland-San Jose-CA	60,165	79	0.13%
Sarasota-Bradenton-FL	5,439	16	0.29%
Seattle-Tacoma-Bremerton-WA	20,299	53	0.26%
St. Louis-MO-IL	10,927	28	0.26%
Tampa-St. Petersburg-Clearwater-FL	17,231	51	0.30%
Tucson-AZ	5,008	23	0.46%
Washington-Baltimore-DC-MD-VA-WV	51,241	275	0.54%
West Palm Beach-Boca Raton-FL	7,508	29	0.39%
Total	734,721	3,958	0.54%

Table 2.4: Loan Origination and Prepayments by MSA
(n=734,721)

MSA Name	# Originated	# Prepaid	% Prepaid
Albuquerque-NM	5,753	970	16.86%
Atlanta-GA	25,854	4,726	18.28%
Austin-San Marcos-TX	6,938	1,618	23.32%
Boston-Worcester-Lawrence-MA-NH-ME-CT	16,349	8,084	49.45%
Buffalo-Niagara Falls-NY	4,849	1,531	31.57%
Charleston-North Charleston-SC	4,051	679	16.76%
Charlotte-Gastonia-Rock Hill-NC-SC	14,263	2,260	15.85%
Chicago-Gary-Kenosha-IL-IN-WI	27,557	12,178	44.19%
Cincinnati-Hamilton-OH-KY-IN	5,514	2,809	50.94%
Cleveland-Akron-OH	5,478	1,927	35.18%
Columbia-SC	4,125	736	17.84%
Columbus-OH	4,213	1,763	41.85%
Dallas-Fort Worth-TX	23,591	5,179	21.95%
Denver-Boulder-Greeley-CO	21,901	14,129	64.51%
Detroit-Ann Arbor-Flint-MI	7,664	3,822	49.87%
Fort Myers-Cape Coral-FL	4,826	1,083	22.44%
Greensboro-Winston-Salem-High Point-NC	5,513	859	15.58%
Greenville-Spartanburg-Anderson-SC	4,804	866	18.03%
Honolulu-HI	7,253	1,860	25.64%
Houston-Galveston-Brazoria-TX	23,130	4,593	19.86%
Jacksonville-FL	6,605	1,103	16.70%
Kansas City-MO-KS	11,059	3,267	29.54%
Los Angeles-Riverside-Orange County-CA	90,400	30,387	33.61%
Miami-Fort Lauderdale-FL	23,263	3,644	15.66%
Milwaukee-Racine-WI	4,824	2,641	54.75%
Minneapolis-St. Paul-MN-WI	34,675	26,044	75.11%
Nashville-TN	5,354	1,190	22.23%
New York-Northern New Jersey-Long Island-NY	26,834	13,401	49.94%
Norfolk-Virginia Beach-Newport News-VA-NC	8,112	1,744	21.50%
Oklahoma City-OK	4,037	823	20.39%
Orlando-FL	9,339	1,832	19.62%
Philadelphia-Wilmington-Atlantic City-PA-NJ	15,102	6,251	41.39%
Phoenix-Mesa-AZ	28,298	11,055	39.07%

Table 2.4: Loan Origination and Prepayments by MSA (continued)
(n=734,721)

Portland-Salem-OR-WA	12,337	6,462	52.38%
Raleigh-Durham-Chapel Hill-NC	5,871	1,564	26.64%
Richmond-Petersburg-VA	7,707	1,782	23.12%
Sacramento-Yolo-CA	15,553	4,623	29.72%
San Antonio-TX	4,977	937	18.83%
San Diego-CA	18,930	7,438	39.29%
San Francisco-Oakland-San Jose-CA	60,165	25,381	42.19%
Sarasota-Bradenton-FL	5,439	825	15.17%
Seattle-Tacoma-Bremerton-WA	20,299	10,294	50.71%
St. Louis-MO-IL	10,927	3,867	35.39%
Tampa-St. Petersburg-Clearwater-FL	17,231	2,519	14.62%
Tucson-AZ	5,008	1,710	34.15%
Washington-Baltimore-DC-MD-VA-WV	51,241	10,811	21.10%
West Palm Beach-Boca Raton-FL	7,508	1,380	18.38%
Total	734,721	254,647	34.66%

Table 2.5: Loan Origination and Defaults by MSA, after Year 1998
(n=486,419)

MSA Name	# Originated	# Defaulted	% Defaulted
Albuquerque-NM	4,493	5	0.11%
Atlanta-GA	20,297	19	0.09%
Austin-San Marcos-TX	5,226	5	0.10%
Boston-Worcester-Lawrence-MA-NH-ME-CT	10,154	2	0.02%
Buffalo-Niagara Falls-NY	706	1	0.14%
Charleston-North Charleston-SC	3,380	2	0.06%
Charlotte-Gastonia-Rock Hill-NC-SC	11,591	8	0.07%
Chicago-Gary-Kenosha-IL-IN-WI	17,648	19	0.11%
Cincinnati-Hamilton-OH-KY-IN	2,776	4	0.14%
Cleveland-Akron-OH	2,170	4	0.18%
Columbia-SC	3,250	1	0.03%
Columbus-OH	2,219	2	0.09%
Dallas-Fort Worth-TX	16,704	19	0.11%
Denver-Boulder-Greeley-CO	10,857	3	0.03%
Detroit-Ann Arbor-Flint-MI	3,458	2	0.06%
Fort Myers-Cape Coral-FL	3,360	0	0.00%
Greensboro-Winston-Salem-High Point-NC	4,575	1	0.02%
Greenville-Spartanburg-Anderson-SC	3,897	1	0.03%
Honolulu-HI	3,899	42	1.08%
Houston-Galveston-Brazoria-TX	15,508	32	0.21%
Jacksonville-FL	5,456	6	0.11%
Kansas City-MO-KS	8,069	5	0.06%
Los Angeles-Riverside-Orange County-CA	66,719	76	0.11%
Miami-Fort Lauderdale-FL	17,727	14	0.08%
Milwaukee-Racine-WI	2,896	0	0.00%
Minneapolis-St. Paul-MN-WI	9,229	7	0.08%
Nashville-TN	4,033	8	0.20%
New York-Northern New Jersey-Long Island-NY	10,046	3	0.03%
Norfolk-Virginia Beach-Newport News-VA-NC	4,889	8	0.16%
Oklahoma City-OK	2,865	5	0.17%
Orlando-FL	6,880	6	0.09%
Philadelphia-Wilmington-Atlantic City-PA-NJ	4,796	2	0.04%
Phoenix-Mesa-AZ	19,136	42	0.22%

Table 2.5: Loan Origination and Defaults by MSA, after Year 1998 (continued)
(n=486,419)

Portland-Salem-OR-WA	6,827	17	0.25%
Raleigh-Durham-Chapel Hill-NC	4,320	1	0.02%
Richmond-Petersburg-VA	4,343	3	0.07%
Sacramento-Yolo-CA	12,599	3	0.02%
San Antonio-TX	3,309	8	0.24%
San Diego-CA	14,985	1	0.01%
San Francisco-Oakland-San Jose-CA	44,791	5	0.01%
Sarasota-Bradenton-FL	4,558	1	0.02%
Seattle-Tacoma-Bremerton-WA	11,397	13	0.11%
St. Louis-MO-IL	8,642	2	0.02%
Tampa-St. Petersburg-Clearwater-FL	14,435	5	0.03%
Tucson-AZ	3,180	7	0.22%
Washington-Baltimore-DC-MD-VA-WV	38,138	30	0.08%
West Palm Beach-Boca Raton-FL	5,986	3	0.05%
Total	486,419	453	0.09%

Table 2.6: Loan Origination and Prepayments by MSA, after Year 1998
(n=486,419)

MSA Name	# Originated	# Prepaid	% Prepaid
Albuquerque-NM	4,493	282	6.28%
Atlanta-GA	20,297	2,159	10.64%
Austin-San Marcos-TX	5,226	985	18.85%
Boston-Worcester-Lawrence-MA-NH-ME-CT	10,154	3,212	31.63%
Buffalo-Niagara Falls-NY	706	15	2.12%
Charleston-North Charleston-SC	3,380	469	13.88%
Charlotte-Gastonia-Rock Hill-NC-SC	11,591	1,132	9.77%
Chicago-Gary-Kenosha-IL-IN-WI	17,648	5,164	29.26%
Cincinnati-Hamilton-OH-KY-IN	2,776	1,023	36.85%
Cleveland-Akron-OH	2,170	666	30.69%
Columbia-SC	3,250	457	14.06%
Columbus-OH	2,219	778	35.06%
Dallas-Fort Worth-TX	16,704	2,263	13.55%
Denver-Boulder-Greeley-CO	10,857	4,851	44.68%
Detroit-Ann Arbor-Flint-MI	3,458	1,078	31.17%
Fort Myers-Cape Coral-FL	3,360	360	10.71%
Greensboro-Winston-Salem-High Point-NC	4,575	559	12.22%
Greenville-Spartanburg-Anderson-SC	3,897	560	14.37%
Honolulu-HI	3,899	465	11.93%
Houston-Galveston-Brazoria-TX	15,508	1,338	8.63%
Jacksonville-FL	5,456	510	9.35%
Kansas City-MO-KS	8,069	1,257	15.58%
Los Angeles-Riverside-Orange County-CA	66,719	12,906	19.34%
Miami-Fort Lauderdale-FL	17,727	1,635	9.22%
Milwaukee-Racine-WI	2,896	1,034	35.70%
Minneapolis-St. Paul-MN-WI	9,229	4,025	43.61%
Nashville-TN	4,033	691	17.13%
New York-Northern New Jersey-Long Island-NY	10,046	2,391	23.80%
Norfolk-Virginia Beach-Newport News-VA-NC	4,889	442	9.04%
Oklahoma City-OK	2,865	362	12.64%
Orlando-FL	6,880	632	9.19%
Philadelphia-Wilmington-Atlantic City-PA-NJ	4,796	1,109	23.12%
Phoenix-Mesa-AZ	19,136	4,160	21.74%

Table 2.6: Loan Origination and Defaults by MSA, after Year 1998 (continued)
(n=486,419)

Portland-Salem-OR-WA	6,827	2,253	33.00%
Raleigh-Durham-Chapel Hill-NC	4,320	628	14.54%
Richmond-Petersburg-VA	4,343	488	11.24%
Sacramento-Yolo-CA	12,599	2,481	19.69%
San Antonio-TX	3,309	366	11.06%
San Diego-CA	14,985	4,394	29.32%
San Francisco-Oakland-San Jose-CA	44,791	12,875	28.74%
Sarasota-Bradenton-FL	4,558	404	8.86%
Seattle-Tacoma-Bremerton-WA	11,397	3,756	32.96%
St. Louis-MO-IL	8,642	2,530	29.28%
Tampa-St. Petersburg-Clearwater-FL	14,435	1,347	9.33%
Tucson-AZ	3,180	441	13.87%
Washington-Baltimore-DC-MD-VA-WV	38,138	4,342	11.38%
West Palm Beach-Boca Raton-FL	5,986	611	10.21%
Total	486,419	95,886	19.71%

Table 2.7: Standard Cox Model for the Risk of Default

	Full Sample	**Year>=98**	**Year>=98**	**Year>=98**
Covariates	n=734,721	n=486,419	n=486,419	n=486,419
LTV ratio	0.0586	0.0767	0.03986	0.0325
	(40.72)**	(12.9)**	(4.49)**	(3.56)**
Log Loan Size	-0.3835	-0.0172	0.42538	0.9786
	(-11.01)**	(-0.15)	(2.31)*	(4.89)**
Points	-0.3142	-0.2383	-0.42321	-0.3933
	(-10.38)**	(-5.15)**	(-5.63)**	(-5.18)**
Contract Rate	0.1795	0.9991	0.99428	1.0033
	(13.63)**	(15.92)**	(8.95)**	(9.07)**
Spread	0.0099	0.0310	0.01631	0.0132
	(5.73)**	(3.61)**	-1.10000	-0.8900
State Unemployment Rate	-0.0865	-1.0697	-1.13914	-1.2490
	(-6.34)**	(-17.12)**	(-10.85)**	(-11.23)**
State House Price Ratio	-8.6148	-15.6195	-15.74563	-16.4153
	(-75.68)**	(-19.89)**	(-12.68)**	(-13.57)**
FICO Score			-0.01222	-0.0126
			(-9.79)**	(-9.99)**
State-Level Annual GDP ($1000)				-0.0259
				(-1.11)
Hispanic percentage per ZIP code				1.1387
				(2.88)**
African-American percentage per ZIP code				-1.6724
				(-2.89)**
Median Household Income per ZIP code				-1.3783
				(-4.29)**

Table 2.8: Standard Cox Model for the Risk of Prepayment

	Full Sample	**Year>=98**	**Year>98**	**Year>98**
Covariates	n=734,721	n=486,419	n=486,419	n=486,419
LTV ratio	-0.01488	-0.0346	-0.03562	-0.0333
	(-113.15)**	(-159.6)**	(-119.27)**	(-102.67)**
Log Loan Size	0.56956	1.0381	1.09730	1.0156
	(144.64)**	(160.75)**	(126.6)**	(-106.89)**
Points	-0.35015	-0.2274	-0.19854	-0.1589
	(-97.96)**	(-60.98)**	(-46.38)**	(-36.67)**
Contract Rate	0.20607	1.3211	1.56353	1.5375
	(121.26)**	(335.75)**	(268.52)**	(262.24)**
Spread	0.01209	-0.0395	-0.04114	-0.0403
	(50.25)**	(-65.33)**	(-44.11)**	(-43.13)**
State Unemployment Rate	-0.16041	-0.3721	-0.24605	-0.2750
	(-89.45)**	(-99.55)**	(-48.64)**	(-53.16)**
State House Price Ratio	-3.32526	-10.6853	-8.88040	-9.4138
	(-363.78)**	(-247.59)**	(-173.03)**	(-177.4)**
FICO Score			0.00273	0.0026
			(30.67)**	(29.37)**
State-Level Annual GDP ($1000)				0.0265
				(34.97)**
Hispanic percentage per ZIP code				1.2774
				(44.76)**
African-American percentage per ZIP code				-1.7611
				(-36.02)**
Median Household Income per ZIP code				0.1805
				(10.03)**

Table 2.9: Shared Frailty Model

Covariates	**Default**	**Prepayment**
	n=734,721	n=734,721
LTV ratio	0.0781	-0.0041
	(47.89)**	(-26.01)**
Log Loan Size	-0.8725	0.3625
	(-22.16)**	(81.23)**
Points	-0.4060	-0.3713
	(-12.66)**	(-109.9)**
Contract Rate	0.0387	0.0927
	(2.48)**	(47.27)**
Spread	0.0116	0.0125
	(6.24)**	(51.36)**
State Unemployment Rate	0.1083	-0.2279
	(5.11)**	(-83.89)**
State House Price Index	-7.9245	-4.0380
	(-65.75)**	(-397.99)**
Theta	0.6426	0.4363
	(5.03)**	(5.17)**
Likelihood Ratio Test - Chi2(1) value	1642	8600

Table 2.10: Expected Frailty Values by MSA

	Expected Frailties	
West	**Default**	**Prepayment**
Albuquerque-NM	0.2695	0.3048
Denver-Boulder-Greeley-CO	0.1860	0.9574
Honolulu-HI	11.1816	0.5002
Los Angeles-Riverside-Orange County-CA	2.4958	1.5506
Phoenix-Mesa-AZ	0.4382	0.6296
Portland-Salem-OR-WA	0.8913	1.8745
Sacramento-Yolo-CA	1.2504	2.3530
San Antonio-TX	0.6820	2.2834
San Francisco-Oakland-San Jose-CA	0.4209	1.7614
Seattle-Tacoma-Bremerton-WA	0.3747	1.8061
Tucson-AZ	0.4053	0.5895
Midwest		
Chicago-Gary-Kenosha-IL-IN-WI	0.8076	1.0227
Cincinnati-Hamilton-OH-KY-IN	0.2305	0.6386
Cleveland-Akron-OH	0.3776	0.4866
Columbus-OH	0.1977	0.5164
Detroit-Ann Arbor-Flint-MI	0.2140	1.0032
Kansas City-MO-KS	0.2072	0.3916
Milwaukee-Racine-WI	0.5222	0.9485
Minneapolis-St. Paul-MN-WI	0.4088	0.6630
St. Louis-MO-IL	0.4168	0.7712

Table 2.10: Expected Frailty Values by MSA (continued)

South		
Atlanta-GA	0.3629	0.5074
Austin-San Marcos-TX	0.0362	0.1478
Charleston-North Charleston-SC	0.2501	0.4599
Charlotte-Gastonia-Rock Hill-NC-SC	0.2351	0.4368
Cincinnati-Hamilton-OH-KY-IN	0.2305	0.6386
Columbia-SC	0.1567	0.4895
Dalla-Fort Worth-TX	0.1161	0.1538
Fort Myers-Cape Coral-FL	1.9918	0.6366
Greensboro-Winston-Salem-High Point-NC	0.3219	0.5220
Greenville-Spartanburg-Anderson-SC	0.2260	0.4919
Houston-Galveston-Brazoria-TX	0.0839	0.1454
Jacksonville-FL	0.2731	0.6615
Miami-Fort Lauderdale-FL	0.3954	0.5789
Nashville-TN	0.2191	0.4349
Norfolk-Virginia Beach-Newport News-VA-NC	0.6061	0.4241
Oklahoma City-OK	0.0437	0.0931
Orlando-FL	0.4750	0.5866
Philadelphia-Wilmington-Atlantic City-PA-NJ	3.0130	1.1795
Raleigh-Durham-Chapel Hill-NC	0.1433	0.5665
Richmond-Petersburg-VA	0.4638	0.4962
San Antonio-TX	0.0803	0.1226
Sarasota-Bradenton-FL	0.7120	0.6354
Tampa-St. Petersburg-Clearwater-FL	0.5400	0.6502
Washington-Baltimore-DC-MD-VA-WV	0.6828	0.6129
West Palm Beach-Boca Raton-FL	0.8042	0.6125
New England		
Boston-Worcester-Lawrence-MA-NH-ME-CT	4.0732	7.7212
Buffalo-Niagara Falls-NY	3.0106	4.0869
New York-Northern New Jersey-Long Island-NY	5.7055	2.4935
Philadelphia-Wilmington-Atlantic City-PA-NJ	3.0130	1.1795

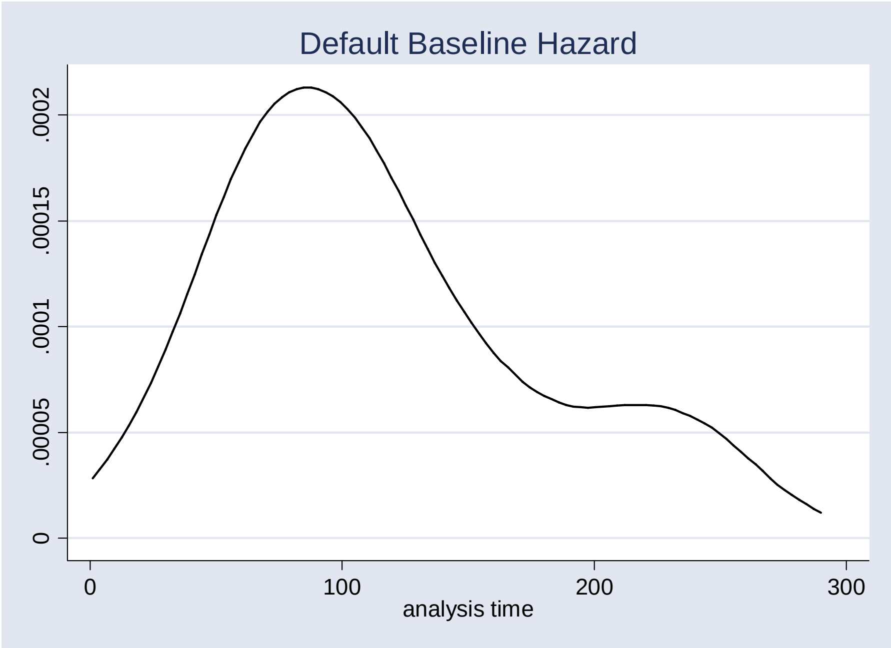


Figure 2.1: Default baseline hazard curve based on the full sample

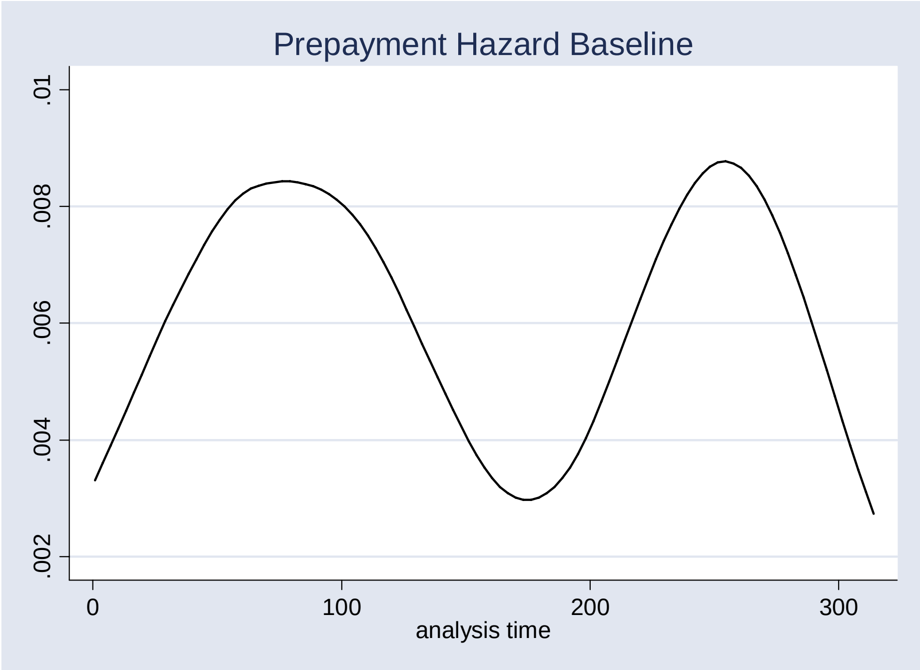
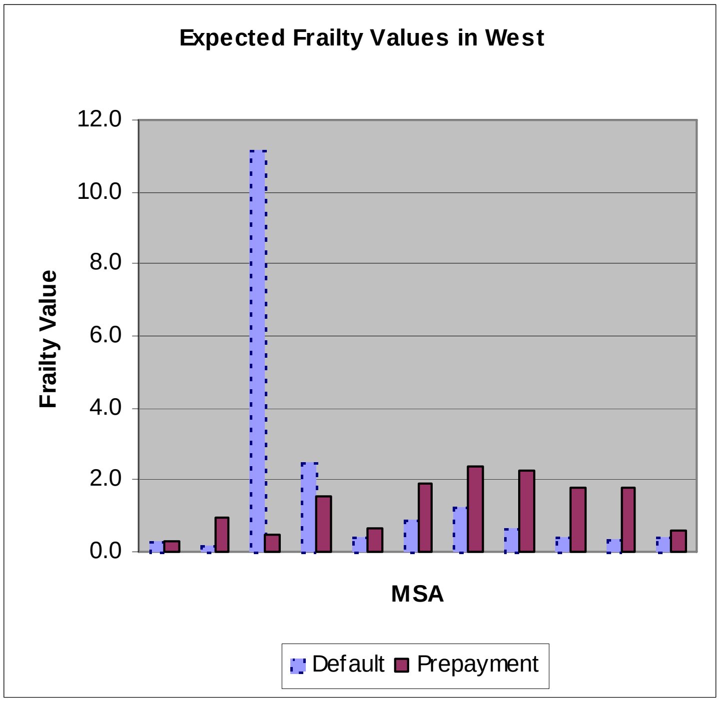


Figure 2.2: Prepayment baseline hazard curve based on the full sample



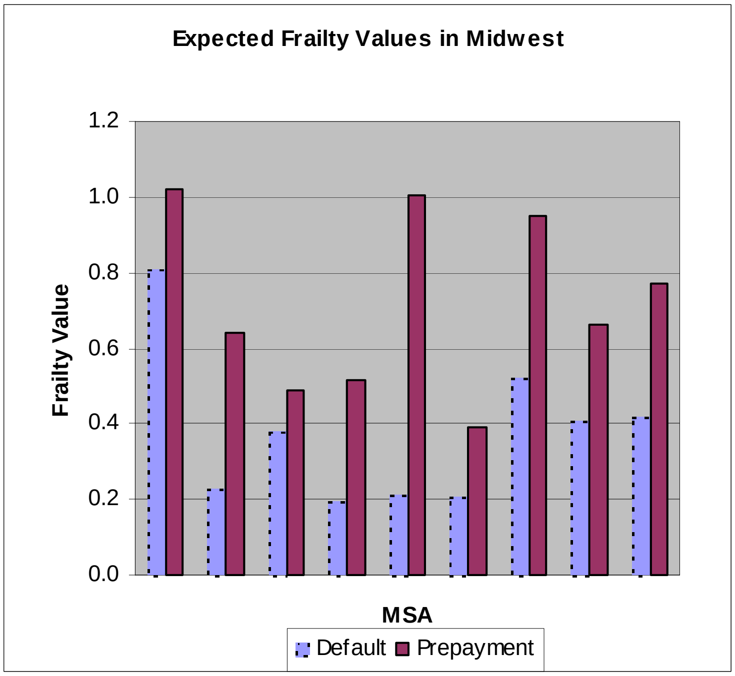
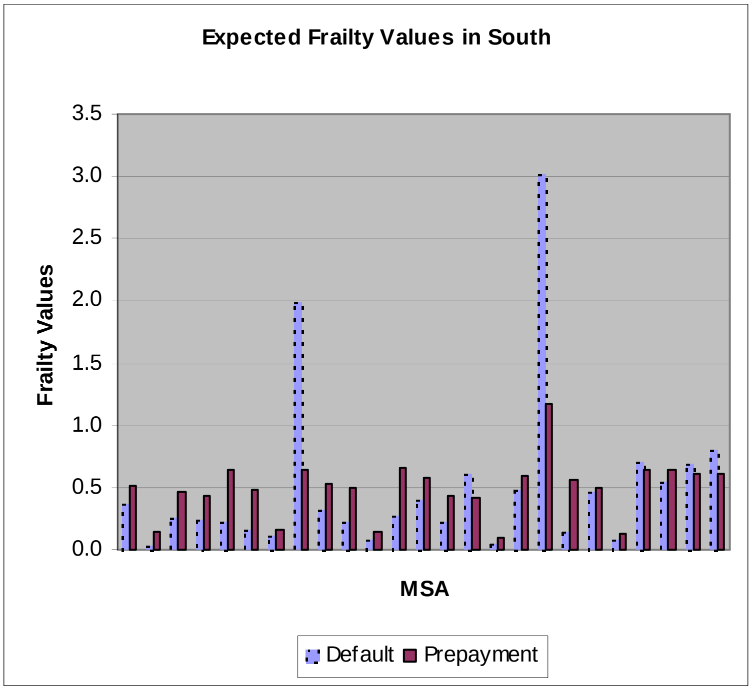


Figure 2.3: Expected frailty values for each MSA



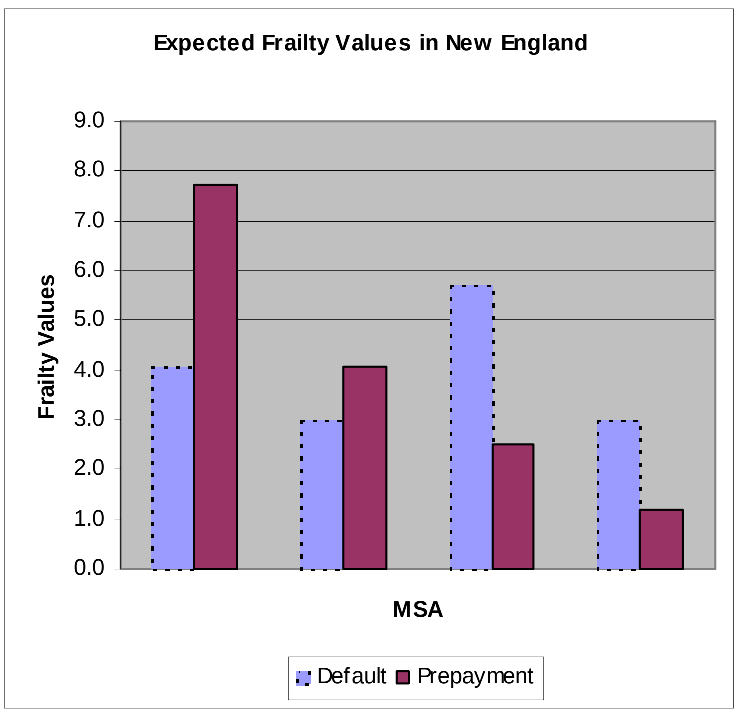


Figure 2.3: Expected frailty values for each MSA (continued)

CHAPTER 3

MORTGAGE TERMINATION RISK AT THE ORIGINATED YEAR LEVEL: A SHARED-FRAILTY APPROACH

3.1. Introduction

Homeownership has often been viewed as the “American Dream”. Both government agencies and private lenders have played an important role in helping families achieve their ideal by providing necessary support of mortgage financing. However, some mortgages are terminated via either default or prepayment. Failure to repay mortgage financing (default) varies substantially with the type of loans and borrowers in different periods across regions. In addition, mortgage default is costly to the lenders and to the federal institutions that guarantee and insure home mortgages. Costs to lenders and institutions are incurred when the net cash regained from the foreclosure proceedings is less than the value of the financial asset. A desire to minimize default costs to both borrowers and lenders are central to default studies. In the meanwhile, the decision to terminate the mortgage agreement can be due to either financial reasons (replacing the loan at a lower rate) or exogenous reasons (divorce or job transfers). Thus, borrowers’ decision of prepayments can be affected by interest rate movements, refinancing costs, and demographic/macroeconomic factors

Mortgage default studies have focused on different aspects of the default decision. Quercia and Stegman (1992) reviewed the residential mortgage default literature and divided the studies into three groups: a study can be categorized either by research perspective (from lender’s perspective, or borrower’s, or institutional), or by the measures used to determine the mortgage risk (e.g. default rates and expected credit loss), or by the primary research approaches (theoretical or empirical).

The mortgage prepayment literature is mostly concerned with individual mortgages and mortgage pools, and has focused on two areas: (1) determining factors that affect mortgage prepayments and modeling prepayment decisions (e.g. Green and Shoven, 1986; Richard and Roll, 1989; Mattey and Wallace, 1998), and (2) the effects of mortgage prepayments on the valuation of mortgage-backed securities (MBSs) (e.g. Dunn and McConnell, 1981; Schwartz and Torous, 1989; Spahr and Sunderman, 1992; and Jegadeesh and Ju, 2000). Results from these studies indicate that mortgage prepayments are positively related to the spread between the contract interest rate and prevailing mortgage rate, loan size, mortgage loan to value (LTV) ratio, and appreciation of house prices.

Few studies have done the temporal analysis of mortgage defaults or prepayments while this issue may have important implications for mortgage pricing. For example, knowledge of whether the default or prepayment rates are similar in adjacent origination years conveys information for risk management of the lending institutions. This essay contributes to the existing literature by conducting an empirical analysis to examine whether the mortgage defaults/prepayments are correlated in terms of sharing common temporal unobserved factors. A related analysis in determining how corporate default events are correlated has been conducted. The remainder of this essay is organized as follows. Section 3 describes the data and methodology. Section 4 presents the empirical results and section 5 concludes.

3.2. Default Correlation

Recent literature from corporate debt has shown the importance of the research on default correlation. Default correlation is a measure of the dependence among risks. Default correlation is defined by Lucas (2004) “as the phenomenon that the likelihood of one obligator defaulting on its debt is affected by whether or not another obligator has default on its debts”. Default correlation is important since default is a rare event and thus, the correlation among defaults might have a large impact on the valuation of the credit portfolios. The concept of default correlation incorporates both economy-wide and industry specific or regional factors. Default risk caused by the macroeconomic variables is like systematic risk which makes the default event to cluster. Coincident default events may be triggered by common underlying factors. Along with default rates and loss severity, default correlation is necessary in estimating the value of portfolio at risk due to credit. Failure to recognize the impact of shocks to the portfolio through default correlation will ultimately underestimate the measures of risks and economic capital required to manage that risk. If mortgage defaults are more heavily clustered in time than envisioned, such as where there is the assumption for independent defaults, then significantly greater capital might be required in order to survive default losses.

The corporate debt literature focuses on default correlation over time among commercial portfolios. The linkage between the initial credit quality of the portfolio and the default correlation of commercial portfolios has been specified. Generally, as credit quality increases, the importance of default correlation decreases. For instance, Zhou (1997) shows the default correlation is almost zero for highly rated firms over short to middle investment horizons but pretty high for low rated firms even for short investment horizons. Using corporate bond and loan portfolios, Lucas (2004) provides numerical examples that default correlations increase as ratings decrease and that default correlations initially increase with time and then decrease with time. However, some recent studies have presented different results. Das et al. (2001) provide empirical evidence that default intensities of high grade firms are more highly correlated than firms of low rating classes within a 3-year time horizon.

Compared with commercial loan and bond portfolios, mortgage loans have similarities and differences. The similarity between corporate debts and mortgage loans are that the process yielding the risk of mortgage default is analogous to the other commercial loans. The differences are reflected in two aspects: (1) the number of loans; mortgage defaults are far more frequently observed than firms going into default; and (2) the homogenous property of mortgage contracts is limited in reflecting the associated risk individually. Thus, it makes hard to apply the approach used in the corporate debt literature to the analysis of the mortgage default. There are few studies regarding the presence of temporally associated mortgage defaults. Cowans (2004) estimated default correlations using a large portfolio of residential sub-prime loans. In their essay, a default correlation coefficient is estimated based on the assumption that all loans within the same risk class have identical default rates. Cowans' approach of calculating default correlation starts from the definition of correlation and can be viewed as a direct measure for the correlation. However, their approach neglects the role played by the observed and unobserved information in determining the correlation. In this essay, I use the survival model with frailty to investigate the potential correlation of mortgage defaults for mortgages originated in a given year. The model that I use includes both observed and unobserved covariates. The approach I follow does not begin with the definition of correlation and thus it can be viewed as an "indirect" measure of the default correlation. To investigate how the initial risk indicators are connected with the default hazard, I stratify the mortgage sample by the borrowers' initial LTV ratio. It is argued that the higher the LTV ratio, the higher the debt service requirements, and hence the higher the probability the borrower will ultimately encounter financial distress (Von Furstenberg, 1969;

Brueckner, 2000; Deng et al, 2000). In addition, I make a further effort of examining the temporal prepayment clustering as prepayment reflects another important source of mortgage hazards from lender's perspective.

3.3. Data and Method

The data for this essay are the detailed loan history file from a large financial lender. The data set consists of monthly data on 30-year fixed rate residential loans originated from 1976 through 2004. Overall, the sample period represents a very mixed economic picture. In general, the economy was enjoying expansion and experiencing economic weaknesses as well. The sample period includes multiple periods of primarily rising or stable house prices in most geographic markets. According to the Federal Home Loan Mortgage Corporation(Freddie Mac's) monthly mortgage rate survey, the sample period includes both periods of declining and rising 30-year mortgage rates, with a peak of 18.45% during the period in October of 1981 and a low of 5.23% in June of 2003. Thus, the time period examined in this essay may reduce the selection bias, especially when compared with previous studies (Cowans, 2004) with limited sample periods.

For each mortgage, the available information includes the year and month of origination and termination (if it has been closed), indicators of termination --- defaulted, prepaid, sold, or censored, and a number of loan-specific characteristics observed at the time of origination. The characteristics available are the original loan amount, the original loan-to-value ratio, the contract rate at origination, the amount of points paid at origination, and the zip code where the property is located. Two additional variables that reflect the state of the general economy, the monthly state-level unemployment rate and the state-level Housing Price Index (HPI) are integrated.

As noted by Jarrow and Turnbull (2000), credit risk time horizons are commonly one year. Thus, I consider examining all the loans on an annual basis. The use of an annual basis provides the benefit of observing a reasonable number of the time periods while at the same time reflecting the general economy. The mortgages in our sample are thus distributed in 29 clusters. The hazard function is a natural approach to model mortgage hazard --- it provides us with information on termination risks of each mortgage known to be alive at the age t and it can be easily adapted to more complicated situations, such as where there is censoring or there is frailty in the model.

Let the random variable T represent the time till the mortgage is terminated via either default or prepayment. The hazard function is defined by

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \le T < t + \Delta t \mid T \ge t)}{\Delta t} \tag{1}$$

The probability of survival at time t can be given in terms of the hazard function:

$$S(t) = P(T \ge t) = \exp(-\int_0^t h(u)du \tag{2}$$

The survival model with heterogeneous frailty was introduced by Clayton (1978) and by Vaupel et al. (1979). Survival analysis focuses on the hazard, which is the instantaneous risk of event at duration t, conditional on survival up to t. In the model, the hazard at time t conditional on observed covariates X and unobserved frailty v is assumed to be the product of frailty and a baseline hazard. Typically, the data are organized as $i = 1,...,n$ groups with $j = 1,...,n_i$ observations in group i. For the jth observation in the ith group, the hazard function is:

$$h_{ij}(t) = h_0(t) v_i \exp(x_{ij}\beta) \tag{3}$$

where v_i is the group-level frailty. Frailties, or the random effects, are not observation specific, but instead are shared across groups of observations, causing those observations within the same group to be correlated. Also, we can use stratified Cox proportional hazards model to control for the within-group correlation. In stratified Cox analysis, the baseline hazards are allowed to differ by groups, but the coefficients are assumed to be the same across different groups (Therneau and Grambsch, 2000). The reason for the use of the Cox model with shared frailty is that we can test within-group correlations in the case of the shared-frailty models. The unobserved frailty effect is represented as a constant relative risk that takes only non-negative values. It is assumed to follow some distribution, $f(v)$, at duration $t = 0$ and, for mathematical convenience of estimable parameters, it is usually assumed to be gamma distribution with a mean equal to one and a variance θ. Covariates are included in the model by specifying the baseline hazard function and the relative risk associated with a set of covariates. Frailty is assumed to be independent with any covariates.

Frailty in this model represents an individual mortgage's susceptibility to the default or prepayment risk. It captures the total effect of all factors that influence the individual's termination risk that are not included in the baseline hazard function or the information set conveyed by the observed covariates. The specific factors that comprise the frailty effect clearly depend on the specific application and on the completeness of the set of observed covariates. In the context of mortgage hazards, the frailty effect reflects a multitude of factors that can be broadly classified as spatial and temporal. In this essay, the frailty v_i is used to model the temporal effect and the origination year is defined as the group variable. The assumption is that mortgages originated in the same year are correlated because they share the same frailty, or common unobserved effects. The estimate of variance θ is used to measure the degree of within-group correlation, and a larger variance implies greater correlation in default or prepayment risk among mortgages in the same group, in addition to implying greater heterogeneity across groups.

3.4. Empirical Results

3.4.1. Default and prepayment rates for subsets of borrowers

First, I document the overall and subsample default and prepayment rates for the 30-year fixed-rate loans. The first and most obvious variable to consider is the risk grade evaluated at the origination. I use the LTV ratio as the comparable rating measure to assess the credit risk of borrowers. This measure would be comparable to the bond rating discussed in the corporate literature. However, in the case of bonds, the bond rating are assigned and tracked over time. For mortgage loans, we only observed the LTV ratio at the time of origination. While one cannot measure how the risk grade might change over time for the borrower, the fact that the borrower started in a particular risk grade is still very indicative of the likelihood of termination risks. Table 1 through table 4 display the default and prepayment rates for the entire sample and for each subgroup.

As can be seen in table 1, the overall default rate is 0.52%. Mortgages with original LTV ratio between 70% and 80% has accounted for 34.2% of the total sample while mortgages with original LTV ratio between 80% and 90% only account for 7.4% of the total sample. LTV ratio appears to accurately capture credit risk. The likelihood of default monotonically increases as the LTV ratio increases: default rates of mortgages with LTV ratio greater than 90% is almost 10 times of the mortgages with LTV lower than 70%. Figure 1 demonstrates the non-parametric smoothed Kaplan-Meier (Kaplan and Meier, 1958) estimates of the default hazard in each month of mortgage life for each LTV category. The shape of the hazard estimates is similar but the scale is different. The shape in figure 1 displays that the default hazard seems to have a tendency to increase monotonically over the first 10 years of mortgage life, and decreases afterwards. The scale of figure 1 presents that the default hazard is consistent with their risk measures at origination, as described above.

Table 2 displays the trends of prepayment rates for the borrowers classified by different initial LTV ratio. The overall prepayment rate for the full sample is about 1/3. Segmented by the initial LTV ratio, borrowers with 80-90% LTV ratio has prepayment rates which are about 10 percentage points higher than the overall prepayment rates. This changing pattern is consistent with the empirical finding by LaCour-Little (1999). Based on a 5 quarter period of data, his results present that LTV ratio is positively associated with the probability of prepayment probability.

3.4.2. Frailty and non-frailty models

The non-frailty model, or the standard Cox model estimating the cause-specific hazard, is a special case of the frailty model when the variance of the frailty is equal to 0. Table 3 and Table 4 present estimates of the frailty and non-frailty models, using the origination year as the group variable, for each LTV category. The likelihood ratio test for each regression shows that the frailty model has improved the overall fit of the model. Most of the coefficients are significant, and coefficients exhibit a similar pattern. Among the statistically significant coefficients, the absolute magnitude for the estimates of *LTV ratio, points, original coupon rate,* and *the state-level housing price index* tend to be smaller under the standard Cox model (non-frailty model). However, the magnitude of the estimate of the log form of *original loan amount* is smaller under the frailty model. In addition, the sign of the *state-level unemployment rate* changes from negative in the non-frailty model to positive in the frailty model. These changes of the estimated coefficients imply that the introduction of the frailty at the originated year level has changed the weights of risk factors on the termination hazard.

The variance of the group-level frailty in the default regressions, which measures the magnitude of frailty that affects the default risk during the observation time periods, displays a mixed pattern for the subsets of borrowers. Whereas initial LTV ratio serve as initial credit quality indicators for the borrowers, the correlation value measured as the variance of the frailty does not follow the trend predicted by the corporate debt literature. The relationship between the credit quality indicators and the magnitude of the heterogeneity of default hazard is mixed. After controlling for the observed covariates, for mortgages classified by different initial LTV ratio, default correlation is highest for mortgages with initial LTV ratio between 70% and 80%, while mortgages with initial LTV ratio between 80% and 90% presents smallest correlation. In Table 4, the variance of the frailty is low for low LTV loans; the variance of the frailty is high for high LTV loans or loans. This shows that the additional risk factor due to frailty follows the change of credit qualities.

Another question of interest is to know whether the estimated frailty values in nearby origination years are similar or follow a correlated pattern. Table 5-6 and figure 3-4 present the expected frailty values for default and prepayment risks. It is interesting to note that most of the expected frailty values from the nearby yearly cohorts are pretty close. This suggests that there exists correlation of the survival times for mortgages originated in nearby years. Furthermore, Table 5 and Table 6 show that mortgages originated between 1999 and 2001 are observed to be subject to higher unobserved temporal risk factors than those originated in other years in our sample. This implies that mortgages originated in these years are subject to termination risks due to the unobserved temporal factors.

3.5. Conclusion

To summarize, this essay applies the shared frailty model, which has its roots in biomedical research, to examine potential temporal correlation between survival times of mortgages originated in the same year. Compared with the method studying the “correlation” issue in the corporate debt literature, this approach offers an alternative to investigate the mortgage termination issue. The temporal frailty effect represents the association among individual mortgages originated in the same year that is a consequence of their shared macroeconomic environment. A large variance of frailty implies greater correlation in default or prepayment risks among mortgages originated in the same year, in addition to implying greater heterogeneity across years. This approach is convenient for statistical analysis of large-sample mortgage data. The shared-frailty model applied in this essay might not thoroughly cover correlations in the real world. However, it provides an attempt to deal with a problem known to cause biased and inefficient estimates. Classified by the initial credit indicators, the regression results show a mixed pattern for default risk but a monotonic trend of the prepayment risk. Finally, the shared frailty model constitutes a significant improvement over the simpler traditional Cox model, which itself conveys substantive information about the mortgage pricing process.

Table 3.1: Default Rates by LTV Ratio and Year of Origination

Year		LTV ratio at the time of loan origination			
	Entire Sample	LTV>=90	80<=LTV<90	70<=LTV<80	LTV<70
	n=1,038,098	n=288,448	n=76,663	n=355,116	n=317,871
1976	0.18%	0.20%	0.27%	0.09%	0.22%
1977	0.42%	0.80%	0.47%	0.21%	0.16%
1978	0.50%	0.77%	0.40%	0.45%	0.25%
1979	1.18%	2.09%	1.05%	0.62%	0.49%
1980	2.16%	2.61%	2.82%	1.62%	1.13%
1981	2.26%	3.53%	2.99%	0.45%	0.55%
1982	2.87%	3.52%	4.82%	1.28%	1.14%
1983	3.35%	4.52%	2.62%	1.34%	1.11%
1984	4.04%	5.00%	2.40%	2.32%	2.53%
1985	5.07%	6.52%	3.17%	2.44%	2.90%
1986	3.59%	5.95%	2.76%	1.53%	0.84%
1987	1.91%	3.00%	1.48%	1.05%	0.70%
1988	2.74%	4.83%	1.36%	1.72%	0.74%
1989	5.03%	6.53%	2.91%	6.28%	1.95%
1990	5.59%	6.64%	4.66%	7.23%	2.82%
1991	3.47%	5.50%	4.36%	3.57%	1.45%
1992	1.71%	3.34%	2.50%	1.14%	0.49%
1993	0.75%	1.33%	1.33%	0.52%	0.19%
1994	1.21%	2.16%	1.25%	0.68%	0.39%
1995	1.10%	1.98%	1.08%	0.51%	0.28%
1996	0.65%	0.94%	0.94%	0.56%	0.21%
1997	0.27%	0.47%	0.43%	0.17%	0.06%
1998	0.17%	0.40%	0.31%	0.07%	0.02%
1999	0.55%	1.20%	0.40%	0.13%	0.05%
2000	0.12%	0.30%	0.06%	0.02%	0.01%
2001	0.01%	0.02%	0.00%	0.00%	0.00%
2002	0.00%	0.00%	0.00%	0.00%	0.00%
2003	0.00%	0.00%	0.00%	0.00%	0.00%
2004	0.00%	0.00%	0.00%	0.00%	0.00%
Total	0.52%	1.15%	0.78%	0.30%	0.12%

Table 3.2: Prepayment Rates by LTV Ratio and Year of Origination

Year		LTV ratio at the time of loan origination			
	Entire Sample	LTV>=90	80<=LTV<90	70<=LTV<80	LTV<70
	n=1,038,098	n=288,448	n=76,663	n=355,116	n=317,871
1976	72.91%	70.95%	70.30%	75.76%	74.15%
1977	71.09%	65.75%	68.53%	74.34%	72.78%
1978	68.62%	59.68%	63.41%	72.61%	73.42%
1979	68.19%	61.91%	67.72%	71.52%	72.87%
1980	65.70%	59.90%	63.56%	70.59%	75.28%
1981	66.70%	64.66%	57.46%	66.03%	74.18%
1982	62.07%	57.99%	66.27%	70.05%	68.18%
1983	62.38%	58.60%	62.45%	71.04%	70.00%
1984	64.75%	62.17%	62.02%	73.35%	75.32%
1985	62.01%	57.61%	62.30%	75.83%	68.60%
1986	64.56%	53.68%	65.44%	76.10%	76.46%
1987	68.46%	61.13%	68.80%	63.05%	75.78%
1988	55.48%	45.63%	60.63%	60.08%	61.96%
1989	55.78%	45.01%	51.74%	63.87%	62.76%
1990	59.83%	47.44%	58.33%	74.61%	69.42%
1991	68.30%	53.69%	64.40%	78.16%	75.32%
1992	71.65%	59.11%	74.35%	64.99%	76.13%
1993	58.21%	47.27%	66.79%	70.74%	36.13%
1994	62.78%	52.83%	70.37%	73.26%	58.62%
1995	64.02%	51.74%	68.62%	59.46%	71.65%
1996	46.78%	28.99%	56.10%	54.91%	54.83%
1997	44.36%	25.93%	49.93%	27.91%	54.44%
1998	26.04%	19.11%	24.99%	62.25%	30.55%
1999	54.63%	47.43%	58.15%	87.43%	55.06%
2000	75.00%	56.98%	76.21%	37.09%	84.20%
2001	23.07%	8.29%	24.76%	0.01%	40.30%
2002	0.01%	0.01%	0.03%	0.00%	0.02%
2003	0.00%	0.00%	0.00%	0.00%	0.00%
2004	0.00%	0.00%	0.00%	36.13%	0.00%
Total	33.76%	45.13%	45.13%	3613.00%	28.38%

Table 3.3: Default Risk Regression by LTV Ratio

	LTV>=90				80<=LTV<90			
	Without Shared Frailty		With Shared Frailty		Without Shared Frailty		With Shared Frailty	
	Coeff	*T*-stat	Coeff	*T*-stat	Coeff	*T*-stat	Coeff	*T*-stat
LTV ratio	0.054	11.040	0.070	13.910	0.044	2.660	0.046	2.750
Log Loan Size	-0.209	-5.130	-0.266	-6.270	-0.319	-3.640	-0.428	-4.650
Points	-0.356	-12.060	-0.397	-12.010	0.047	0.440	0.023	0.210
Original Contract Rate	0.239	17.440	0.250	8.560	0.091	2.520	0.227	4.060
Spread	0.005	3.310	0.005	1.580	0.011	2.940	0.008	1.640
State Unemployment Rate	-0.124	-8.950	-0.068	-4.420	-0.067	-2.050	-0.027	-0.720
State-level House Price Ratio	-8.196	-67.060	-9.162	-68.660	-8.207	-29.650	-8.890	-30.430
Theta			1.228	3.779			0.611	2.971
	70<=LTV<80				LTV<70			
	Without Shared Frailty		With Shared Frailty		Without Shared Frailty		With Shared Frailty	
LTV ratio	Coeff	*T*-stat	Coeff	*T*-stat	Coeff	*T*-stat	Coeff	*T*-stat
Log Loan Size	-0.002	-0.230	0.004	0.350	0.077	11.800	0.077	11.740
Points	-0.335	-5.480	-0.412	-6.410	-0.340	-3.440	-0.298	-2.920
Original Contract Rate	-0.071	-1.080	-0.186	-2.730	-0.075	-0.990	-0.124	-1.520
Spread	0.249	9.370	0.474	8.590	0.296	6.810	0.457	7.310
State Unemployment Rate	0.014	3.780	-0.001	-0.100	-0.001	-0.120	-0.004	-0.680
State-level House Price Ratio	-0.006	-0.230	0.055	1.890	0.042	0.940	0.079	1.680
Theta	-7.581	-38.250	-8.149	-37.140	-6.789	-22.800	-7.466	-23.070
			1.109	3.258			0.576	2.644

Table 3.4: Prepayment Risk Regression by LTV Ratio

	LTV>=90				80<=LTV<90			
	Without Shared Frailty		With Shared Frailty		Without Shared Frailty		With Shared Frailty	
	Coeff	*T*-stat	Coeff	*T*-stat	Coeff	*T*-stat	Coeff	*T*-stat
LTV ratio	-0.070	-77.620	-0.069	-73.350	0.021	9.910	0.004	1.790
Log Loan Size	0.680	90.320	0.851	107.280	0.606	51.620	0.663	54.200
Points	-0.350	-63.790	-0.227	-44.440	-0.364	-28.600	-0.247	-21.500
Original Contract Rate	0.190	69.010	0.210	34.030	0.167	34.090	0.257	24.120
Spread	0.011	31.990	0.001	2.440	0.012	21.370	-0.001	-1.400
State Unemployment Rate	-0.197	-69.110	-0.291	-97.750	-0.175	-37.250	-0.262	-51.290
State-level House Price Ratio	-3.927	-230.240	-4.969	-257.660	-3.781	-136.340	-4.537	-146.280
Theta			2.679	4.469			1.938	4.286
	70<=LTV<80				LTV<70			
	Without Shared Frailty		With Shared Frailty		Without Shared Frailty		With Shared Frailty	
	Coeff	*T*-stat	Coeff	*T*-stat	Coeff	*T*-stat	Coeff	*T*-stat
LTV ratio	0.012	12.520	-0.012	-11.830	-0.010	-29.640	-0.017	-52.670
Log Loan Size	0.598	107.310	0.691	119.360	0.333	53.820	0.507	80.870
Points	-0.317	-55.830	-0.235	-46.050	-0.229	-38.690	-0.111	-20.570
Original Contract Rate	0.256	109.380	0.323	54.240	0.215	74.250	0.349	50.550
Spread	0.009	27.160	-0.006	-11.720	0.007	17.990	-0.005	-8.450
State Unemployment Rate	-0.187	-75.350	-0.225	-84.630	-0.156	-52.630	-0.181	-58.080
State-level House Price Ratio	-3.239	-252.800	-4.009	-266.620	-3.401	-230.030	-4.188	-247.900
Theta			2.371	4.418			2.313	4.406

Table 3.5: Estimated Frailty Values for Each Originated Year --- Default Risk

Year	**LTV ratio at the time of loan origination**			
	LTV>=90	80<=LTV<90	70<=LTV<80	LTV<70
1976	2.520	1.671	1.525	2.115
1977	4.023	1.418	1.366	1.581
1978	0.756	0.509	0.757	0.931
1979	0.759	0.274	0.262	0.395
1980	0.341	0.414	0.257	0.589
1981	0.495	0.703	0.065	0.269
1982	0.203	1.432	0.091	0.669
1983	0.484	1.264	0.131	0.993
1984	1.178	0.929	0.347	1.641
1985	1.924	2.018	0.917	2.344
1986	1.777	1.382	1.574	1.101
1987	0.974	0.752	0.948	0.794
1988	0.609	0.240	0.355	0.274
1989	0.417	0.320	0.631	0.354
1990	0.459	0.431	0.655	0.433
1991	0.756	0.800	1.034	0.649
1992	1.136	1.120	1.227	0.753
1993	1.400	1.608	1.990	1.072
1994	1.752	1.506	2.547	1.980
1995	2.138	2.260	3.027	2.349
1996	1.276	2.194	3.795	2.160
1997	0.637	1.217	1.304	0.906
1998	0.642	0.974	0.733	0.463
1999	1.810	1.817	2.300	1.553
2000	0.443	0.438	0.447	0.515
2001	0.028	0.220	0.167	0.630
2002	0.008	0.249	0.144	0.480
2003	0.011	0.188	0.111	0.317
2004	0.044	0.653	0.295	0.692
Estimated variance	1.227	0.611	1.109	0.576

Table 3.6: Estimated Frailty Values for Each Originated Year --- Prepayment Risk

Year	**LTV ratio at the time of loan origination**			
	LTV>=90	80<=LTV<90	70<=LTV<80	LTV<70
1976	12.741	8.367	6.236	5.895
1977	5.478	3.606	3.052	2.958
1978	1.618	1.182	1.246	1.048
1979	0.637	0.521	0.383	0.308
1980	0.339	0.225	0.132	0.143
1981	0.236	0.186	0.072	0.098
1982	0.172	0.429	0.079	0.112
1983	0.282	0.370	0.116	0.221
1984	0.396	0.265	0.186	0.245
1985	0.393	0.439	0.329	0.289
1986	0.306	0.419	0.520	0.412
1987	0.357	0.463	0.466	0.353
1988	0.120	0.147	0.151	0.117
1989	0.070	0.110	0.102	0.092
1990	0.081	0.117	0.098	0.089
1991	0.135	0.219	0.233	0.193
1992	0.270	0.394	0.438	0.377
1993	0.409	0.661	0.716	0.635
1994	0.425	0.731	0.806	0.690
1995	0.522	0.927	1.047	1.070
1996	0.418	0.974	1.118	1.108
1997	0.452	1.196	1.394	1.514
1998	0.365	0.664	0.855	1.154
1999	1.074	1.986	2.383	2.553
2000	1.506	3.504	5.015	4.865
2001	0.199	0.890	1.824	2.459
2002	0.000	0.002	0.001	0.002
2003	0.000	0.001	0.000	0.000
2004	0.001	0.007	0.001	0.001
Estimated variance	2.679	1.938	2.371	2.313

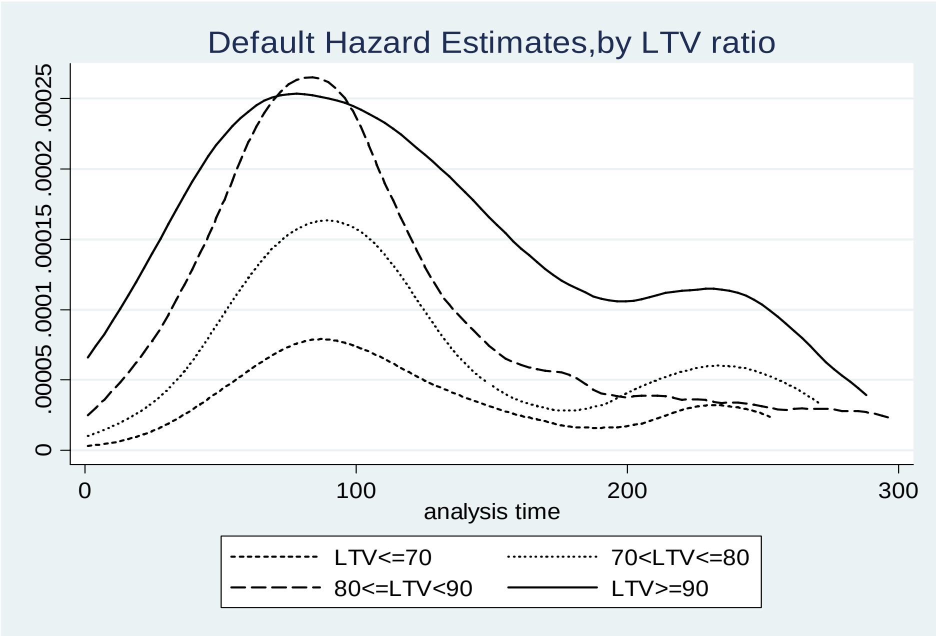


Figure 3.1: Default hazard curves by LTV Ratio

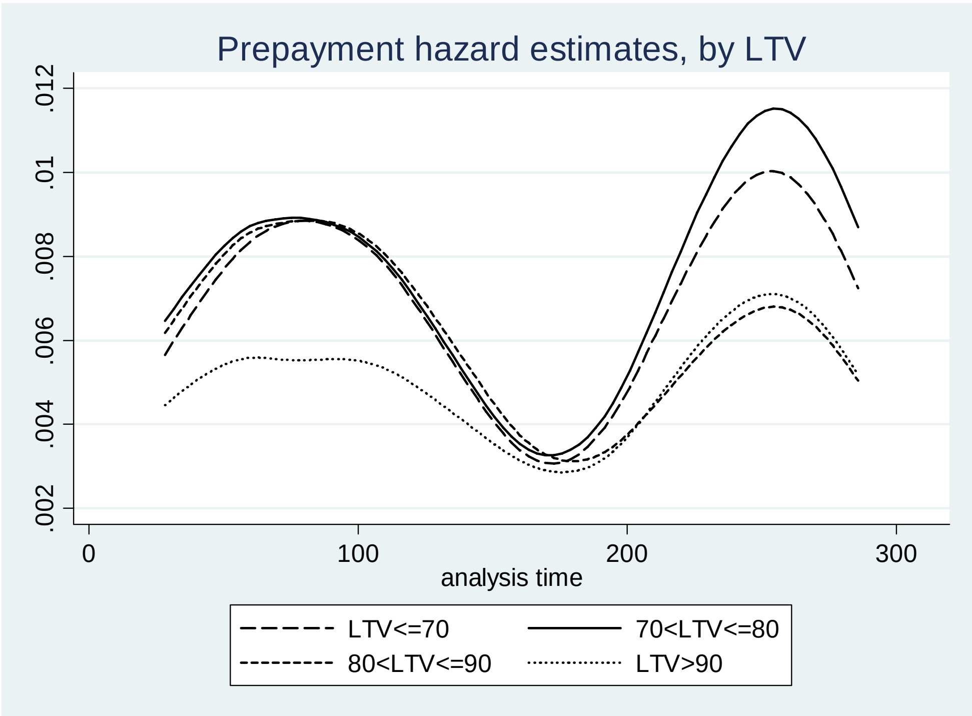


Figure 3.2: Prepayment hazard curves by LTV Ratio

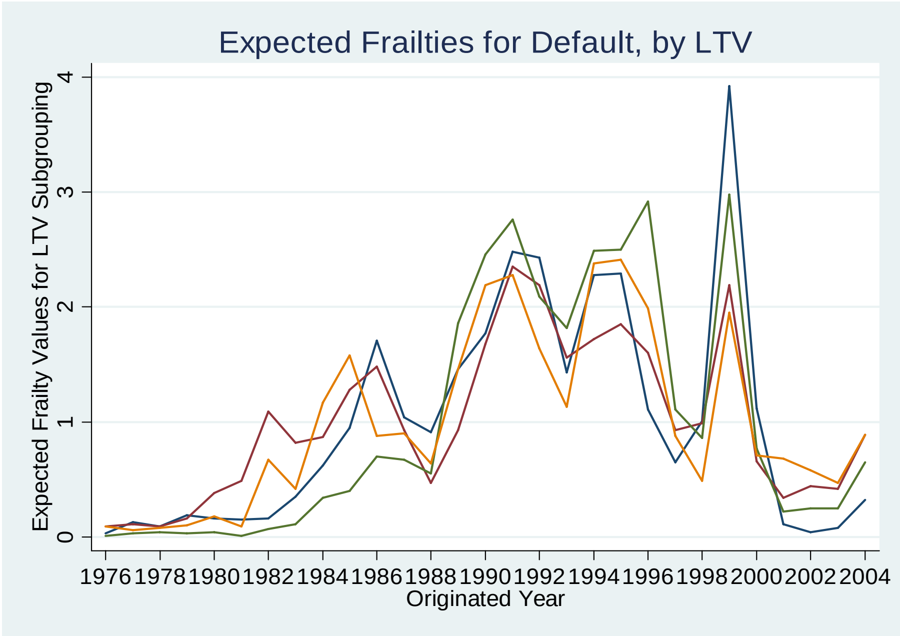


Figure 3.3: Expected frailties for default hazard by LTV ratio

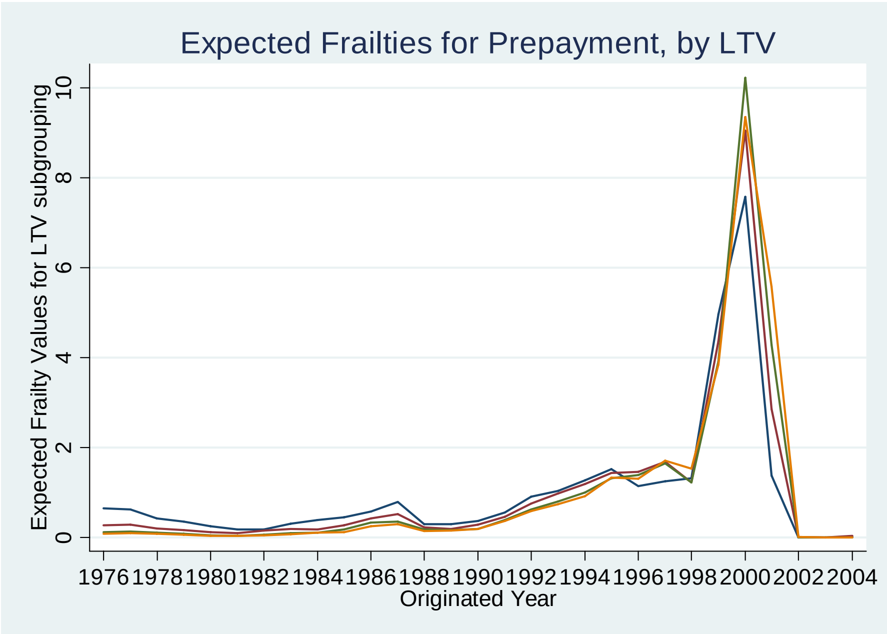


Figure 3.4: Expected frailties for prepayment risks by LTV ratio

CHAPTER 4

DO MORTGAGE INTEREST RATES, DEFAULT RATES, AND PREPAYMENT RATES VARY BY SOCIOECONOMIC CHARACTERISTICS OF NEIGHBORHOODS?

4.1. Introduction

Great attention has been focused on differential rejection rates in regard to the redlining, the discrimination against the low-income and minority neighborhoods. The redlining behavior may depress property values and the homeownership rates in those neighborhoods. The concerns about possible discriminatory behavior in mortgage markets have led to passage of the Community Reinvestment Act (CRA) in 1977, a statute designed by the federal government to encourage lending by financial institutions to nearby lower-income neighborhoods. The federal government has expressed a strong interest in assuring that potential homebuyers in minority and low-income neighborhoods have full access to credit. The CRA's focus was to assure that borrowers in low- and moderate-income neighborhoods were not disadvantaged in terms of credit access because of their location.

Most of redlining studies find little evidence of differential treatment of race (Avery and Buynak, 1981; Gabriel and Rosenthal, 1991; Schill and Wachter, 1993; Holmes and Horvitz, 1994). However, even though loan denial rates may not vary by neighborhood characteristics, redlining may occur in a more subtle form, such as a variation in loan pricing by neighborhood. Schill and Wacheter (1994) find evidence of loan concentration effects in lower-income neighborhoods and Evanoff and Segal (1996) find increased application and origination flows in low-income neighborhoods. Variation in mortgage interest rates and origination fees by the racial composition or relative income of a neighborhood may reflect legitimate economic considerations relating to credit risk, origination cost, or expected mortgage life. In other words, racial geographic disparities in mortgage lending may result from a number of factors other than discriminations. Neighborhood risk factors, insufficient information (Lang and Nakamura, 1993), an imbalance between supply and demand of housing units, as well as regulatory policy itself may affect geographic lending patterns. Kau et al (2008) find the evidence that borrowers in predominately black neighborhoods pay significantly higher interest rates than is consistent with evidence of their behavior.

To test the presence and source of neighborhood disparities in lending decisions, the present study uses a nationally representative sample of 30-year fixed-rate residential mortgages with detailed loan origination data for the years 1998-2004 to examine how the mortgage rates vary by neighborhoods, after controlling for the loan-specific and demographic variables. The mortgage data with rich information on the individual loan characteristics, combined with other external macroeconomic variables, such as the unemployment rate and the state level housing price index at the loan origination, and the neighborhood characteristics, are used to determine whether mortgage rates on fixed-rate loans vary by the income and racial composition of the neighborhood. In particular, the question—whether poor and minority concentrated neighborhoods under CRA are subject to lower mortgages rates – is of interest.

In addition, this essay adds to the existing literature by examining the neighborhood distribution of default and prepayment rates. A unique feature of the mortgage data used in this essay is that mortgage termination behavior via either default or prepayment was recorded. With this advantage, I am able to conduct an analysis about the impacts of demographic variables and CRA policy on the variation of defaults and prepayments, at both loan-level and the ZIP level.

Few studies have documented the effects of neighborhood characteristics on default and prepayment rates. Among the few studies, Cotterman (2001) found that lower-tract income and higher black racial composition are associated with higher default rates. Archer et al (1996) analyze the prepayment behavior of low-income borrowers using American Housing Survey Data and found low-income households are not significantly different from others with regard to mortgage terminations.

This essay is structured as follows. The next section describes the data and methodology used in the analysis. The section thereafter presents the results. A concluding section summarizes the analysis.

4.2. Data and Analysis

This analysis is based on the mortgage performance data from a large financial service institution with the origination year between 1998 and 2004. Each loan recorded in the data contains information including the principal loan amount, origination coupon rate, the loan-to-value (LTV) ratio, the borrowers' FICO score at loan origination, the loan amount, the points paid at origination, and the five-digit ZIP code location of the property. A dummy variable indicating whether the loan is subject to the CRA policy is also available. For this essay's purpose, each loan was matched to an external dataset containing the characteristics of the five digit ZIP code area based on data collected as part of the 2000 census of Population and Housing; the census data include variables of racial composition and median household income in each ZIP code. In order to control the effects from general and local economy, I extracted macroeconomic level variables, namely, the state-level per capita annual GDP and state-level monthly unemployment rate, from the Bureau of Economic Analysis (BEA) and Bureau of Labor Statistics (BLS) respectively. The House Price Index (HPI) indicating the movement of single-family house prices, which is publicly available at the Office of Federal Housing Enterprise Oversight (OFHEO), is used in this analysis as well. After merging the mortgage performance data and the external datasets, the final sample includes 677,764 30-year fixed-rate loans in 21,186 ZIP codes; this represents more than 50% of all valid ZIP codes in the U.S.

The main purpose of this essay is to investigate how the neighborhood characteristics and the CRA policy affect the mortgage contract rates as most research on racial discrimination in home finance accepts the assumption that interest rates in home mortgage markets do not differ across borrowers. To conduct this research, I examine the effects of neighborhood characteristics on mortgage contract rates after controlling for the loan-level characteristics and the macroeconomic information. Standard multiple regressions are conducted at both the loan-level and the ZIP code level. These two regressions can be used to validate the robustness of the estimation as each of them represents disaggregate and aggregate level analysis respectively. For the loan-level multiple regression, the dependent variable is the mortgage contract rate and the independent variables are intended to capture the effects of neighborhood characteristics (ZIP code racial composition and ZIP code median household income), the loan-level differences (the original LTV ratio, the original loan amount, the points at origination, the FICO score at loan origination, CRA, and the spread[11]), and the macroeconomic variables (the state-level per capital GDP, the state-level house price index, and the state-level unemployment rate at the loan origination). For the ZIP code level analysis, the dependent and independent variables are imputed as the average of those loan-level variables for the entire ZIP code. In order to determine the variation of default and prepayment rates in the neighborhood, I use the average default and prepayment rates for each ZIP code as the dependent variables. A complete list of model variables including a brief description is included in Table 1 and Table 2. Table 3 and

[11] Spread is defined as the relative difference between the loan's contract rate and the current 10-year treasury rate.

Table 4 present the summary statistics for each variable. The value of mortgage contract rates ranges from 3.75 to 13.25 for our sample, with the sample mean value equal to 6.51. The summary of the racial composition shows that the Hispanic population account for approximately 14% of the total population in the included neighborhoods while African-American population account for about 10% of the total population. More than 14% of all the neighborhoods have population with more than 30% of the Hispanic population and less than 10% of all the neighborhoods have population with more than 30% of the African-American population. About 28% of loans in our sample are CRA covered and more than 90% of loans were originated in the metropolitan areas.

4.3. Results

The mortgage performance data allow a unique opportunity to examine how mortgage loan features vary by the relative income[12] and minority share[13] of neighborhoods. Especially the data set contains FICO[14] score which allows for the investigation of the connection between the credit risks and the neighborhood characteristics. Table 5 compares the LTV and FICO distribution based on the income and minority composition for the ZIP code. Over 60% originated loans' LTV ratio is equal or below 80. The upper panel contains the LTV and FICO distribution within each relative income bucket. The share of high-LTV loans decreases as the relative income of the neighborhood increases while the share of low-FICO loans increases as the relative income of neighborhood increases. Less than 7% of loans in high-income ZIP codes had LTVs above 90%, compared to 29% in the lowest income areas. It is noteworthy that the share of low-LTV loans increases as the relative income of the neighborhood increases. For

[12] The relative income threshold is defined as the quartile of the median household income. The cutoff points are: 38,187, 48,452, and 61,475.
[13] The minority share is defined as the percentage of non-White population for each ZIP code.
[14] The FICO score threshold is defined as the quartile of the FICO score recorded. The cutoff points are: 693, 736, and 771.

example, the group with the highest proportion (close to 90%) of low LTV loans (less than 80) was the high-income ZIP code group (the median household income above $61,475). At the same time, high-FICO loans appear to increase with neighborhood income. In low-income areas, 21% of originated mortgages were low-FICO loans, compared to 12% of loans in high-income areas.

The lower panel of Table 5 shows the variation in the LTV and FICO distributions based on the minority composition. Here, the share of high-LTV loans increases with minority concentration. Over 25% of loans in areas where at least 30% of the population are minorities have high LTVs (LTV above 90%) while only 16% of loans have high LTVs for areas where the minority share is less than 10%. Also the share of low-LTV loans decreases with minority concentration. This pattern reflects that minorities tend to have lower wealth. In addition, areas with high minority concentration have a larger share of low-FICO (FICO<=693) loans. For areas with less than 10% minority concentration, the share of low-FICO loans was 14.5% while for areas with more than 50% minority concentration the low-FICO loans have account for more than 24%. This pattern is most likely due to the fact that minorities have lower credit scores.

An additional advantage of the current mortgage dataset is that it provides the information on the termination events via default or prepayment. Thus it allows one to perform a distribution analysis of defaults and prepayments based on the distribution of the ZIP code characteristics. Table 6 presents the results. It is interesting to note that default rates increases monotonically as the share of minority increases while default rates decreases monotonically as the neighborhood income increases. The average default rate for the high-income areas in only 0.02%, compared to 0.16% in low-income areas. At the same time, the average default rate of the high-minority concentrated areas is three times as that of the low-minority concentrated areas.

The pattern of prepayment rates demonstrates an inverse trend compared with that of the default rates: the prepayment rates appear to increase as the relative income of neighborhood increases, and it appears to decrease as the minority concentration increases.

The regression analysis examines the determinants of mortgage interest rates and points a percentage of loan amounts. As Table 7 shows, borrower race does significantly affect the level of mortgage rates: holding other factors constant, borrowers in Hispanic concentrated neighborhood (with at least 30% Hispanic population) pay higher rates --- about 11 basis points higher. Borrowers in neighborhood with at least 30% African-American population, however, pay about 5 basis points lower in interest rates. These results are consistent with the finding by Northaft and Perry (2002) who indicated that Hispanic neighborhoods pay more in interest rates while borrowers in African-American neighborhoods pay about the same or less interest rates. The regression shows that borrowers in low-income neighborhoods generally pay slightly higher rates. However, it is worth noting that borrowers in high-income neighborhoods also pay higher interest rates --- about 3.6 basis points more. After controlling for the property attributes and neighborhood characteristics, the CRA policy demonstrates a strong significant effect on the interest rates: the CRA covered borrowers typically pay about 14 basis points lower in interest rates. This shows direct evidence that CRA has been implemented to lessen racial discrimination in neighborhoods. In addition, borrowers with CRA subsidies are observed to pay less amount of fee. The owner-occupied homes are charged with lower interest rates and less fees.

The loan-related variables demonstrate expected effects on the interest rates. Evidence of pricing by credit risk is evident in the LTV and FICO variables. Loans with a LTV above 90% (HILTV) cost borrowers 10 basis points more, while loans with a LTV below 80% (LOLTV) cost borrowers 5 basis points lower. Borrowers with FICO score more than 771 pay less basis points in interest rates than borrowers with lower FICO scores, after controlling for all other factors. The original loan amount variable (LOAN) is strong negative, consistent with evidence that average fixed costs of originating and servicing loans decline with loan size. Loans secured by homes in metropolitan areas had mortgages rates that were 5 basis points higher. The POINTS variable is inversely related to the contract interest rate, as expected.

When examining the effect of the general and local economy variables, the unemployment rate at the loan origination, is negatively related to the interest rates while the state per capita GDP is positively related to the level of interest rates. This suggests that mortgage rates are sensitive to the change of the macroeconomic environment.

Table 8 shows different regression results using ZIP code as the unit of analysis. The dependent variables chosen for the ZIP code regression are average mortgage interest rates, the average default rate and prepayment rate for the ZIP code. The major finding of the research is that mortgage rates and prepayment rates do vary by the demographic makeup of the neighborhood. The pattern of the variation of mortgage rates is consistent with the loan-level regressions in Table 7. Table 8 displays that racial composition does affect prepayment rates and default rates. For example, borrowers in areas where at least 30% of population are African-American or Hispanic had lower prepayment rates. It is interesting to note that residents in both high-income areas (median household income above $61,475) and low-income areas (median household income below $38,187) had lower prepayment rates. However, the coefficient on LOINCOME is insignificant. For the level of default rates, BLK30 is significant and positive, indicating that areas with at least 30% African-American population tend to have higher default rates; HIS10 is negative and significant, suggesting that areas with no more than 10% of Hispanic population tend to have lower default rates. This pattern is mostly due to the fact that minorities have lower wealth and are most likely to have default events. The income distribution, however, shows no evidence of effect on the level of default rates.

The coefficient of CRA effect on the prepayment rates implies that the CRA-covered loans tend to have higher prepayment rates, compared to those without CRA coverage. The MOCCUP is also significant and positive, meaning that owner-occupied homes may have higher prepayment rates as well.

A discussion of the interpretation of the coefficients may be useful. As an example, consider the coefficient on MLTV of -0.004 on the prepayment rates. This can be interpreted as a one percent increase of LTV ratio would decrease the predicted value of prepayment rates by 0.004. The interpretation of this variable is consistent with the individual mortgage termination behavior: high LTV ratio is expected to decrease the probability of prepayment because high LTV ratio makes refinancing most costly and less attractive.

4.4. Conclusions

This essay estimates the mortgage interest rate differences by borrowers with a national home mortgage lender during the years 1998-2004. After controlling for all relevant factors, this research finds evidence of variation in loan pricing by the relative income and racial mix of a neighborhood. In particular, borrowers in Hispanic concentrated areas pay higher mortgage rates, with a similar effect for borrowers in low-income areas. The CRA lending programs may overlap with predominately African-American neighborhoods, resulting in lower observed interest rates in these neighborhoods. The mortgage dataset used contains information on CRA coverage and the creditworthiness of borrowers, which provide a unique opportunity to explain the differentials we observe. For example, generally speaking, both the CRA effect and the FICO score are observed to be negative on the level of mortgage interest rates. This reflects the effects of both external policy and internal credit risk on the variation of mortgage rates. In addition, there exist negative significant effects of minority concentration and income distribution on the prepayment rates across neighborhood. This result confirms with the observation that the probability of prepayments is negatively associated with the wealth effect.

Table 4.1: Variable Definition for Loan-level Regressions

Variable Name	Description
ORIGINALRATE	The loan's contract rate (dependent variable)
ZIP code variables	
HISP	The percentage of residents that are Hispanic
HIS1030	1 if Hispanic share between 0.1 and 0.3; 0 otherwise
HIS30	1 if Hispanic share 0.3 or greater; 0 otherwise
BLACK	The percentage of residents that are African-American
BLK1030	1 if African-American shared between 0.1 and 0.3; 0 otherwise
BLK30	1 if African-American shared 0.3 or greater; 0 otherwise
INCOME	The log form of median household income
METRO	1 if ZIP code located within MSA; 0 otherwise
Loan-related variables	
LTV	Loan-to-Value ratio
HILTV	1 if LTV>=90; 0 otherwise
LOLTV	1 if LTV<=70; 0 otherwise
POINTS	Points paid at the loan origination
LOAN	Log of the original loan amount
ORIGFICO	The borrower's FICO score at loan origination
CRA	1 if the loan if under CRA; 0 otherwise
OCCUP	1 if the property is owner-occupied; 0 otherwise
SPREAD	The relative difference between the contract rate and the current 10-year treasury rate
Macroeconomic variables	
UNEM	The monthly state-level unemployment rate
PAGDP	The annual state-level per capita GDP
STATEHPI	The state-level house price index

Table 4.2: Variable Definition for ZIP code-level Regressions

Variable Name	Description
MCOUPON	The average loan contract rate per ZIP code (dependent variable)
ZIP code variables	
MHISP	The percentage of residents that are Hispanic
MHIS1030	1 if Hispanic share between 0.1 and 0.3; 0 otherwise
MHIS30	1 if Hispanic share 0.3 or greater; 0 otherwise
MBLACK	The percentage of residents that are African-American
MBLK1030	1 if African-American shared between 0.1 and 0.3; 0 otherwise
MBLK30	1 if African-American shared 0.3 or greater; 0 otherwise
MINCOME	The log form of median household income
MMETRO	1 if ZIP code located within MSA; 0 otherwise
Loan-related variables	
MLTV	The average LTV ratio per ZIP code
MPOINTS	Average points paid at the loan origination
MLOAN	Average log form of the original loan amount
MFICO	Average FICO score for consumers located in the given ZIP code
MCRA	The percentage of CRA-related loans per ZIP code
MOCCUP	The percentage of owner-occupied properties per ZIP code
MSPREAD	The average spread values per ZIP code
MDEFAULT	Average default rate
MPREPAY	Average prepayment rate
Macroeconomic variables	
MUNEM	The monthly state-level unemployment rate
MAGDP	The annual state-level per capita GDP
MSTATEHPI	The state-level house price index

Table 4.3: Summary Statistics for Loan-level Regressions

Variables	Mean	Std. Deviation	Minimum	Maximum
ORIGINALRATE	6.513	0.945	3.75	13.25
HISP	0.135	0.170	0	0.986
HIS1030	0.256	0.436	0	1
HIS30	0.142	0.349	0	1
BLACK	0.098	0.158	0	0.985
BLK1030	0.173	0.378	0	1
BLK30	0.097	0.295	0	1
INCOME	10.796	0.332	7.824	12.206
METRO	0.903	0.295	0	1
LTV	74.321	18.557	20	125
HILTV	0.241	0.428	0	1
LOLTV	0.439	0.496	0	1
POINTS	-0.056	0.706	-3	8
LOAN	11.757	0.553	9.903	13.021
ORIGFICO	728.326	53.829	427	849
CRA	0.282	0.450	0	1
OCCUP	0.896	0.305	0	1
SPREAD	28.218	6.761	-37.600	63.020
UNEM	5.246	1.225	2.1	13
PAGDP	34870.7	7015.3	22395	116441
STATEHPI	273.369	76.980	135.140	674.320

Table 4.4: Summary Statistics for ZIP code-level Regressions

Variables	Mean	Std. Deviation	Minimum	Maximum
MCOUPON	6.685	0.621	4	9.75
MHISP	0.078	0.145	0	0.986
MHIS1030	0.115	0.319	0	1
MHIS30	0.131	0.337	0	1
MBLACK	0.081	0.158	0	0.985
MBLK1030	0.118	0.322	0	1
MBLK30	0.144	0.351	0	1
MINCOME	10.626	0.350	7.824	12.206
MMETRO	0.642	0.479	0	1
MLTV	77.054	11.467	20	113.5325
MPOINTS	-0.093	0.457	-3	6
MLOAN	11.563	0.441	9.903	13.021
MFICO	723.893	34.373	469	839
MCRA	0.228	0.280	0.000	1.000
MOCCUP	0.860	0.245	0	1
MSPREAD	28.261	3.919	-20.6	52.42857
MDEFAULT	0.001	0.018	0	1
MPREPAY	0.178	0.240	0.000	1.000
MUNEM	4.509	1.087	2.1	13
MAGDP	32957.5	4880.8	22395	116441
MSTATEHPI	228.437	57.259	135.140	662.770

Table 4.5: ZIP code demographic characteristics and distribution of loan originations by LTV and FICO

ZIP code characteristics	LTV			FICO			
	<=80	80-90	>=90	>=771	736-771	693-736	<=693
Relative Income							
<=25%	60.17%	10.79%	29.04%	42.86%	17.08%	18.58%	21.47%
25-50%	66.31%	9.98%	23.71%	45.69%	17.67%	18.01%	18.63%
50-75%	71.77%	8.96%	19.27%	46.72%	18.63%	17.77%	16.88%
>75%	88.62%	4.75%	6.62%	50.48%	21.17%	16.46%	11.89%
Minority Percent							
<=10%	74.21%	10.14%	15.65%	51.03%	17.85%	16.62%	14.50%
10-30%	71.43%	9.00%	19.58%	46.56%	18.85%	17.90%	16.69%
30-50%	65.96%	8.13%	25.91%	41.63%	18.67%	19.06%	20.63%
>50%	61.79%	8.56%	29.64%	38.49%	17.94%	19.41%	24.16%

Table 4.6: ZIP code demographic characteristics and distribution of Defaults and Prepayments

ZIP code characteristics	Default	Prepayment
Relative Income		
<=25%	0.16%	12.91%
25-50%	0.11%	17.39%
50-75%	0.08%	21.07%
>75%	0.02%	26.97%
Minority Percent		
<=10%	0.06%	22.26%
10-30%	0.09%	19.56%
30-50%	0.12%	16.89%
>50%	0.18%	14.26%

Table 4.7: OLS Estimates of Determinants of the Interest rates and Points

INDEP.VAR	ORIGINALRATE	POINTS
BLK10	0.188	-0.073
	(79.87)**	(32.45)**
HIS10	-0.192	0.027
	(90.01)**	(13.18)**
BLK30	-0.051	0.107
	(13.66)**	(30.14)**
HIS30	0.102	0.065
	(33.77)**	(22.55)**
HIINCOME	0.036	-0.049
	(8.55)**	(12.43)**
LOINCOME	0.011	-0.009
	(4.66)**	(4.31)**
HILTV	0.11	0.051
	(44.81)**	(21.84)**
LOLTV	-0.054	0.045
	(23.86)**	(20.79)**
HIFICO	-0.203	-0.025
	(84.20)**	(11.00)**
LOFICO	-0.069	-0.029
	(28.66)**	(12.73)**
LOAN	-0.097	-0.045
	(58.57)**	(28.61)**
CRA	-0.135	-0.073
	(63.28)**	(35.52)**
OCCUP	-0.189	-0.089
	(64.46)**	(31.62)**

Table 4.7: OLS Estimates of Determinants of the Interest rates and Points (continued)

METRO	0.052	0.008
	(16.45)**	(2.73)**
UNEM	-0.368	-0.043
	(433.68)**	(46.78)**
STATEHPI	-0.003	-0.001
	(198.31)**	(80.47)**
PAGDP	0.015	0.002
	(104.52)**	(15.18)**
ORIGINALRATE		-0.183
		(160.55)**
POINTS	-0.201	
	(160.55)**	
INTERCEPT	9.997	2.213
	(521.82)**	(103.09)**
Observations	677611	677611
R-squared	0.42	0.06
Absolute value of t statistics in parentheses		

Table 4.8: OLS estimates of ZIP code-level Regressions

INDEP.VAR	ORIGINALRATE	DEFAULT RATE	PREPAYMENT RATE
BLK10	0.077	0.001	0.022
	(7.39)**	-1.71	(6.83)**
HIS10	-0.034	-0.001	-0.011
	(3.22)**	(3.97)**	(3.31)**
BLK30	0.002	0.001	-0.029
	-0.13	(2.53)*	(6.09)**
HIS30	0.046	0	-0.019
	(2.75)**	-0.48	(3.76)**
HIINCOME	0.142	0	-0.02
	(6.40)**	-0.52	(3.00)**
LOINCOME	-0.03	0	-0.001
	(3.23)**	-0.58	-0.5
MLTV	0.008	0	-0.004
	(19.72)**	(4.67)**	(33.40)**
MPOINTS	-0.206	0	-0.005
	(25.79)**	-0.04	(2.19)*
MFICO	-0.001	0	0
	(12.50)**	-0.76	(13.58)**
MLOAN	-0.299	-0.001	0.139
	(27.52)**	(3.16)**	(41.47)**
MCRA	-0.116	0.001	0.032
	(6.80)**	-1.55	(6.17)**
MOCCUP	-0.275	0	0.012
	(19.12)**	-0.89	(2.69)**

Table 4.8: OLS estimates of ZIP code-level Regressions (continued)

METRO	0.121	0	-0.002
	(13.76)**	-0.3	-0.75
UNEM	-0.146	0	-0.012
	(39.81)**	-0.82	(10.17)**
STATEHPI	-0.001	0	0
	(16.00)**	-1.48	(6.50)**
PAGDP	0.011	0	0
	(12.80)**	-0.5	-0.12
ORIGINALRATE		0.001	0.218
		(4.89)**	(97.67)**
MSPREAD		0	-0.005
		-0.26	(16.71)**
INTERCEPT	11.032	0.012	-0.437
	(76.53)**	(2.67)**	(8.04)**
Observations	18592	18592	18592
R-squared	0.23	0.01	0.17
Absolute value of t statistics in parentheses			
* significant at 5%; ** significant at 1%			

CHAPTER 5

CONCLUSIONS

In the analysis of mortgage termination using the hazard model approach, situations where the survival times of mortgages are not independent are often encountered. In particular, for example, mortgages originated from the same area or same year may be similar in terms of duration times. The shared-frailty model provides a method for modeling survival data when the survival times are not independent. The frailty, representing the effects of measurement errors and missing variables, is modeled as a non-negative latent random variable that acts multiplicatively on the hazard function. This dissertation uses MSA and the originated year as the group variables. The empirical results suggest that it is important to control for the group level frailty to account for the within-group correlation among individual mortgages. Differences in environment and in macroeconomic setting are likely to have an important influence on mortgage termination risks.

This dissertation also finds evidence of variation in loan pricing by the relative income and racial mix of neighborhoods. In particular, borrowers in Hispanic concentrated areas pay higher mortgage rates, with a similar effect for borrowers in low-income areas. The CRA lending programs may overlap with predominately African-American neighborhoods, resulting in lower observed interest rates in these neighborhoods. The mortgage dataset I use in this dissertation contains information on CRA coverage and the creditworthiness of borrowers, which provide a unique opportunity to explain the differentials we observe. For example, generally speaking, both the CRA effect and the FICO score are observed to be negative on the level of mortgage interest rates.

This reflects the effects of both external policy and internal credit risk on the variation of mortgage rates. In addition, there exist negative significant effects of minority concentration and income distribution on the prepayment rates across neighborhood. This result confirms with the observation that the probability of prepayments is negatively associated with the wealth effect.

REFERENCES

Ambrose, B. W., and A. B. Sanders. 2003. "Commercial Mortgage-backed Securities: Prepayments and Defaults," *The Journal of Real Estate Finance and Economics* 26(2-3): 179-196.

Ambrose, B. W., and A. Pennington-Cross. 2000. "Local Economic Risk Factors and the Primary and Secondary Mortgage Markets," *Regional Science and Urban Economics:* 683-701.

Ambrose, B. W., M. LaCour-Little, and A. B. Sanders. 2004. "The Effect of Conforming Loan Status on Mortgage Yield Spreads: A Loan Level Analysis," *Real Estate Economics* 32 (4): 541-569.

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

Avery, R.B. and Buynak, T.M., 1981. "Mortgage redlining: some new evidence," *Economic Review*. Federal Reserve Bank of Cleveland, 15-31.

Brueckner, J. K. 2000. "Mortgage Default with Asymmetric Information," *Journal of Real Estate Finance and Economics* 20: 251-275.

Calem, P. S. and M. LaCour-Little. 2004. "Risk-Based Capital Requirements for Mortgage Loans," *Journal of Banking and Finance* 28(3): 647-672.

Capozza, D. R., D. Kazarian, and T. A. Thomson. 1997. "Mortgage Default in Local Markets," *Real Estate Economics* 25(4): 631-655.

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

Clayton, D.G. 1978 "A Model for Association in Bivariate Life Tables and its Application in Epidemiological Studies of Familial Tendency in Chronic Disease Incidence," *Biometrika* 65(1): 141-151.

Ciochetti, B.A., Y. H. Deng, B. Gao, and R. Yao. 2002. "The Termination of Commercial Mortgage Contracts through Prepayment and Default: A Proportional Hazard Approach with Competing Risks," *Real Estate Economics* 30(4): 595-633.

Cotterman, R. 2001. "Neighborhood Effects in Mortgage Default Risk," U.S Department of Housing and Urban Development, Office of Policy Development and Research.

Cowan, A. M. and C.D. Cowan. 2004. "Default Correlation: An Empirical Investigation of a Subprime Lender," *Journal of Banking and Finance* 28: 753-771.

Das, S., L. Freed, G. Geng, and N. Kapadia. 2001. "Correlated Default Risk," working paper, Santa Clara University.

Dempster, A.P., N.M. Laird, and D.B. Rubin. 1977. "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society, Series B (Methodological)* 39(1): 1-38.

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

Duffie, D., A. Eckner, G. Horel, and L. Saita. 2006. "Frailty Correlated Default," working paper, Stanford University.

Dunn, K.B., and J.J. McConnell. 1981. "Valuation of Mortgage-backed Securities," *The Journal of Finance*, 36: 599-617.

Evanston, D. D and Segal, L.M., 1996. "CRA and fair lending regulations: resulting trends in mortgage lending," *Economic Perspective* 20 (November/December): 19-46.

Follain, J. R., J. Ondrich, and G. P. Sinha. 1997. "Ruthless Prepayment? Evidence from Multifamily Mortgages," *Journal of Urban Economics*, 41, 78-101.

Foster, C., and R. Van Order. 1984. "An Option-based Model of Mortgage Default," *Housing Finance Review*, 3, 351-372.

Gabriel, S. A. and Rosenthal, S. S., 1991. "Credit rationing, race, and the mortgage market," *Journal of Urban Economics* 20(3): 371-379.

Green J., and J. Shoven. 1986. "The Effect of Interest Rates on Mortgage Prepayments." *Journal of Money, Credit and Banking* 18 (1): 41-59.

Heckman, J., and B. Singer. 1984. "A Method for Minimizing the Impact of Distributional Assumption in Econometric Models for Duration Data," *Econometrica* 52 (2): 271-320.

Holmes, A. and Horvitz, P., 1994. "Mortgage Redlining: Race, Risk and Demand," *Journal of Finance* 49 (March): 81-99.

Jarrow, R. A. and S. M. Turnbull. 2000. "The Intersection of Market and Credit Risk," *Journal of Banking and Finance* 24: 271-299.

Jegadeesh, N. and X. Ju. 2000. "A Non-Parametric Prepayment Model and Valuation of Mortgage-backed Securities," *Journal of Fixed Income* 10(2): 50-68.

Kalbfleisch, J.D., and R. L. Prentice. 2002. The Statistical Analysis of Failure Time Data. Second Edition, Willey Publications.

Kaplan, E. L., and P. Meier. 1958. "Non-parametric Estimation for Incomplete Observations," *Journal of the American Statistical Association* 53(282): 457-481.

Kau, J.B., D. C. Keenan, W. J. Muller, and J. F. Epperson. 1990. "Pricing Commercial Mortgages and their Mortgage-backed Securities," *The Journal of Real Estate Finance and Economics* 3(4), 333-356.

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

------ 1995 "The Valuation at Origination of Fixed-rate Mortgages with Default and Prepayment," *The Journal of Real Estate Finance and Economics* 11 (1): 5-36.

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

Kau, J.B., D. C. Keenan, and A. Simurov. 2006. "Reduced Form Mortgage Pricing as an Alternative to Option-Pricing Models," *The Journal of Real Estate Finance and Economics* 33 (3): 183-196.

Kau, J. B., D. C. Keenan, and H. Munneke. 2008. "Racial Discrimination and Mortgage Lending," Working paper, University of Georgia.

Klein, J. 1992. "Semiparametric Estimation of Random Effects Using the Cox Model Based on the EM Algorithm," *Biometrics* 48: 795-806.

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

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

Lang, W.W. and Nakamura, L.I., 1993. "A Model of Redlining," *Journal of Urban Economics* 33: 223-234.

Li, D. X. 2000. "On Default Correlation: a Copula Approach," *Journal of Fixed Income* 9: 43-54.

Loffler, G. 2003 "The Effects of Estimation Error on Measures of Portfolio Credit Risk," *Journal of Banking and Finance* 27: 1427-1253.

Lucas, D. J. 2004. "Default Correlation: From Definition to Proposed Solution," UBS CDO research working paper.

Mattey, J, and N. Wallace.1998. "Housing Prices and the (In)stability of Mortgage Prepayment Models: Evidence from California." Federal Reserve Bank of San Francisco Working Paper 98-05.

Mattey, J., and N. Wallace. 2001. "Housing Price Circles and Prepayment Rates of U.S. Mortgage Pools," *The Journal of Real Estate Finance and Economics* 23(2): 161-184.

Megbolugbe, I. F., and M. Cho. 1993. "An Empirical Analysis of Metropolitan Housing and Mortgage Markets," *Journal of Housing Research* 4(2): 191-224.

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

Nothaft, F. E. and V. G. Perry. 2002. "Do Mortgage Rates Vary by Neighborhood? Implications fro Loan Pricing and Redlining," *Journal of Housing Economics* 11: 244-265.

Gutierrez, R.G. 2002. Parametric Frailty and Shared Frailty Survival Models. *STATA Journal* 2, 22-44.

Pennington-Cross, A. 2003. "Credit History and the Performance of Prime and Non-prime Mortgages," *Journal of Real Estate Finance and Economics* 27(3): 279-301.

Quercia, R. G. and M. A. Stegman. 1992. "Residential Mortgage Default: A Review of the Literature," *Journal of Housing Research* 3: 341-379.

Quigley, J. M., and R. Van Order. 1991. "Defaults on Mortgage Obligations and Capital Requirements for U.S. Savings Institutions," *Journal of Public Economics* 44: 353-369.

Richard, S. F., and R. Roll. 1989. "Prepayments on Fixed-rate Mortgage-backed Securities," *Journal of Portfolio Management* 15: 73-82.

Schwartz, E. S., and W. N. Torous. 1992. "Prepayment and valuation of mortgage-backed securities," *The Journal of Finance* 44(2): 375-392.

Spahr, R. W. and M. A. Sunderman. 1992. "The Effect of Prepayment Modeling in Pricing Mortgage-backed Securities," *Journal of Housing Research* 3(2): 381-400.

Sastry, N. 1997. "Family-level Clustering of Childhood Mortality Risk in Northeast Brazil," *Population Studies* 51(3): 245-261.

Schill, M.H. and Wachter, S.M., 1993. "A Tale of Two Cities: Racial and Ethnic Geographic Disparities in Home Mortgage Lending in Boston and Philadelphia," *Journal of Housing Research* 4: 245-27.

Schill, M.H. and Wachter, S.M., 1994. "Borrower and Neighborhood Racial and Income Characteristics and Financial Institution Mortgage Application Screening," *Journal of Real Estate Finance and Economics* 9 (November): 223-239.

Schwartz, E. S., and W. N. Torous. 1989. "Prepayment and Valuation of Mortgage-backed Securities," *The Journal of Finance* 44(2), 375-392.

Stanton, R., and N. Wallace. 1998. "Mortgage Choice: What is the Point?" *Real Estate Economics* 26: 173-205.

Survival Analysis and Epidemiological Tables Reference Manual. STATA Press.

Titman, S., and W.N. Torous 1989. “Valuing Commercial Mortgages: An Empirical Investigation of the Contingent-claims Approach to Pricing Risky Debt,” *The Journal of Finance* 44(2): 345-373.

Therneau, T. M. and P. M. Grambsch. 2000. Modeling Survival Data: Extending the Cox Model. Springer.

Zhou, C. 1997. “Default Correlation: an Analytical result”, FEDS paper 1997–27, Federal Reserve Board.