Reduced-Form Mortgage Valuation with Stochastic Home Prices

by

Shannon D. Clark

(Under the direction of James B. Kau)

Abstract

In recent years, the reduced-form approach to valuation has become widely used in asset pricing. Unlike earlier structural models, reduced-form models do not require that data on the underlying asset be available, which makes them a convenient tool for empirical applications.

I develop a simple reduced-form model of mortgage pricing where both default and prepayment are exogenous, and empirically estimate it using historical data on about one half million fixed-rate residential mortgages. The empirical analysis proceeds in several stages. First, individual mortgage histories are used to estimate effects of exogenous variables, those being loan-specific characteristics, on hazards of termination. In the second stage, particle filtering is employed to estimate stochastic hazard processes as well as a correlated house price process. Once the parameters of hazard processes, as well as those of the conventional risk-free terms structure, are obtained, I use calibration to convert physical processes into risk-neutral ones. After that, tests of pricing performance are conducted.

Index words: Mortgage, Asset pricing, Prepayment, Default

Reduced-Form Mortgage Valuation with Stochastic Home Prices

by

Shannon D. Clark

B.B.A., Augusta State University, 2002

M.S.M.F., University of North Carolina at Chalotte, 2004

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

2010

Reduced-Form Mortgage Valuation with Stochastic Home Prices

by

Shannon D. Clark

Approved:

Major Professor: James B. Kau

Committee: Henry J. Munneke
Richard Martin

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
December 2010

Reduced-Form Mortgage Valuation with Stochastic Home Prices

Shannon D. Clark

December 6, 2010

Contents

List of Figures

Chapter 1

INTRODUCTION

With few exceptions, the structural approach to option pricing has been the predominate way previous studies have built models for mortgage valuation and option pricing. The structural approach, proposed by Merton [78], treats default as a call option. Effectively, both prepayment and default are embedded options in the mortgage. Thus, a mortgage is simply a fancy bond with embedded options that need to be priced.

In structural models, both prepayment and default are determined endogenously. They are the result of decisions that are being made by borrowers based on the value of the underlying options. To correctly value these underlying options, the underlying asset (the house) must be valued. This constitutes a difficult challenge for the structural approach, or any approach for that matter due to the lack of relaible housing data[60].

Reduced-form models has its origins in some of the same literature that created the structural approach. However, in the reduced-form framework, default and prepayment enter the model exogenously. In other words, default and prepayment always come as a surpise to the model. Borrowers may be exercising their options optimally, or sub-optimally.

Also, the reduced-form model is extremely flexible when it comes to exogeneous variables that are likely to play a major role in termination. Namely, the current level of interest rates and house price appreciation. These are able to be used without comprimising the general pricing theory of taking the risk-neutral expectation of all future cash flows.

My goal in this dissertation is to introduce a reduced-form model similar to those introduced by Duffie and Singleton [40] or Lando [71][72].[1] These models allows for several types of termination, with my focus on two that are the most important, default and prepayment. In addition to exogenously determined terminations on a data set of fixed-rate mortgages, I will incorporate a two factor stochastic interest rate in the CIR framework and stochastic home price evolution that is correlated with those interest rates.[2]

Both the interest rate process and the house price process should be driving factors in termination. The interest rate process will model the spread in the market influencing prepayment, whereas the house price model will help model the price of the default option. The house prices model are at the county level providing very rich data influencing borrowers decisions to default.

Once I have built and adequetly described the model, I will estimate the parameters of the model empirically. Even though the mortgages used in the analysis have originations ranging from 1975 to 2002, the observation window is from 1995 to 2002. The data contains information on termination as well several key loan specific characteristics that will aid in the model's estimation. The data also contains macroeconomic variables such as evolution of t-bills and treasury bonds for the interest rate process and fip specific house price indices for the house price process.

I will unite several methods from mortgage termination literature. By dividing the data into county specific strata, I will enable the double stochasticity of the model. Not only the event will be unknown, but also the probability of the event occuring is also unknown.

[1] While these models were applied to corporate debt, the reduced-form structure is the same. For a thorough analysis, see Madan and Unal[74], Duffie[38], Duffie and Singleton[40], and Bielecki and Rutkowski[8]

[2] While these are fixed rate mortgages, similar analysis can be found on adjustable rate mortgages in Buser, Hendershott, and Sanders[19], Cox, Ingersoll, and Ross[30], Kau et al.[61][64][66], Schwartz and Torous[93].

By making use of the stochastic baselines, I allow idiosyncratic risk of the borrowers to influence the actual number of terminations that occur. Once this is accomplished, I am able to estimate the market's assessment of risk by implicitly including multiplicative and additive risk parameters in the valuation equation.[3]

The dissertation is organized as follows. Chapter 2 summarizes existing theoretical and empirical literature on mortgage pricing. Chapter 3 describes the theoretical pricing model used for empirical estimation. Chapters 4 to 7 analyze the empirical analysis involved. In Chapter 4, I estimate a mortgage duration model using the discrete stratified partial likelihood approach. Individual mortgage duration data are used to obtain estimates of the effects of exogenous variables, such as original loan-to-value ratio, loan size, etc., on loan termination. In Chapter 5, I build a two-factor extended CIR process for the estimation of the term structure. Chapter 6 describes the house price process that will be used. Chapter 7 provides the description of the termination processes that will be used in simulation. Finally, Chapter 8 provides results from the calibration procedure

[3] It should be stated that the general model that follows is directly the result of previous work don by Kau and Keenan.

Chapter 2

MORTGAGE PRICING: A LITERATURE REVIEW

In this chapter, I provide a short review of relevant literature related to mortgage valuation, mortgage termination, and house price evolution. The first section analyzes papers that have used the structural approach to mortgage pricing. The second looks at attempts to to use the reduced-form approach. The final sections looks at multiple approaches to mortgage valuation and how house prices were incorporated into their analysis. Refences to previous works related to specific techniques or individual steps used in this analysis will be included in the relavant chapter.

2.1 THE STRUCTURAL APPROACH

The structural approach, sometimes called the option-based approach, has been used in the literature for some time. It is based on the work of Black and Scholes[10] and Merton[78]. Both prepayment and default are embedded options in the mortgage instrument. Thus, the structural approach attempts to incorporate optimal behavior on behalf of the borrower in the exercising of those options. I should point out that this actually means "financially" optimal.

For default, the option is most like a European put option due to specific times in which one would default[1]. For prepayment, the story is a little different. It is most like an American-style call option. It could be optimal to prepay between payment dates, unlike the default option. What has caused a lot authors grief is that these two options are intertangled. They represent competing termination states. Due to the difficulty, both from data and estimation, early works would include only one option.

The earliest works included the prepayment call option, but not the default option. Examples of such models with only a call option present include Dunn and McConnell [41][42], Buser and Hendershott [18] and Brennan and Schwartz [14]. Eventually, an attempt was made to model the default put option in Cunningham and Hendershott[32], although it did not have a stochastic term strucure. This is an unrealistic assumption on securities with average duration longer than 1 year. Epperson[43] did finally add stochastic interest rates, but still with only the default option. Finally, Kau et al. [62], Leung and Sirmans [73], and Kau et al. [65] developed models that incorporated the joint option of prepayment and default.

I draw certain concepts from each of the works listed above. Hense, further details are necessary to get a complete picture of the evolution of mortgage valuation. In the Buser and Hendershott[18] model, they introduce taxes and refinancing cost into their prepayment model. Refinancing cost are an important hurdle that influence the speed of prepayment and payments are a taxable event to the lender. Both are important factors in a mortgage valuation model. Other examples include Hall[53], Pozdena and Iben[85]. Schwartz and Torous[92] value MBSs using the now famous Brennan and Schwartz term-strucure model.

[1] A borrower would always choose to default on payment dates rather than in between payments.

Their major addition, though, is that they included a deterministic prepayment function where the interest rates enter as a covariate. This had rarely been done previously.

Epperson et al.[43] introduced the idea of modeling default as a compound European put option and modeled it in a stochastic interest rate environment. Kau et al. [62] use the same approach to build a pricing model that is applied to Freddie Mac's multifamily participation certificates backed by commercial mortgages. The same study is extended by Kau et al. [63], who provide more detailed numerical results of the same model.

Titman and Torous[98] build a model of commercial mortgages. The main focus in commercial mortgages is default, so prepayment is prohibited. The reason for their inclusion here is the other state variables included in their model. They include a square-root interest rate diffusion and a lognormal house price process. They model both as correlated processes. While the parameters of the interest rate process are estimated, the house price parameters are set by the authors.[2]

Leung and Sirmans [73] model default and prepayment options in a discrete-time framework using a lattice approach. Overall, their results are similar to those of Kau et al. [65]. This is important because of the discreteness of observed data in mortgage duration. Kau et al.[65] developed what is now considered a complete option-theoretic model of residential mortgage pricing. This paper is likely the reason for the lack of other works employing the strucural approach in mortgage pricing in recent history. The work assumed that borrowers had the joint right to prepay or default. In addition to the financially optimal stopping times, they also have "non-financial termination". Kau et al.[67] also includes the joint option, but without the "sub-optimal" prepayment or default.

Kau and Keenan[60] provide a discussion of option-theoretical work on mortgage pricing. While not speaking to specific papers, their emphasis is on the curent state of literature

[2]I follow a similar model in my estimation of house prices. However, I empirically estimate the parameters in both processes.

in mortgage pricing along with possible extensions of the general framework. A study by Chatterjee et al.[24] compares different option-pricing models.

2.2 THE REDUCED-FORM APPROACH

The reduced-form approach differs from the strucutal approach in that default and prepayment are not options, but stochastic processes varying over time. These processes may or may not be linked to the underlying asset. This enables the termination to be determined exogenously so it comes as a "surprise" to the model.

In a general reduced-form model, the valuation equation for a defaultable zero coupon bond that promises to pay \$1 at time T can be expressed as

$$X(r,t) = E^{\mathbb{Q}}[(\mathbb{1}_{\{\tau>T\}} + W(\tau)\mathbb{1}_{\{\tau\leq T\}})\mathbf{e}^{-\int_t^T r(s)ds}] \tag{2.1}$$

where τ is the stopping time, $W(\tau)$ is the recovery on default, and $\mathbb{1}$ is the usual indicator function. Particular attention could be focused on W, which could be stochastic or a constant. In the event there is no default, the bondholder gets the originally promissed payment of one dollar. Its analogy to mortgages is clear.

Reduced-Form mortgage valuation has received a lot of attention over the last few years and is becoming increasingly popular, not only in mortgage duration literature, but in several fields.

Reduced-Form modeling can be further broken down into different formulations of 2.1. Those being the hazard, the intensity, and the compensator of the process. Most literature to date has used the intensity-based approach. For example, Duffie and Singleton [39] [40], Lando [71] [72], Duffee[37], or Gorovoy and Linetsky[51] have used this approach. Hazard-based models are becoming more common as well such as Schönbucher [91] and Bélanger et al. [6]. Even with the explosion of reduced-form literature, compensator-based modelling is still rather uncommon. Notable exceptions are Giesecke[46] [47], Kau et al.[68], and Kau, Keenan, and Smurov[69].

Lando[72] provides a complete discussion of doubly stochastic Poisson processes in the bond-pricing literature. This is similar to work by Jarrow[57]. Both of these works are important because of the doubly stochastic nature of their models. Lando[71] also provided literature on possible connections between the reduced-form models and stuctural ones. He noted that the price of the underlying security (the house), which is the most important variable in a structural model, can easily be incorporated in a reduced-form model.[3]

Empirical studies of reduced-form modelling such as Duffie and Singleton[39] and Duffie[40] model defaultable securities using intensity based approaches and incorporate such things as liquidity and taxes. Duffie, in particular, was able to isolate a substantial credit risk component, as well as a liquidity premium component.[4]

A final paper built in the intensity-based reduced-form model framework that is important to this work is Driessen[36]. The model incorporates a two-factor risk free term structure, as well as firm specific effects. The parameters are estimated using an extended Kalman Filter.[5] In the paper, they expore the validity of the conditional diversification hypothesis.[6] Their empirical results provide evidence against the diversification hypothesis since their estimated value for their jump adjustment parameter is around six. Kau, Keenan, and Smurov[69] estimate this parameter to be around nine.

---

[3]This is the case in this analysis.

[4]An interesting note about Duffie's work is that it allows the possibility of negative default rates. In fact, some of the estimated values are negative.

[5]This is similar to my estimation technique. I use a more general particle filter.

[6]This states that default jumps can be diversified away and are, therefore, not priced by the market.

2.3 HOUSE PRICE EVOLUTION

There have been numerous papers written on the evolution of house prices. I will restrict my review to those of particular importance to the valuation process described in this paper.

House prices are a major part of the valuation process. The default option is directly affected by the house price. Therefore, proper inclusion and evolution of house prices are paramount. Downing, Stanton, and Wallace[35], and Kau et al.[65] and Stanton[94] all model the evolution of house prices incorporating interest rates and service flow. Their formulation is as follows:

$$dh_t = (r_t - q_H)H_t dt + \phi_H H_t dW_{H,t} \tag{2.2}$$

where r_t is the instantaneous interest rate and q_H is the estimated value of service flow. This is the standard model for modelling home prices in the mortgage pricing and default literature.[7]

In works such as Mattey and Wallace[75], they argue that if your goal is to model termination patterns, then including some evolution of the dynamics of home prices and how they will affect terminations is likely to improve results. In Krainer et al.[70], they explicitly model service flow and then model the price of housing as the expectation of all future service flow discounted to the present.

In Kau, Keenan, and Smurov[69], house price are held constant during the observation period. However, they were working with fixed rate mortgages over a period of increasing house prices, so this is less of a concern when estimating default.

Titman and Torous[98] built a model of commercial mortgages that included an evolution of the price of the underlying asset similar to 2.2. However, they model the asset price as

[7] See Deng, Quigley, and VanOrder[33], Downing, Stanton, and Wallace[35], and Kau and Keenan[60].

being correlated with interest rates. Their analysis provides strong arguments that interest rates should be included in the evolution of the price of the underlying asset.

In general, the way house prices have been treated in the literature has been relatively consistent. They need to be included in any analysis of mortgage termination or mortgage valuation. Also it is clear that interest rates play some role in house price evolution and need to be included.

Chapter 3

THE PRICING MODEL

We start with a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Denote by $\{\mathcal{F}_t : t \geq 0\}$ a filtration of σ-algebras satisfying the usual conditions.[1] The instantaneous interest rate r (the short rate) is modeled as a positive continuous process adapted to $\mathcal{F}$ so that it will be $\mathcal{F}_t$-measurable. This is done out of necessity and simplicity.

The objective of this chapter is to describe the valuation process, in particular, the process of valuing a fixed-rate mortgage security capable of both prepayment and default. To track information over time, assume the security pays $\mathbb{1}_{\{\tau<t(i)\}}M$ at $t(i) > 0$, where τ_d is the time of default and τ_p is the time of prepayment, $\tau = \tau_d \wedge \tau_p \equiv \min(\tau_p, \tau_d)$, $\mathbb{1}_{\tau>t(i)}$=1 if $\tau > t(i)$, $\mathbb{1}_{\tau>t(i)}$=0 if $\tau \leq t(i)$, and M is the resulting cash flow from the associated termination. The two stopping times, τ_p,τ_d are also assumed to be non-negative, bounded and $\mathcal{F}_t$-measurable.

The two termination processes, default and prepayment, will be modeled as two "latent" stochastic baseline hazard processes ${\lambda_0}^d(t)$ and ${\lambda_0}^p(t)$. Due to the obvious need for the non-negativity of both of these processes, they will be modeled in the Cox-Ingersoll-Ross framework with time-varying trends. This is done because both prepayment and default exhibit strong differences in their mean reverting values over time.

[1]For a more rigorous exposition, see, for example, Brémaud [13].

$$d\lambda_0^\ell(t) = \kappa_\ell(\theta_\ell(t) - \lambda_0^\ell(t))dt + \sigma_\ell\sqrt{\lambda_0^\ell(t)}dz_\ell^{\mathbb{P}} \qquad \ell = d, p \tag{3.1}$$

The trends will be discussed, in detail, in a later chapter. The white noise processes are assumed to be independent here for simplicity. The most obvious choice for observation of these processes is once per month, or equivalently, on the payment date. At each of these dates, following Prentice and Kalbfleisch[58], the process yields the "realized" hazard:

$$\lambda_i^\ell = \mathbf{e}^{\beta_\ell' \mathbf{x}(t(i))}\lambda_0^\ell(t(i)) \qquad \ell = d, p \tag{3.2}$$

where $\mathbf{x}(t(i))$ represents the vector of exogenous variables affecting termination and β is the vector of coefficients to be estimated. It is important to have the same variables for both the prepayment and default hazard to model them in a competing risk environment.

Since the absorbing states of this model is only observable once per month, the "intensity" that is viewed is not well defined. To this end, assume a stopping time $\tau = \tau_d \wedge \tau_p \equiv \min(\tau_p, \tau_d)$, an indicator process $\mathcal{H}$, (a one-jump process equal to zero before termination and jumping to 1 at τ), a filtration $\mathcal{D}$=$\mathcal{D}_t, t \geq 0$ generated by $\mathcal{H}$, and an enlarged filtration $\mathbb{G}$=$\mathcal{G}_t, t \geq 0, \mathcal{G}_t = \mathcal{F}_t \wedge \mathcal{D}_t$. Then the compensated process $\mathcal{H}$ is a $(\mathbb{G}, \mathbb{Q})$-martingale. Schönbucher [91] and Bélanger et al. [6] develop the theory of reduced-form default in terms of hazard processes rather that intensities, while Giesecke[47] has introduced the terminological distinction between intensity-based and compensator-based default modeling.

The termination processes are still under the real probability measure $\mathbb{P}$. Through the Girsanov theorem and both additive and multiplicative risk adjustments, there is an equivalent martingale measure $\mathbb{Q}$.

$$dz_\ell^{\mathbb{Q}} = dz_\ell^{\mathbb{P}} - \nu_t^\ell dt \tag{3.3}$$

$$d\Lambda_\ell^{\mathbb{Q}} = \mu_t^\ell d\Lambda_\ell^{\mathbb{P}} \qquad \ell = d, p \tag{3.4}$$

(3.3) is the usual drift adjustments and (3.4) is the multiplicative adjustments for jump risk. I take ν_t to be of the form $\nu\sqrt{\lambda_0(t)}/\sigma$, leading to the usual drift adjustment $-\nu\lambda_0(t)dt$ for $d\lambda_0(t)$. Since both the prepayment and default event carry a risk premium, I take the $\mathbb{Q}$-intensity, for some constant μ, to be equal to $\lambda^{\mathbb{Q}} = \Sigma_i(\mu\lambda_i) = \mu\Sigma_i\lambda_i = \mu\lambda^{\mathbb{P}}$.[2] This multiplicative constant is to compensate for the risk premium of default and prepayment. Jarrow et al shows that if this event can be diversified away, then μ=1.

The term structure that will be used in the valuation equation will be a two-additive-factor extended CIR so that under the risk-neutral measure $\mathbb{Q}$, it takes the form

$$dy_\imath(t) = (\kappa_\imath(\theta_\imath - y_\imath(t)) - \nu_\imath y_\imath(t))dt + \sigma_\imath\sqrt{y_\imath(t)}dz_\imath^{\mathbb{Q}} \qquad \imath = 1, 2 \tag{3.5}$$

$$r = y_1 + y_2 + x \tag{3.6}$$

where the x is a deterministic shift to the two-factor process, therefore making the process "extended". This extension is desired for many reasons, mainly for analytical tractability.[16]

Using the stopping times $\tau = \tau_d \wedge \tau_p \equiv \min(\tau_p, \tau_d)$ with the associated martingale hazard process and the usual balancing equation for mortgages,

$$V(t(0))/L - (1 - \delta) = 0. \tag{3.7}$$

[2]See Bjork et al. [9] and Schönbucher [91], who use the cited Girsanov result, though the former works within an intensity framework.

where L is the loan amount, the valuation equation becomes

$$V(t(0)) = E^{\mathbb{Q}}_{t(0)}\Bigg[\sum_{i=1}^{I} \mathrm{e}^{-\int_{t(0)}^{t(i)}((1-\tau_F)r(s)+\ell)ds}\Bigg(\prod_{j=1}^{i-1}(1-\lambda_j^d-\lambda_j^p)\Bigg)$$
$$\Bigg(\lambda_i^d D(i) + \lambda_i^p P(i) + (1-\lambda_i^d-\lambda_i^p)M\Bigg)\Bigg] \quad I = 360 \tag{3.8}$$

where D(i) is the recovery value upon default, P(i) is the tax-adjusted principal unpaid balance due at prepayment, M(i) is the tax-adjusted mortgage payment indicating that no stopping has been reached, ℓ is the liquidity premium, and T is the federal tax-rate.

In the following chapters, I describe the estimation procedure and give results at each stage, concentrating the analysis on the estimation of the termination processes and the estimation of the multiplicative and additive risk adjustments used in the valuation equation.

Chapter 4

ESTIMATION OF COVARIATE EFFECTS USING A DISCRETE STRATIFIED PROPORTIONAL HAZARD MODEL

In this chapter, I focus on the effects of exogenous variables on the hazards of both prepayment and default. A proportional hazard model was a logical choice for the estimation of these parameters. The proportional hazard imposes no structure on the underlying baselines which is important for the doubly stochastic nature of the model.[29] Even though no explicit structure is specified, it allows one to estimate the effects of the covariates on the termination hazards.

The proportional hazard also allows many conveniences in the estimation procedure. The most important is the capability to handle discretely observed data. Prentice and Kalbfleisch [58] show that in this discrete time setting, the Breslow estimator of covariate effects is consistent, though the standard errors must be adjusted. This is necessary because the observed data appears as ties which is not allowed in the pure form of the proportional hazard. The Breslow estimator also allows for more efficient estimation of the covariates, which is important in large data set such as the one used in this analysis.

Secondly, the phreg procedure allows time-dependent parameters to be used almost effortlessly. These time-dependent covariates are included, in the form of the interest rate spread and mark-to-market loan to value ratio. Both of these are extremely important to the termination processes. Due to the inclusion of time-dependent parameters, the hazards are not proportional any more. However, it has become standard to refer to this model as a proportional hazard even though it is not. I will follow that tradition as well.

Third, stratification is very simple in a proportional hazard framework. This has a few additional effects. Stratification requires an additional stratum set up for every possible value of the categorical variable. If there are many categories the data is stratified by, it is easy to end up with some strata that may only have one observation. Therefore, choice of strata is important in that sense. Also, for large data sets, stratification can lead to large computational gains.

Fourth, the hazard framework easily allows for entering and exiting the risk set. Indeed, there are several issues such as left truncation, right censoring, and uninformative censoring that has no effect on the consistency of the estimates. In fact, as in most types of this analysis, I convert calendar time into mortgage time. So, while care is taken to know what mortgages are at risk at any point in calendar time, I am able to explore the effects of mortgage time, or seasoning, on the mortgage.

Finally, I model default and prepayment as competing events. This is done by modeling each separately with each having its own hazard. In other words, default and prepayment are conditionally independent of each other.

4.1 MODEL SPECIFICATION

Assume the complete probability space, the indicator process $\mathbb{H}$, and the enlarged filtration $\mathbb{G}$ described in chapter 3.

Also, assume the covariate process X and the censoring process Y, where Y takes the value of one if an individual is at risk at time t, and zero otherwise.[1] Following Prentice and Kalbfleisch[58], I define the multiplicative hazard intensity model[2] as

$$\lambda_i(dt) = \lambda_0(dt)\mathbf{e}^{\beta' \mathbf{x}_i(t)} \tag{4.1}$$

where $\lambda_0(dt) = \lambda_0(t) - \lambda_0(t^-)$ if t is a mass point of the failure distribution in the discrete case.

As previously mentioned, the Breslow approximate partial likelihood estimator

$$L(\beta) = \prod_{i=1}^{n} \frac{\mathbf{e}^{\beta' \mathbf{s_i}(t)}}{\left[\sum_{l \in R_i(t)} \mathbf{e}^{\beta' \mathbf{x}_l(t)}\right]^{d_i}} \tag{4.2}$$

is a logical choice for handling tied data while still providing consistent estimates of the covariates.[15] Adjustments will need to be made to the covariance matrix. Use of the Breslow approximation allows us to still have consistent estimates while accounting for time-lapse between observations.

The hazard of prepayment and default are derived by multiplying the covariate effects by its associated baseline. While prepayment and default are modeled as competing events, they can be estimated separately due to the assumption of conditional independence.[3]

[1]Here, the censoring process is assumed to be an independent process. For details, see Anderson et al.[4]

[2]The original multiplicative intensity model was developed by Aalen[2]

[3]Partial likelihood models of prepayment can be found in Green and Shoven [52], Quigley and Van Order [87], as well as Fu et al. [45]. Examples of partial likelihood models of default include Harmon [55], Vandell et al. [99], as well as Quigley and Van Order [88]

4.2 STRATIFICATION

The stratification component turns out be crucial to my specification of the model. The baselines of both prepayment and default are modeled as stochastic processes. This enables the doubly-stochastic nature of the model. In fact, each stratum represents another realization of the underlying processes. Therefore, all mortgages in each stratum share a common baseline.[4] It would be preferable if each mortgage had its own realization of the process, but not only is this not possible given current data and computational resources, it would cause estimation problems in the proportional hazard. In fact, given how granular the stratification is, particular care had to be given to ensure that no stratum contained only one mortgage.

Given that the house price process described in chapter 6 is realized by fip codes, it seems logical to stratify the hazard process by fip codes as well. Therefore, from each county level stratum, one draw of a realized house process, a default termination baseline, and a prepayment termination baseline will emerge. The termination baselines will be seen explicitly in the valuation formula, while the house price process will be imbedded as a covariate in the proportional hazard model.

An attempt was made to keep the covariates in the hazard model specific to the individual mortgages themselves. This was done because every macro-economic variable must also then be modeled into the future for any valuation procedure to work correctly. However, the interest rate spread between the contract rate of the mortgage and rate in the market is likely to be a major driver of prepayment, just as whether a specific borrower has negative equity (meaning their default option is in the money) is likely to be a major driver of default. Luckily, these are easily accounted for with the CIR interest rate model described in the next chapter and the house price model described in chapter 6.

[4]For a complete explanation of partial likelihoods via stratification, see Ridder and Tunali[90]

4.3 DATA

The data in this analysis consist of 578,165 fixed rate mortgages originated between 1975 and 2002. Table 4.1a contains summary statistics of all loans in the observation universe. Table 4.1b and table 4.1c contain the same statistics, but restricting the data to those loans that either prepaid or defaulted respectively. These loans were observed from January 1995 to March 2002 for purposes of estimation. Table 4.2 and Table 4.3 report the distributions of prepayment and default over the observation period.

As previously mentioned, the data is stratified by fip code. In this data set, that results in 12,434 stratums across the United States. Figure 4.1 gives a geographic count by fip codes. The data is widely dispersed across the United States ensuring that the data is geographically very rich. Figure 4.2 and figure 4.3 provide a similar graph, but with counts of prepayment and default.

During this observation period, 4,248 loans were defaulted and 306,291 were prepaid. 16,147 were prematurely terminated due to uninformative censoring such as third party sales, release of servicing rights, or some other unknown event. 250,438 reached the end of the observation period with no termination. 1,041 terminated normally. Under the counting style format for phreg, this results in 6,752,418 loan observation months.[5]

The house price data used in the covariates was obtained from Case-Schiller. It is in the form of quarterly data segmented by state and county fip code. In the event that there is not enough data in a particular fip, or strata, the state fip replaces the county fip. This allows data to be grouped into the smallest possible groupings, given data limitations. An explanation of fip is described in chapter 6.

[5] The original proportional hazard model built to explicitly deal with multiple reasons for leaving the risk set was proposed by Prentice and Gloecker [86], and later extended by Meyer [79], Narendranathan and Stewart [80]. It was further developed and applied to competing risks by Han and Hausman [54], Sueyoshi [96][97], and McCall [76][77].

4.4 CHOICE AND SPECIFICATION OF VARIABLES

In general, I follow literary convention for my choice of variables included in the estimation procedure. Several key variables have been identified as being key drivers of prepayment and default. For example, option-based theories have shown that the value of the underlying options to be integral in termination prediction. Those options take the form of the interest rate spread between the contract rate on the mortgage and prevailing rates in the market in the case of prepayment and the amount of negative equity minus any transaction cost in the case of default.

All of the other variables used in estimation are directly visible in the mortgage data. This is important to the valuation procedure described later in this paper. Any variables not directly found in the mortgage data will have to be modeled separately so that they enter the model at the appropriate time to shift the baseline. The remainder of this section will describe the covariate specification in detail. In all cases, both the covariate described and their quadratics are used in estimation.[6]

A helpful predictor, though rarely found in mortgage data, is the amount of points initially paid on the mortgage. While in a technical sense, points are defined as prepaid interest, they indicate intentions of the borrower. Positive points should be purchased by borrowers who expect to stay in the mortgage for an extended period of time. Therefore, points should have a negative impact on the prepayment hazard. Likewise, negative points, usually called credits by banks, should rapidly increase the hazard.[95]

Secondly, I use the original loan amount of the mortgage. The loan amount can have multiple effects making it unclear how this will affect termination probabilities. The loan

[6]It should be stated that there is nothing in this model that can account for unobserved heterogeneity. Deng et al.[33] and Follain et al.[44] provide evidence that estimated coefficents are adversely affected by its omission.

amount will likely enter as a rough proxy of credit quality given I do not have individual FICO scores. It could also affect prepayment by changing the incentive to prepay.

As already mentioned, the interest rate spread is traditionally an important driver of prepayment. The spread here is defined as the difference between the mortgage contract rate and the current 10-year rate in the market divided by the contract rate.[7] The 10-year rate is lagged by two months to allow for rates to permeate the market and allow borrowers time to refinance.

The final covariate is the mark-to-market loan to value ratio, MTMLTV. Updated house prices are via Case-Schiller home price index at the county fip level. Since 100 percent of the loans observed are fixed-rate mortgages, it is trivial to calculate scheduled balances after the origination date by simply reducing the original balance by the scheduled reduction in principal from payments. Ideally, one would want to use the unpaid principal on a loan which would differ from scheduled balances in the event of delinquent borrowers or possibly loan extensions due to financial hardship, but without a state transition model, scheduled balance is a close approximation. Once I have updated balances, a simple conversion will yield the updated ltv. Borrowers who are under water have an incentive to exercise their put option and default.

4.5 RESULTS

This section presents empirical results from estimating two conditionally independent risk of termination. This analysis uses the formulation by Prentice and Kalbfleisch which makes use of the Breslow estimator for discretely observed tied data. The data is stratified, by

[7] The formulation of the spread varies from paper to paper. Often, it is normalized by the contract rate (Calhoun and Deng[20]) or the current rate (Campbell and Dietrich[22]). Others simply specify a dependence between the two rates, such as the ratio (Richard and Roll[89], or Pavlov[81])

both origination quarter and county fip location. The covariates used in this analysis are the original loan size, spread, MTMLTV, and Points, and their quadratics.

The results of the prepayment and default termination models are presented in Table 4.2. I include the hazard estimates, their standard errors, and their corrected standard errors. Their corrected standard errors vary only slightly from their normal standard errors. This is largely due to the size of the data set used in this analysis. Figures 4.4 and 4.6 provide Kaplan-Meier estimates of both prepayment and default[59].

All the covariates are statistically significant except for Orig Loan Amount Squared in the default model, though their joint significance is high. Since the total effect of each covariate is the combined effect from both itself and its quadratic, it is important to look at each of these in tandem.

Figures 4.8 and 4.9 show the combined effect from each covariate, $\mathbf{e}^{\beta_k^{linear}\mathbf{x}_k+\beta_k^{quadratic}\mathbf{x}_k}$ for default and prepayment respectively. One can easily see the importance of including the quadratic terms as the effects on their hazards shows. Each covariate has effects that are direct and effects that are indirect.

Mark-to-Market LTV is the refreshed loan balance divided by the refreshed house value. Those that have negative equity, meaning they owe more than their house is worth, should be highly likely to default.[8] Figure 4.8(a) shows exactly this effect. The shape is consistent with current theory and is as expected. In fact, borrowers at 100 percent MTMLTV are about 6 times as likely to default as those borrowers at only 50 percent. If MTMLTV were around 120 percent, then this ratio reaches almost 15 times as likely to default. For prepayment, the effect is reversed and slightly muted. This was expected since those borrowers that are underwater still may have a prepayment option “in the money”, but the higher the MTMLTV, the higher the transaction cost of prepayment.

[8]Studies such as Ambrose and Sanders[3] suggest that high origination LTV might have a negative affect on default hazards.

Loan Size has many effects on termination. Since we do not have FICO information on borrowers, the loan size is a proxy for the credit quality of the borrower. Indeed, lower loan size results in a slightly higher default probability. After around 220,000 loan size, the default hazard appears to flatten and thus has no effect relatively speaking. For prepayment, there seems to be a bigger effect. As loan size increases, the hazard of prepayment increases up to around 280,000.

Points that are paid on a mortgage help to indicate the preferences of the borrower. Those who plan to stay in their mortgage for a long period will buy down their rate, given certain conditions, by paying positive points. Those that plan to stay for only a short time will typically have negative points. These negative points will usually show as a credit on the closing documents. Thus points should help to predict prepayment, or lack thereof, by borrowers. As 4.9(c) shows, the more negative the points, the greater the likelihood of a prepayment occuring. A borrower with one negative point is about 10 percent more likely to prepay than a borrower with 0 points, ceteris parabis. Coincidently, we see the same effect in default. When the points paid are positvie or 0, there is virtually no effect. However, when points are negative, there is a considerable shift in the default hazard.

The final covariate used in the analysis is the interest rate spread. This spread is modeled as the sum of the contract rate minus the 2 month lagged 10-year Treasury Bond rate divided by the contract rate. With this formulation of the spread, the spread has a ceiling of 1, indicating the rate in the market has went to 0. On the downside, there is no floor. Most of the mortgages during the observation period ranged from -0.5 to .5. This covariate should be the driver of the prepayment termination. Figure 4.9(d) confirms that this is the case. A borrower with a spread of 0.5, indicting rates have fallen to half of the original contract rate, is about 40 times as likely to prepay as someone with a spread value of 0. In the negative direction, rising rates have the opposite effect on prepayment. Once the spread reaches -0.15, prepayment rates are driven virtually to zero. This confirms my expectation of how this spread should effect prepayment and matches conventional theory. For most of the mortgages in this analysis, the spread effect was almost negligible.

TABLE 4.1

Panel A

Summary Statistics: All Loans

Variable	Mean	Std.Dev.	Min.	Max.
Original Loan Size ($)	111,750.00	63,779.00	20,000.00	452,000
Spread (%)	5.00	0.89	-111.23	62.75
MTMLTV (%)	68.16	17.33	0.25	129.75
Contract Rate (%)	7.77	0.890	3.75	13.88
Points (%)	0.01	0.67	-3.00	3.00

Panel B

Summary Statistics: Defaulted Loans

Variable	Mean	Std.Dev.	Min.	Max.
Original Loan Size ($)	102,652.00	56,016.00	20,000.00	450,000.00
Spread (%)	5.00	0.89	-111.23	42.00
MTMLTV (%)	83.30	12.70	5.21	127.32
Contract Rate (%)	8.52	1.18	3.75	13.50
Points (%)	-0.02	0.64	-3.00	3.00

Panel C

Summary Statistics: Prepaid Loans

Variable	Mean	Std.Dev.	Min.	Max.
Original Loan Size ($)	127,963.00	69,812	20,000.00	452,000.00
Spread (%)	5.00	0.89	-33.42	62.75
MTMLTV (%)	69.40	16.74	1.23	133.87
Contract Rate (%)	7.99	0.81	3.75	13.88
Points (%)	-0.02	0.71	-3.00	3.00

Descriptive statistics presented in Panel A are constructed using the entire sample of 578,165 loans. Descriptive statistics presented in Panels B and C are constructed using the sub-samples of 4,248 loans that defaulted during the period of 1995-2002 and 306,291 loans that were prepaid during the same period, respectively. All variables were obtained from the original mortgage databases.

TABLE 4.2

Panel A

Loan Origination, Default and Prepayment by Year of Origination

Year	# Originated	# Defaulted	% Defaulted	# Prepaid	% Prepaid
1975	1836	2	0.11	1235	67.27
1976	3384	4	0.12	2044	60.4
1977	6278	13	0.21	3634	57.88
1978	6262	26	0.42	3683	58.82
1979	3547	37	1.04	2110	59.49
1980	1227	19	1.55	697	56.81
1981	273	2	0.73	158	57.88
1982	196	4	2.04	86	43.88
1983	675	10	1.48	355	52.59
1984	562	9	1.6	333	59.25
1985	634	16	2.52	307	48.42
1986	4142	84	2.03	1966	47.46
1987	6178	113	1.83	2912	47.13
1988	2828	83	2.93	1553	54.92
1989	3373	181	5.37	1964	58.23
1990	4562	278	6.09	2835	62.14
1991	10621	372	3.5	7391	69.59
1992	38862	640	1.65	27767	71.45
1993	77159	556	0.72	44383	57.52
1994	36919	425	1.15	23016	62.34
1995	41571	440	1.06	26221	63.08
1996	27830	188	0.68	13094	47.05
1997	29885	86	0.29	13091	43.8
1998	90644	154	0.17	24033	26.51
1999	72958	415	0.57	40089	54.95
2000	70977	88	0.12	53528	75.42
2001	33715	3	0.01	7800	23.14
2002	1067	0	0	6	0.56
Total	578165	4248	0.73	306291	52.98

Panel B

Loan Termination, Default and Prepayment by Year of Termination

Year	# Terminated	# Defaulted	% Defaulted	# Prepaid	% Prepaid
1995	3708	89	2.4	3398	91.64
1996	16022	406	2.53	14515	90.59
1997	20312	696	3.43	19547	96.23
1998	63200	732	1.16	49953	79.04
1999	40506	667	1.65	39525	97.58
2000	28954	634	2.19	27990	96.67
2001	120758	801	0.66	117388	97.21
2002	34267	223	0.65	33975	99.15
Total	327727	4248	1.3	306291	93.46

Figure 4.1: **Geographic Distribution of Loan Originations**

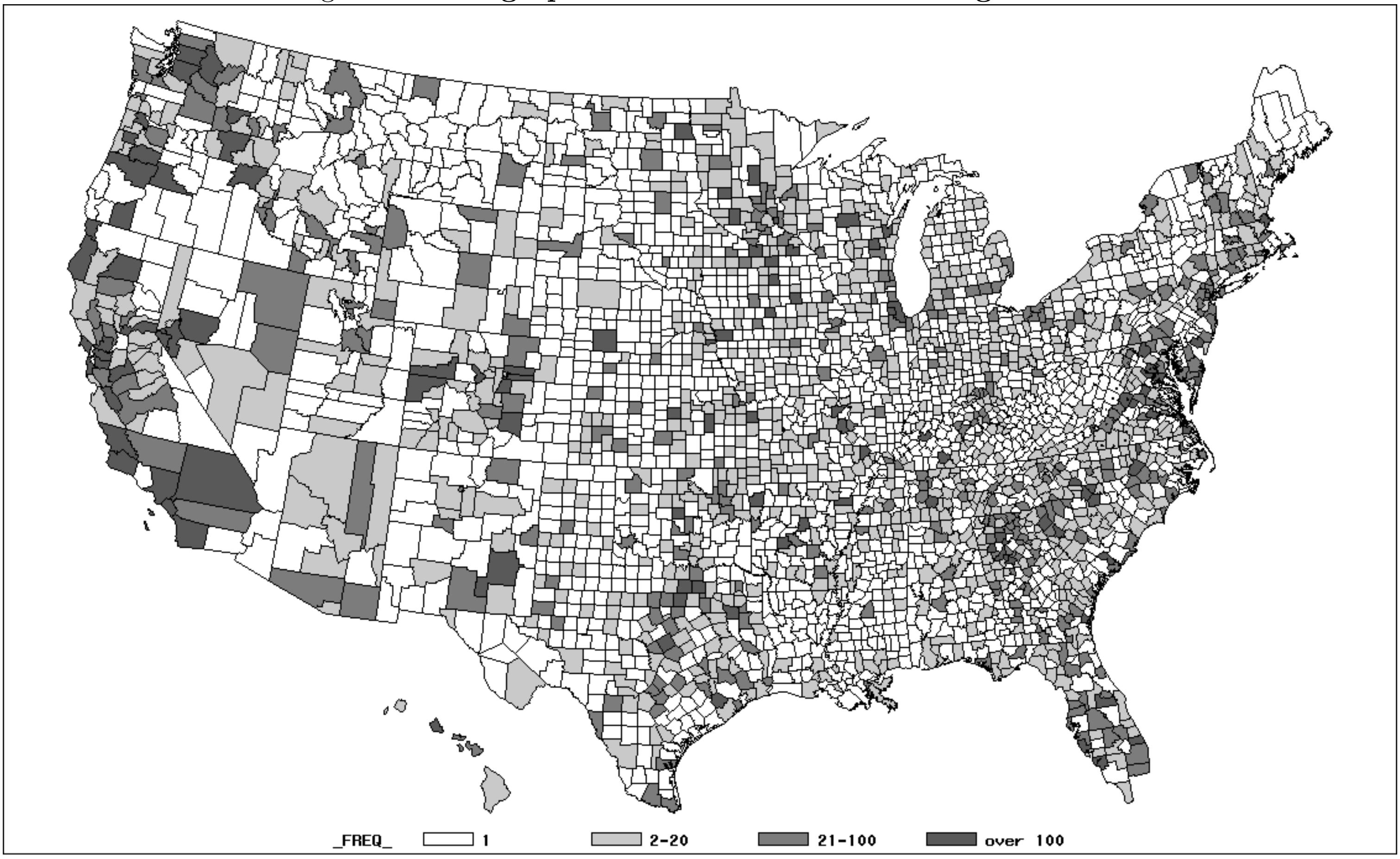


Figure 4.2: **Geographic Distribution of Defaulted Loans**

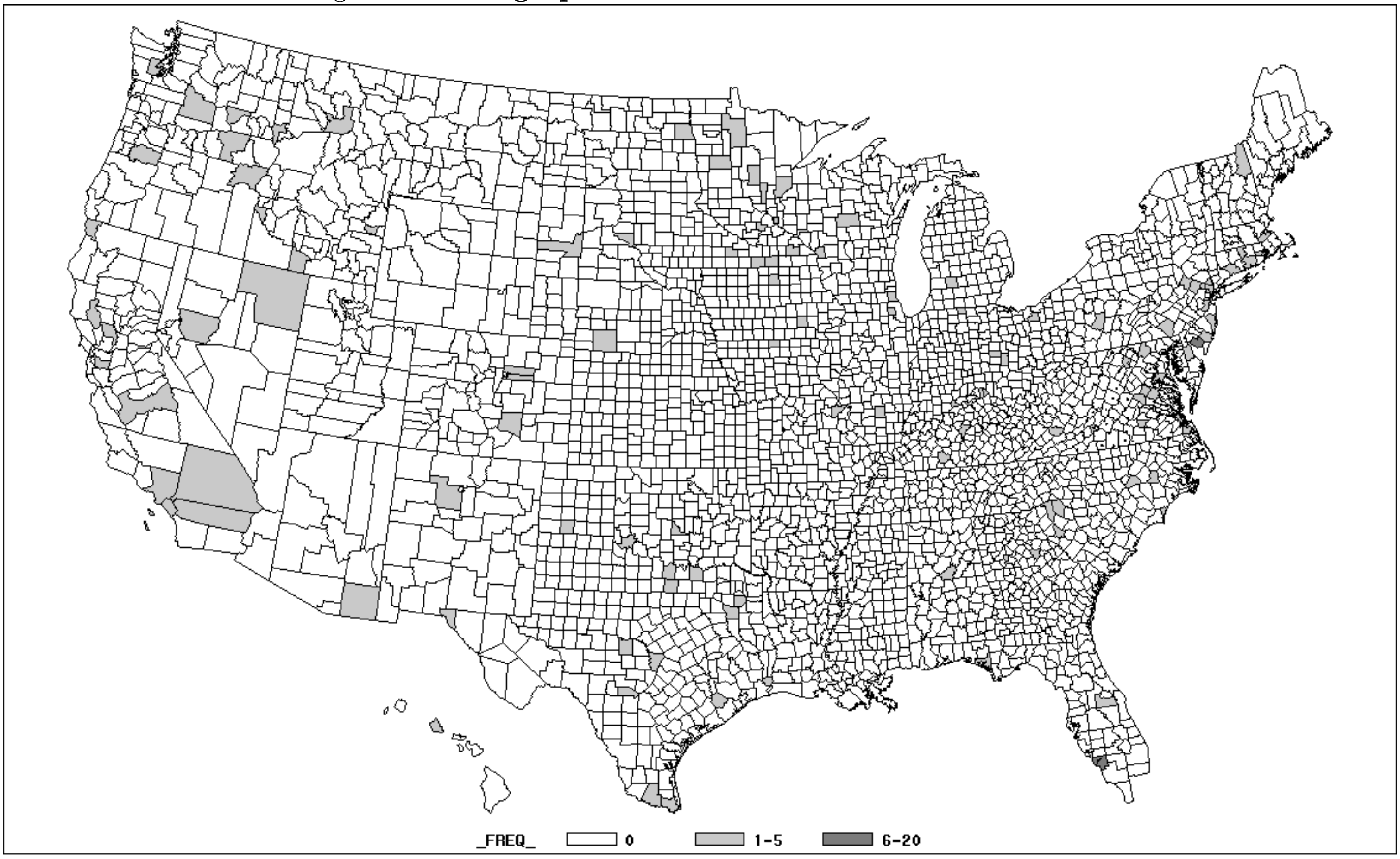


Figure 4.3: **Geographic Distribution of Prepaid Loans**

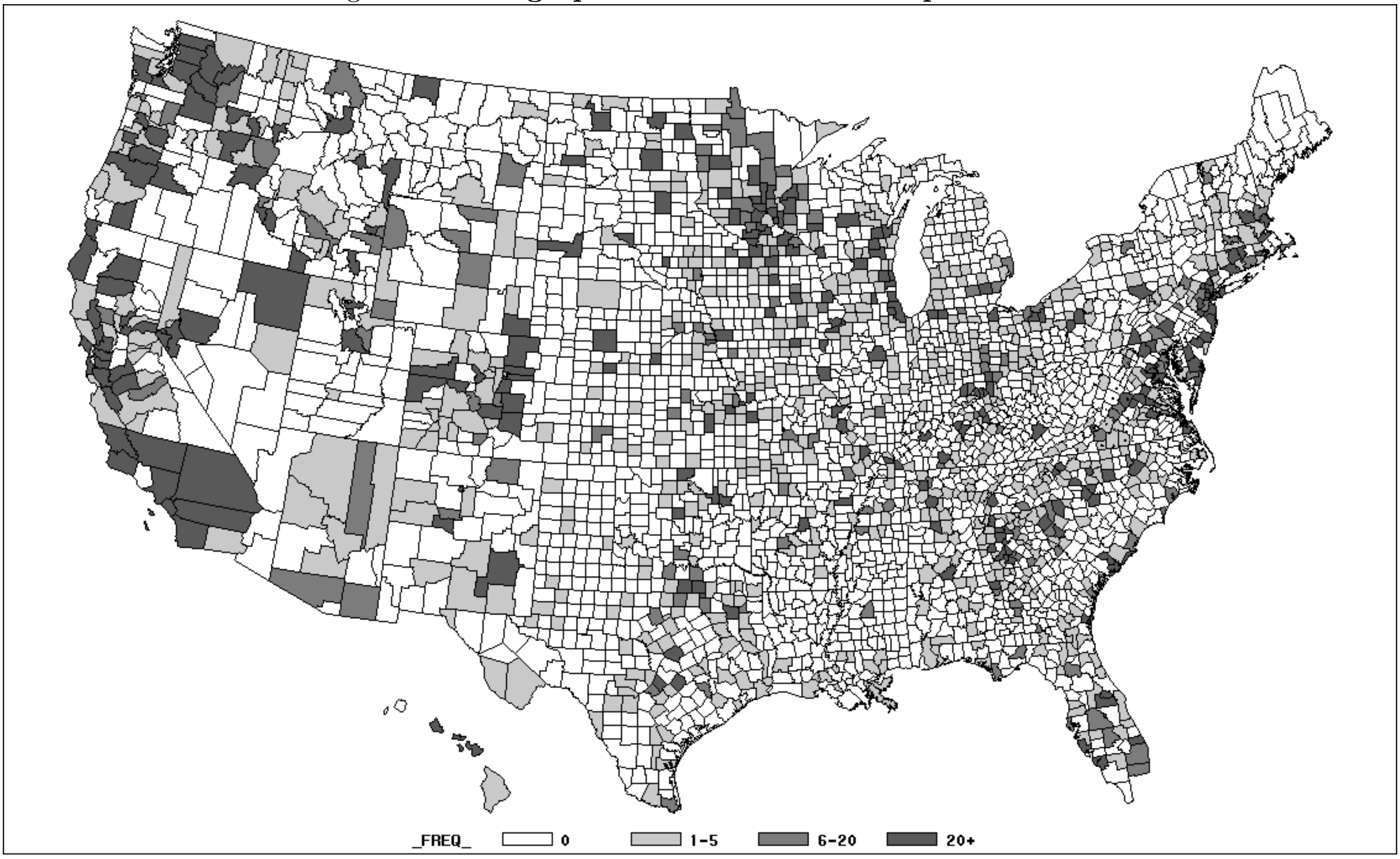


TABLE 4.2

Stratified Proportional Hazard Estimates for

Competing Risks of Default and Prepayment

Sample Size: N=578,165

	Default Model	Prepayment Model
Variable	**Estimate**	**Estimate**
	(Standard Error)	**(Standard Error)**
	(Corrected Standard Error)	**(Corrected Standard Error)**
MTM LTV	0.05268	-0.00032
	(0.00195)	8.93×10^{-5})
	(0.00194)	1.012×10^{-4}
MTM LTV Squared	-.0000538	-1.8969×10^{-6}
	6.2755×10^{-6}	(8.8474×10^{-8})
	6.2755×10^{-6}	(6.7578×10^{-6})
Orig Loan Size	-0.00283	0.01357
	(0.0010)	(0.00010)
	(0.0012)	(0.00011)
Orig Loan Size Squared	-5.0868×10^{-5}	-2.4100×10^{-5}
	(2.9530×10^{-6})	2.7596×10^{-7}
	(0.00372)	1.6356×10^{-6}
Points	-0.25587	-0.04489
	(0.03072)	(0.00306)
	(0.02789)	(0.00311)
Points Squared	0.08688	0.03579
	(0.01444)	(0.00153)
	(0.01421)	(0.00154)
Spread	2.34459	6.37792
	(0.70098)	(0.17148)
	(0.70215)	(0.17148)
Spread Squared	10.93795	-1.44745
	(1.26325)	(0.28265
	(1.26322)	(0.34222)

Note that the MTMLTV variable was formed by dividing the refreshed loan balance by the refreshed house value. The spread variable is the difference between the loan's contract rate and the current ten-year Treasury rate, as a percent of the contract rate. Both the MTMLTV and the spread are time-varying covariates. The first listed standard error is the conventional one for proportional hazard models, whereas the second includes a correction required by the Prentice and Kalbfleisch modification of the Cox proportional hazard model.

Figure 4.4: **Kaplan-Meier Estimates of Monthly Prepay Hazard**

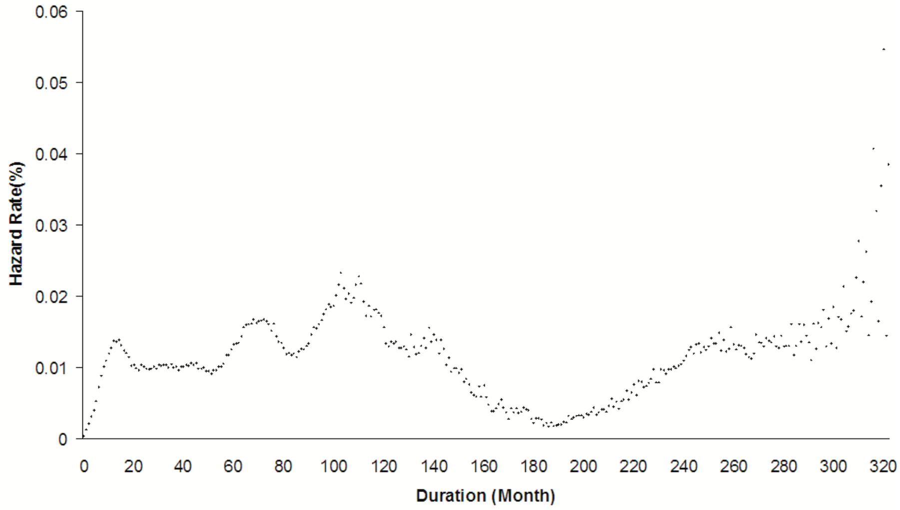
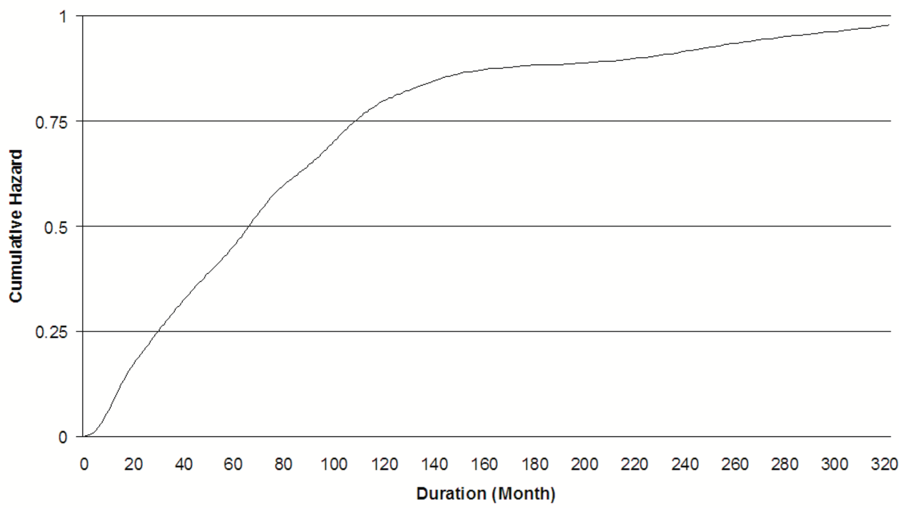


Figure 4.5: **Kaplan-Meier Estimates of Cumulative Prepayment Hazard**



The non-parametric Kaplan-Meier estimator (product limit estimator) is a strictly empirical approach to hazard function estimation. The estimator of the hazard of default(prepayment) is $\frac{\#\ of\ defaults(prepayments)\ at\ time\ i}{\#\ of\ observed\ loansat\ risk\ at\ time\ i}$, where $i = 1, 360$ is the month in mortgage life. For presentation purposes, the last 40 observations in the prepayment figure are truncated due to the very low number of surviving loans, resulting in extremely high hazard estimates.

Figure 4.6: **Kaplan-Meier Estimates of Monthly Default Hazard**

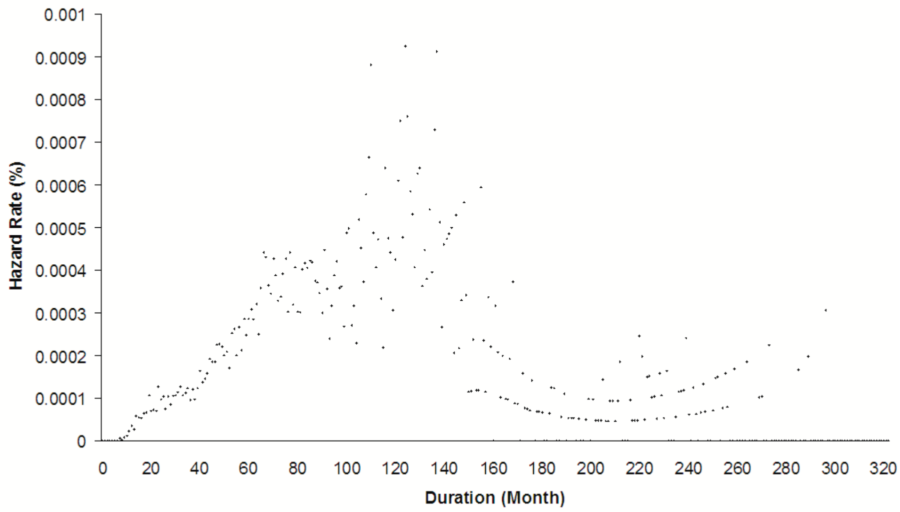


Figure 4.7: **Kaplan-Meier Estimates of Cumulative Default Hazard**

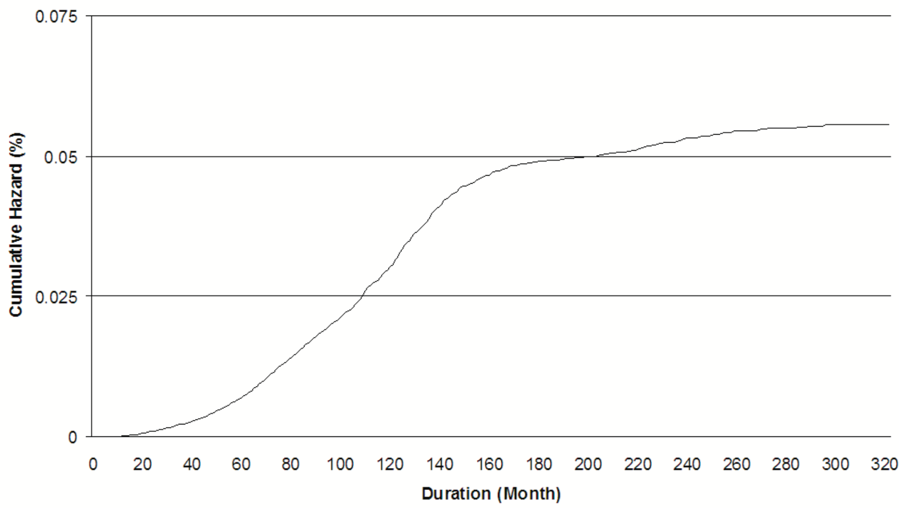
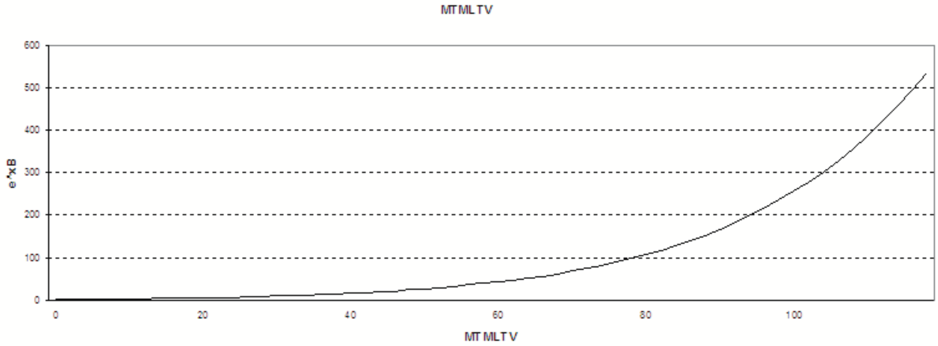
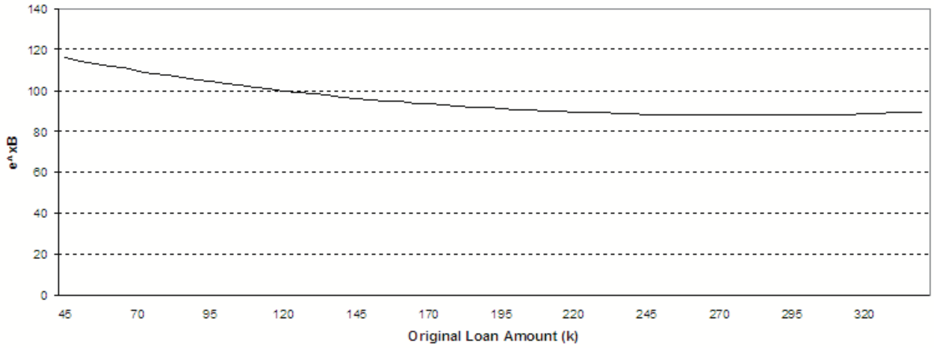


Figure 4.8: **Effects of Covariates on Hazard of Default**

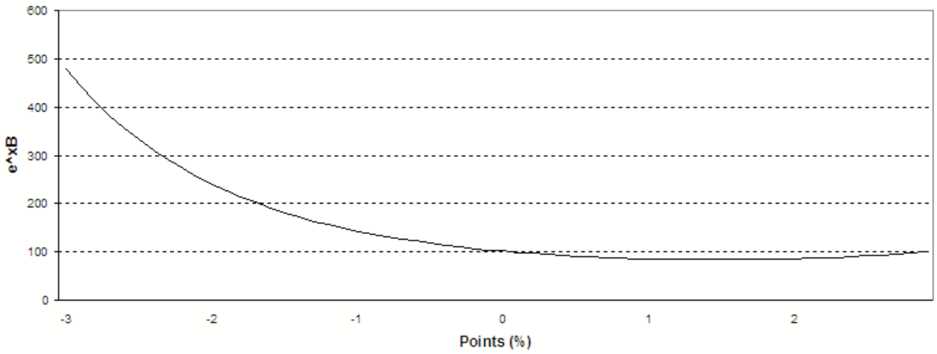
(a) Effect of MTMLTV on Hazard of Default



(b) Effect of Loan Size on Hazard of Default



(c) Effect of Points on Hazard of Default



(d) Effect of Interest Rate Spread on Hazard of Default

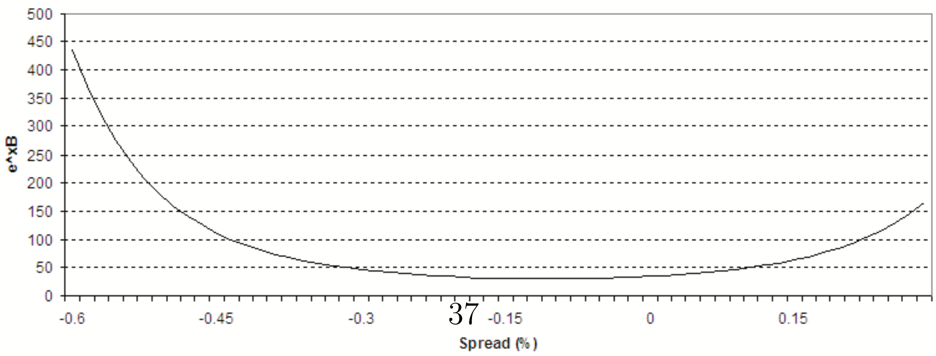
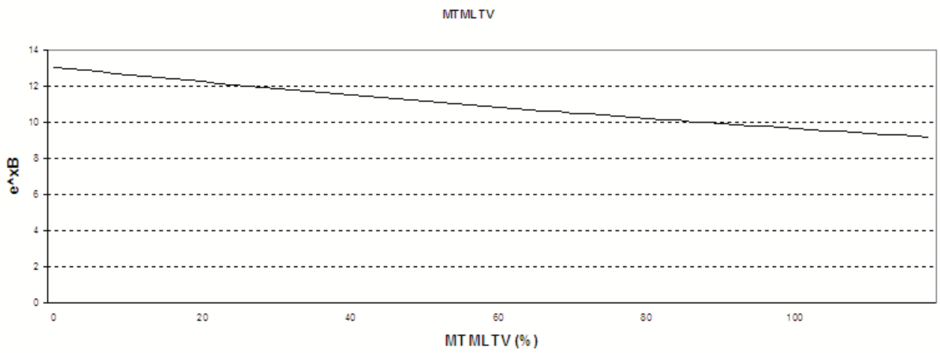
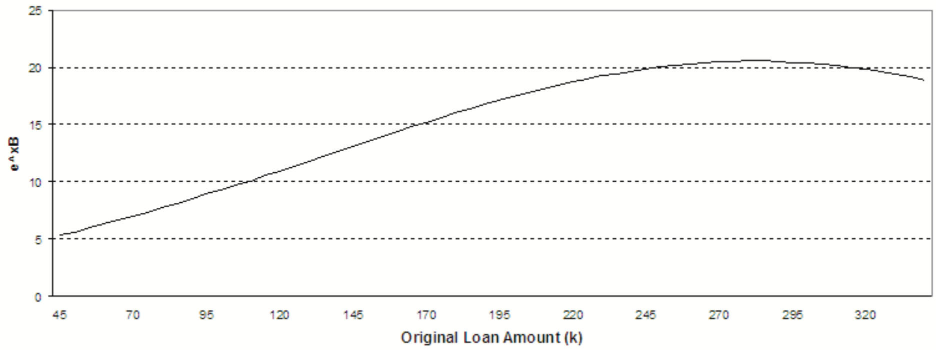


Figure 4.9: **Effects of Covariates on Hazard of Prepayment**

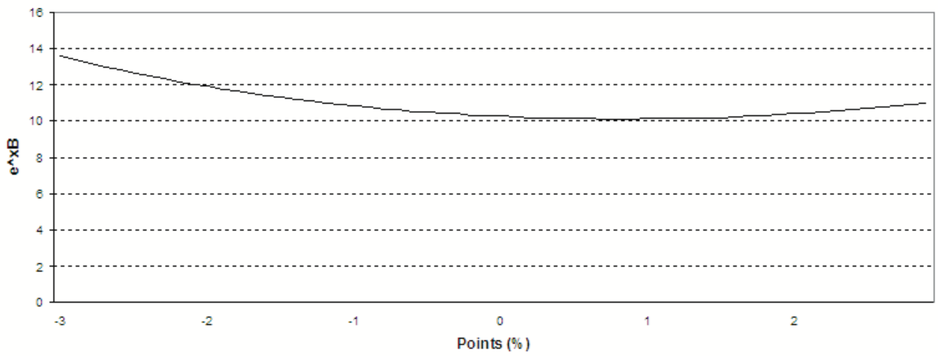
(a) Effect of MTMLTV on Hazard of Prepayment



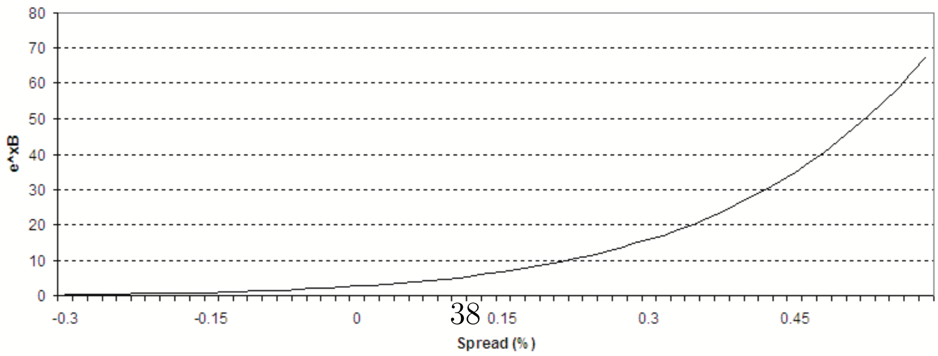
(b) Effect of Loan Size on Hazard of Prepayment



(c) Effect of Points on Hazard of Prepayment



(d) Effect of Interest Rate Spread on Hazard of Prepayment



Chapter 5

ESTIMATION OF A TWO-FACTOR EXTENDED CIR AFFINE TERM STRUCTURE MODEL OF INTEREST RATES

It is well known that the spread between the current mortgage contract rate and the prevailing rate in the market is a major driving factor of prepayment. Its affect on default is likely to be very muted comparatively speaking. I model a two factor extended CIR model of interest rates.[31] I will describe the model in detail and then describe the estimation procedure.[1]

[1] One of the first multi-factor models was developed by Pearson and Sun[82] who estimate a two factor CIR. Chen and Scott[26] developed a similar model, but introduced errors.

5.1 A STATE-SPACE MODEL OF INTEREST RATES WITH TWO INDEPENDENT FACTORS

Chen and Scott[28] have found that at least two factors must be used to adequately capture the dynamics of interest rates.[2] Interest rates themselves are not the primary focus of this paper, but mortgages are greatly influenced by their movements. Thus, I will model interest rates as a two factor extended CIR process.

Each factor will be modeled independently and will evolve according to its own diffusion process. I define the instantaneous spot rate, under the risk neutral measure, by

$$dy_\imath(t) = (\kappa_\imath(\theta_\imath - y_\imath(t)) - \nu_\imath y_\imath(t))dt + \sigma_\imath \sqrt{y_\imath(t)} dz_\imath^{\mathbb{Q}} \qquad \imath = 1, 2 \tag{5.1}$$

$$r = y_1 + y_2 + x \tag{5.2}$$

where y_1 and y_2 are the independent CIR processes and x is a deterministic shift. The shift, or extension, allows for exact fit of the term structure and analytical formulas for bond prices.[**?**]

At each point in time, a series of discount bonds with differing maturities are observed. With the normal bond price formulas, the state-space model will be fully specified. I then use the formulas to examine the likelihood and estimate the parameter vector:

$$\Omega = (\kappa_1, \kappa_2, \theta_1, \theta_2, \sigma_1, \sigma_2) \tag{5.3}$$

The data are obtained using the Bliss-Nelson-Siegel method[11], which fits the discount rate function R(s) directly to observed bond prices by using their approximating function:

[2] See Brunsun et al.[17] for extending the term structure to include more state variables. Chen and Scott show that the third factor only provides marginal improvement. They also used two factor CIR models in option pricing. See [25][27] for examples.

$$R(s) = \beta_0 + \beta_1 \left[\frac{1 - \mathbf{e}^{-s/\tau_1}}{s/\tau_1} \right] + \ \beta_2 \left[\frac{1 - \mathbf{e}^{-s/\tau_2}}{s/\tau_2} - \mathbf{e}^{-s/\tau_2} \right] \tag{5.4}$$

Using 5.4, I can obtain prices of riskless zero coupon bonds for the estimation procedure.

5.2 DATA AND RESULTS

The data used in the estimation of the two factor CIR model are an extensive set of treasury securities covering the period from January 1993 to March 2002, the observation period of the mortgages. Following Kau, Keenan, and Smurov[69], I use the 3-month and 10-year maturities. The 3-month Treasury bill will capture the shorter end of the yield curve for discounting purposes and the 10-year Treasury bond will help capture prepayment decisions. The 10-year bond will also be used in the formulation of the house price model in the next chapter.

Overall the results are consistent with literature and with previous similar works. The parameters are significant with the exception of the speed of mean-reversion on the short rate, though this is likely due to the short time period under estimation. While longer time periods could easily be used, it would contaminate the valuation procedure discussed later with data that these loans were never subject to.

Table 5.2 contains estimates of the parameters and their standard errors. The first factor can be thought of as a gauge of the general level of interest rate. The second factor is more of a guage of the spread of the interest rates used in this analysis. It will also be used as an input to the drift process described in chapter 6.

TABLE 5.1

Summary Statistics of Interest Rate Data

Observation Period: January 1995 - March 2002

Sample size: N = 87

Variable	**Mean**	**Median**	**Std.Dev.**	**Min.**	**Max.**
3-month Treasury Bill	0.049	0.050	0.013	0.017	0.064
10-year Treasury Bond	0.059	0.062	0.010	0.045	0.078

TABLE 5.2

Maximum Likelihood Estimates

of a Two-Factor Square-Root Model of Term Structure

Observation Period: January 1995 - March 2002

Sample size: N = 87

	First Factor		**Second Factor**
Variable	**Estimate**	**Estimate**	**Estimate**
	(Std. Error)	**(Std. Error)**	**(Std. Error)**
κ	0.1191		0.7714
	(0.1173)		(0.3890)
θ	0.0229		0.0492
	(0.0092)		(0.0180)
σ	0.095		0.1327
	(0.007)		(0.019)
ν	-0.036		-0.0825
	(0.0130)		(0.0326)
x		-0.0210	
		(0.0041)	
Log-Likelihood		1,143.03	

Estimates of Parameter Combinations for Asset Pricing

$\kappa\theta$	0.0027	0.03795
	(0.001)	(0.006)
$\kappa + \nu$	0.0831	0.6889
	(0.014)	(0.066)

As always, κ is the CIR reversion coefficient for the state variable and σ is the volatility parameter, while here, θ is the constant trend, ν is the additive risk adjustment, and x is a constant, which when added to the sum of the two state variables, yields the spot rate, and makes this an "extended" two-factor CIR model. Note that the half-life, $\ln 2/\kappa$, of the first factor is found to be 5.82 years, whereas the half-life of the second is 0.89 years.

TABLE 5.3

Correlation Coefficients

Variables	**Correlation**
First Factor and Spread	-0.617
Second Factor and Spread	-0.942
First Factor and 10-year Treasury Bond	0.969
Second Factor and 10-year Treasury Bond	-0.528

Chapter 6

ESTIMATION OF DIFFUSION PROCESS FOR MODEL OF HOUSE PRICE INDEX

I model home price appreciation as a Ornstein-Uhlenbeck process. I follow previous literature such as Downing, Stanton, and Wallace[35] with some slight modifications. The ease of which simulation can be achieved with this type of model made it very attractive since it will need to be modeled separetely and used in the Monte Carlo pricing formula. It is also important to have the same rates that are driving prepayment be the same rates that are driving the drift of the house price diffusion. This will be achieved through correlated diffusions of interest rates and house prices. The rest of the chapter is organized as follows. In the first section, I describe my estimation technique in detail. In section 2, I provide results from the estimation.

6.1 THE HOUSE PRICE MODEL

The price of the underlying asset (the house price) is assumed to follow the risk-adjusted diffusion

$$dH = (\mu_H - \lambda_H \sigma_H)dt + \sigma_H dz_H \tag{6.1}$$

and:

$$dr(t) = (\kappa(\theta - r(t)) - \nu r(t))dt + \sigma \sqrt{r(t)} dz_r^{\mathbb{Q}} \tag{6.2}$$

is the instantaneous interest rate, with $dz_H dz_r = \rho(r, H, t)dt$.[1]

Using the argument that the expected risk-adjusted return on a house is simply the risk-free rate,

$$(\mu_H - \lambda_H \sigma_H)/H + s = r$$

6.1 can be reformulated as

$$dH = (r_t - s(r, H, t))Hdt + \sigma_H dz_H \tag{6.3}$$

where s is the house rental rate (service flow).[2]

Since my goal is to model 6.1 and 6.2 as correlated processes, estimation and simulation of 6.3 requires some formulation. From the linear transformation property, I need only find a special formulation of the covariance matrix between the two processes. In particular, simulation of the correlated house price process reduces to finding a special matrix A for which $AA^T = \Sigma$ where Σ is the covariance matrix between the two observed processes. A

[1] In this analysis I, I hold $\rho(r, H, t) = \rho$ for simplicity.
[2] I also hold service flow to be a constant.

particular convienient choice for A is one that is lower triangular. This decomposition is known as the Cholesky decomposition. If the covariance matrix is specified as follows:

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{bmatrix}$$

and $\sigma_1 > 0$ and $\sigma_2 > 0$, the Cholesky factorization is

$$A = \begin{bmatrix} \sigma_1 & 0 \\ \sigma_2\rho & \sqrt{1-\rho^2}\sigma_2 \end{bmatrix}$$

The Cholesky decomposition is a square root matrix that when multiplied with itself yields the original correlation matrix of the two Wiener processes. This decomposition, A, multiplied by a random vector produces an additional vector of correlated random variables. Initial analysis must be on estimation of this matrix, though not of direct concern, and the vector of parameters $\Omega = (s, \rho, \sigma_H)$.

The parameters of 6.2 are known. In fact, 6.2 is one of the factors of the interest rate process estimated in chapter 5.[3] Its inclusion here has two important affects. First, it is important to have the rates that are driving house prices be the same rates, at least in part, that are driving interest rate spreads. Second, since the parameters of the interest rate process are already estimated, the procedure for estimating the house price process collapses to estimating the covariance matrix, Σ , the Cholesky decomposition, A and the drift of the house process, $\mu_H = (r - s(r, H, t))H$. For simplicity, I let $s_t = s$.[4]

Data for Home Prices are observed from the Case-Schiller House Price Index. House prices are indexed for quality and by the fip code in which they reside. Fip codes are a five-digit Federal Information Processing Standard (FIPS) that uniquely identifies state and

[3] In fact, it is the second factor. I chose this factor since it most closes resembles rates borrowers would expect to see in the market.

[4] It is straightforward to allow the service flow to be time-dependent. Since it is not of direct concern, I hold it as a constant as is very common in the literature.

county. These are not necessarily zip codes. In fact, in many cases, zip boundaries cross fip boundaries.

Estimation of A, and the parameter vector $\Omega = (s, \rho, \sigma_H)$ can be quite difficult due to the likelihood not generally being available for diffusions that are discretely observed. This is particularly problematic with house price indices that are only observed on a quarterly basis. Common filtering techniques rely on the parameters to be linear and gaussian. In most cases in this analysis, neither of those requirements are true. I, therefore, make use of the particle filter for estimation of parameters where there are not viable alternatives.

6.2 PARAMETER ESTIMATION BY PARTICLE FILTERING

Parameter estimation has long been an important and difficult issue in stochastic differential equations. Given certain conditions[5], it has been showed that regular kalman filtering is the optimal estimation technique. When those initial conditions are not true, or not known to be true, then other methods must be explored. The method I have chosen is the particle filter.[6]

Particle filters are a monte carlo method for model and parameter estimation.[7] They make use of the discretely observed data and create posterior distributions that are then used for estimation. This is accomplished by the use of a set of particles that start off equally weighted. As the procedure partially estimates some of the moments of the posterior

[5]Namely, that all the random varaible statistics are Gaussian. This breaks down in most real-world applications.

[6]Particle filter techniques are not limited to mortgage literature. Berzuini et al.[7] used particle filters in the medical profession. Gordon et al.[50][49] and Pitt and Shephard[83] used bayesian tracking on a navigation model.

[7]Particle filters are actually known by several different names. Bootstrap filtering (Gordon et al.[50]), sequential Monte Carlo, Bayesian tracking (Arulampalam et al.[5]), and the condensation algorithm (Izard and Blake[56]) are other names that are synonymous. The name particle filter is due to Capenter et al.[23]

distributions, some of the particles become “more important” and are re-weighted. This happens recursively until a full picture of the posterior distribution becomes visible and estimation of the parameter vector is possible.[8]

Particle filters have an edge over the normal filtering techniques since they are accurate to third-order approximations, whereas the kalman filter is only a first-order approximation. Doucet et al.[34] show that MSEs for particular styles of the particle filter are significantly smaller than those generated by other types of filters.

This procedure is not without its problems. It appears it will always suffer from a degeneracy problem, where all but a few particles become meaningless. This can be partially offset by resampling, but extreme care must be taken not to contaminate the existing posterior. The procedure is also very slow so care must be taken when deciding how many particles are optimal.

The dynamical system of a nonlinear, non-Gaussian is formulated as

$$x_k = f(x_{k-1}, u_{k-1}, v_{k-1}) \textit{ state equation} \tag{6.4}$$

$$y_k = h(x_k, u_k, v_k) \textit{ observation equation} \tag{6.5}$$

where the state equation is assumed to follow a first order Markov process and observations are assumed to be independent given the states. Recursively estimating the posterior distributions reveal the weights of the particles that will be used to estimate the parameter vector.

With the data from Case-Schiller, I estimate the parameters of the model. These parameters will be used when it is necessary to model the evolution of home prices in the valuation formula.

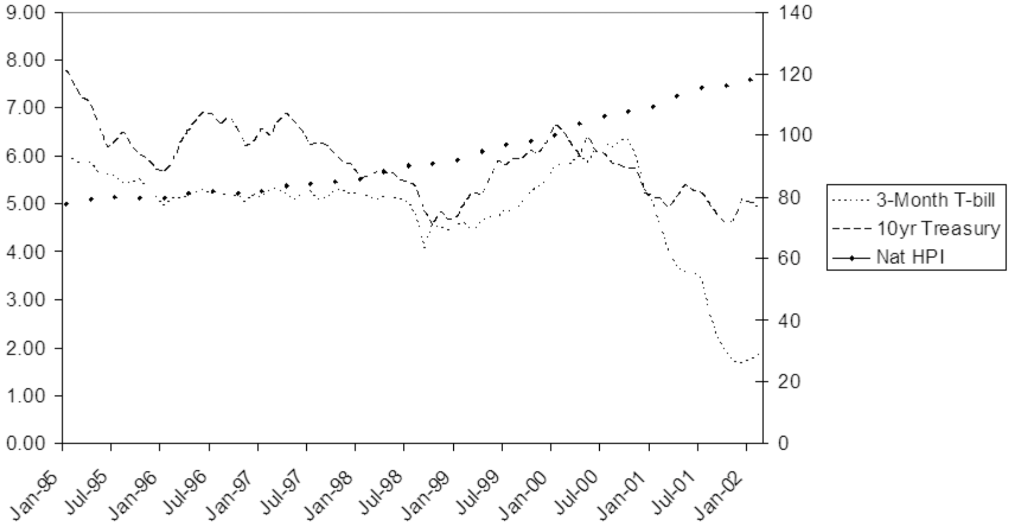
[8] A collection of different simulated models applicable to econometric time series can be found in Pitt[84]

6.3 RESULTS

The data used in this estimation is from Case-Schiller's Home Price Index. Figure 6.1 shows the national data along with interest rates over the particular observation period for the mortgages. As expected, interest rates are negatively correlated with house prices. Though, this analysis does not cover the most recent house price crisis that is combined with abnormally low interest rates. Estimation of service flow, s in the drift term, is 2.96 %. The average service flow estimates from BLS is 3.2 % over the same period.

Results are listed in table 6.1. The variable of most importance here is ρ which allows me to generate correlated diffusion processes during the valuation procedure. The particle filter estimate for ρ is -.3203. For comparison, Titman and Torous[98] found there correlation to fluctuate between -0.1 and -0.3. The estimate of the volatility of house prices, σ_H, is likely to understate the "true" volatility of house prices. This is because estimates of these parameters are built on an index that should attenuate the "true" volatility. My estimate of the volitility is 8.77%. While this is similar to many other estimates of volatility in the literature, it should not be taken as completely accurate since these all suffer from the same attenuation problem. Interest rates and house prices will evolve according to 6.1 and 6.2, respectively with each level of stratification representing one "realized" path.

Figure 6.1: **National HPI Index**



<u>TABLE 6.1</u>

Particle Filter Estimates for the House Price Diffusion Process

Sample Size: N=11,985

Propagation Method: Transition Density and Euler Discretization

Monthly Time Subintervals: $1/\Delta = 30$

Variable	**Estimate (Std. Error)**
s	0.0296
	(0.0204)
ρ	-.3203
	(0.00567)
σ_H	0.0877
	(0.018)
Log-Likelihood	-11,112.97

s here is the estimated value of the service flow parameter, ρ is the estimated correlation between the two stochastic processes, and σ_H is the estimated house price variance

Chapter 7

ESTIMATION OF DIFFUSION PROCESS FOR PREPAYMENT AND DEFAULT HAZARD

The termination processes for prepayment and default will enter the valuation model exogenously. This enables the model to be doubly stochastic. In a technical sense, the house price process is another level of stochasticity. However, as previously described, it is embedded in the hazard covariates described in chapter 4. Both termination processes will be modeled as square-root diffusions in the CIR framework. The need for both hazards to produce only positive values was paramount.

The rest of this chapter is organized as follows. In section 1, I describe the models of the baseline termination process. In section 2, I describe the estimation technique of the vector of parameters of both processes. In the final section, I describe and discuss the results.

7.1 THE STATE-SPACE TERMINATION PROCESS

As discussed in chapter 4, the baselines for the Cox Proportional Hazard will be stochastic. In each quarter and each fip zone, a different realization of the baseline termination process will be drawn. Due to the importance of the non-negativity of the realized baselines and the desire to have them mean-revert, the CIR process seemed the most applicable. Thus, the specifications of the termination baselines are as follows:

$$d\lambda_0^p(t) = \kappa^p(\theta^p(t) - \lambda_0^p(t))dt + \sigma^p\sqrt{\lambda_0^p(t)}dW(t) \tag{7.1}$$

$$d\lambda_0^d(t) = \kappa^d(\theta^d(t) - \lambda_0^d(t))dt + \sigma^d\sqrt{\lambda_0^d(t)}dW(t) \tag{7.2}$$

where $\theta^p(t)$ and $\theta^d(t)$ are time varying mean-reverting levels. These functions are critically important as these are fundamental to every subsequent step.

At each point in time, the baseline functions are not visible. What is known, however, is the number of mortgages that either prepaid or defaulted. The hazards estimated in chapter 4 are assumed to influence the termination baselines in the following multiplicative manner.

$$\lambda_i^\ell = \mathbf{e}^{\beta_\ell' \mathbf{x}(t(i))}\lambda_0^\ell(t(i)) \qquad \ell = d, p \tag{7.3}$$

where x's are mortgage specific variables and the B are the covariates estimated from those variables.

The number of mortgages that prepay or default in a specific month can then be approximated by use of a Poisson distribution with means equal to the hazard rates. If the termination of an individual mortgage is an independent event, the number of prepayments and defaults is then binomially distributed with probability equal to the hazard rate. When the probability of an event occurring is small relative to the mortgage universe, the Poisson distribution will provide a good approximation to the binomial distribution.

$$Pr(N_p = k) = e^{\lambda_t^p} \frac{(\lambda_t^p)^k}{k!} \tag{7.4}$$

$$Pr(N_d = k) = e^{\lambda_t^d} \frac{(\lambda_t^d)^k}{k!} \tag{7.5}$$

where N_p and N_d are the number of prepayments and defaults respectively.

7.2 ESTIMATION OF PARAMETERS OF CIR TERMINATION PROCESS

In chapter 4, a proportional hazard was estimated with stratification based on the origination quarter and the origination fip. All mortgages in the same strata are assumed to share one realization of the underlying baseline. The mortgages in separate strata are assumed to have a different realization of the underlying baseline, but from the same diffusion process.

In each strata, I can now formulate a log-likelihood of observing N_p prepayments and N_d defaults at time t in each strata. These can be summed over time and over strata to get a total log-likelihood during the life of the mortgages. This is formulated in the following total log-likelihoods.

$$log f_p = \sum_{k=1}^{K} \sum_{t=1}^{T_k} log f_p(N_{t,k}^p) \tag{7.6}$$

$$log f_d = \sum_{k=1}^{K} \sum_{t=1}^{T_k} log f_d(N_{t,k}^d) \tag{7.7}$$

where T_k is the max duration in months for stratum k.

All mortgages are pooled into 12,434 strata.[1] Those strata with unusually low number of loans are then changed from their county fip code to their state fip code. By doing this, if it is possible to estimate at the county level, then there is a baseline drawn for the fip and origination quarter. By rolling it up to a state fip code, all counties in a particular state with low numbers of mortgages can be grouped together and then a baseline can be drawn at the state level. This logic also attaches the correct house price index to those mortgages, whether they are at the county or state level.

In each stratum, the number of prepayments and defaults at time t and the corresponding covariate terms are calculated. With the indicator functions of prepayment and default, I can estimate an intensity hazard at each time t.

The CIR processes used in the termination processes are similar to those used in previous chapters. However, here we take the mean reverting values to be time-varying. In particular, I follow Kau, Keenan, And Smurov[69] and define the trends to be deterministic functions of time with only the need to estimate the particular parameters of the functions.[2]

In chapter 4, I estimated the time series of the baseline hazard functions for both prepayment and default. These baseline hazards act like a constant in a normal regression at any particular time t. In effect, they are means at any particular time t without the effect of any of the covariates. I fit the mean-reverting functions to this time series and thus, identify the specification for the full termination process.

Following Tong and Kau, Keenan, and Smurov[69], the functional forms I use for the baseline are a 6th degree polynomial function for prepayment and an un-normalized chi-

[1]While this is true in a technical sense, there are several strata that contain too few mortgages for any estimation procedure to be viable. Therefore, any strata that contains fewer than 200 mortgages will roll up to their state counterpart. Even though I have house price data at their local fip level, I use the state level HPI Index instead to enable all low-level fips in the same state to be merged into one larger strata. In fact, there are only 685 strata used in the final procedure.

[2]It has become standard in the literature to model prepayment as polynomial function and model default as an un-normalized chi-square.

square for default. These represent the typical functional forms of prepayment and default found in the literature and in industry.

$$\theta^p(t) = \sum_{k=1}^{6} \alpha_i t^k \qquad - \text{ Polynomial} \tag{7.8}$$

$$\theta^d(t) = \frac{\alpha_d}{\Gamma(\rho_d)2^{\rho_d}} t^{\rho_d - 1} e^{-t/2} \qquad - \text{ Unnormalized chi} - \text{square} \tag{7.9}$$

Using the particle filtering technique, I estimate the coefficients and drift of the diffusion processes including the parameters of the mean-reverting functions 7.8 and 7.9. Thus, the set of parameters for each termination are:

Prepayment Hazard: - $\kappa_p, \sigma_p, \alpha_0, ...\alpha_6$

Default Hazard: - $\kappa_d, \sigma_d, \alpha_d, \rho$

When dealing with hazard rates that are close to zero, particularly with default, it is important to monitor levels as they relate to the multiplicative form of the baseline. Indeed, it became necessary to factor the estimates in the particle filtering technique. For prepayment, the estimates are scaled up by 1.0×10^5. For default, no scaling was necessary.

Tables 7.3 and 7.4 present useful comparisons of predicted counts of defaults and prepayments to their actual levels. The Counts are grouped into count buckets to be able to perform a pearson statistic.[21] The fifth column in each represents each buckets contribution to the pearson statistic. From looking at table 7.3, the model performs well overall, but does miss on particular high counts, namely the 51-60 bucket. It is easy to see that for a default model to be able to pass this test, it need to be able to predict lower counts of default quite accurately. Since prepayment is a much more common event, the buckets for Table 7.4 are much wider. Even though, prepayment does not pass the Cameron-Trivedi test, it performs remarkable well. Note that the major flaw in the prepayment predictions occurs in the second bucket of 501-1000. In fact, 67% of the test statistic is contributed by this bucket. At first glance, it may appear that this prediction is not that bad. However, when you account for the massive number of loans that are prepaying, the discrepency is quite large.[3] In general, I am pleased with the performance of the prepayment model. It performs remarkably well, even though it did not pass the Cameron-Trivedi test that is known to be very strict.

[3]The reason this contribution is so large is the number of loans is a factor in this staistic, therefore, prepayment is particular sensitive to errors.

TABLE 7.1

Particle Filter Estimates

of a Diffusion Process for Prepayment

Observation Period: January 1995 - March 2002

Sample size: N = 87

Variable		**Estimate (Std. Error)**
κ		0.7875
		(0.3890)
σ		1.8791
		(0.0180)
α_1		6.1436×10^{-2}
		(0.019)
α_2		1.1001×10^{-3}
		(0.0526)
α_3		-2.6376×10^{-5}
		(0.0041)
α_4		1.7358×10^{-7}
		2.875×10^{-8}
α_5		-4.7674×10^{-10}
		1.9897×10^{-10}
α_6		4.8242×10^{-13}
		3.0165×10^{-13}
Log-Likelihood	2,487.65	

κ is the usual speed of reversion. σ is the volatility of the termination process and $\alpha_1 - \alpha_6$ are the parameters of the time-varying trend.

<u>TABLE 7.2</u>

Particle Filter Estimates

of a Diffusion Process for Default

Observation Period: January 1995 - March 2002

Sample size: N = 87

Variable		**Estimate (Std. Error)**
κ		1.1006
		(0.2590)
σ		1.4456
		(0.1189)
α_d		1.1618×10^{-4}
		2.5512×10^{-5}
ρ		.1254
		(0.0366)
Log-Likelihood	2,092,47	

κ is the usual speed of reversion, σ is the volatility of the termination process, α_d and ρ are the parameters of the unnormalized chi-square.

TABLE 7.3

Tests of Model Predictions

Default Model

Panel A

Predicted vs. Actual Probabilities

Counts	Actual	Predicted	|Diff.|	Pearson
0-10	0.6656	0.6832	0.0176	1.9328
11-20	0.0557	0.0495	0.0062	3.2879
21-30	0.0619	0.0560	0.0060	2.6963
31-40	0.0805	0.0814	0.0009	0.0465
41-50	0.0774	0.0772	0.0002	0.0025
51-60	0.0495	0.0431	0.0064	4.0665
> 61	0.0093	0.0000	0.0007	0.2154

Panel B

CM Tests

Degrees of Freedom: Q = 7

Covariance Used	Test Statistic	P-Value
Newey-West	12.05	0.099

The predicted probability of a termination count is its combined probability over cohorts and months, using one-step particle-filter forecasts and the now estimated parameters. This prediction is then compared to the actual percent of occasions that terminations for some cohort and month was, indeed, of that count, with the difference yielding the unstandardized residual (|Diff|). The final column calculates each category's contribution to the total Pearson statistic P, which is then compared to the degrees of freedom $Q = N - K$, with a lower value of the statistic taken to be an indication of underdispersion. The second table presents the Conditional Moment (CM) tests of the overall model fit. Under the null hypothesis of an adequate fit, the test statistic has a chi-square distribution with the number of degrees of freedom equal to the number of count cells specified.

TABLE 7.4

Tests of Model Predictions

Prepayment Model

Panel A

Predicted vs. Actual Probabilities

Counts	**Actual**	**Predicted**	**|Diff.|**	**Pearson**
0-500	0.6533	0.6520	0.0012	0.6986
501-1000	0.0310	0.0327	0.0017	26.3021
1001-1500	0.0774	0.0771	0.0003	0.4598
1501-2000	0.0279	0.0277	0.0001	0.2008
2001-2500	0.0681	0.0683	0.0002	0.2367
2501-3000	0.0372	0.0381	0.0010	7.2508
>3000	0.1053	0.1041	0.0012	3.9576

Panel B

CM Tests

Degrees of Freedom: Q = 7

Covariance Used		
Newey-West	39.1063	0.000

The predicted probability of a termination count is its combined probability over cohorts and months, using one-step particle-filter forecasts and the now estimated parameters. This prediction is then compared to the actual percent of occasions that terminations for some cohort and month was, indeed, of that count, with the difference yielding the unstandardized residual (|Diff|). The final column calculates each category's contribution to the total Pearson statistic P, which is then compared to the degrees of freedom $Q = N - K$, with a lower value of the statistic taken to be an indication of underdispersion. The second table presents the Conditional Moment (CM) tests of the overall model fit. Under the null hypothesis of an adequate fit, the test statistic has a chi-square distribution with the number of degrees of freedom equal to the number of count cells specified.

Chapter 8

CALIBRATION OF RISK ADJUSTMENT PARAMETERS AND PRICING

8.1 THE MONTE CARLO CALIBRATION MODEL

Excluding transaction cost and assuming an arbitrage free environment, every mortgage must adhere to the following balancing equation:

$$V(t_0) = L(1 - \delta) \tag{8.1}$$

where V is the value of the mortgage at origination, L is the loan amount, and δ is the amount points taken on the loan.

Given the specification of any FRM, it can be valued in the risk-neutral world according to the valuation formula described in chapter 3. 3.8 can be expressed in a more intuitive form as

$$
\begin{aligned}
V(t(0)) = E^{\mathbb{Q}}_{t(0)}\Bigg[\sum_{i=1}^{I} \mathrm{e}^{-\int_{t(0)}^{t(i)}((1-\tau_F)r(s)+\ell)ds}\Bigg(\prod_{j=1}^{i-1}(1-\lambda^d_j-\lambda^p_j)\Bigg)\Bigg]\lambda^d_i W(i) + \\
E^{\mathbb{Q}}_{t(0)}\Bigg[\sum_{i=1}^{I} \mathrm{e}^{-\int_{t(0)}^{t(i)}((1-\tau_F)r(s)+\ell)ds}\Bigg(\prod_{j=1}^{i-1}(1-\lambda^d_j-\lambda^p_j)\Bigg)\Bigg]\lambda^p_i A(i) + \\
E^{\mathbb{Q}}_{t(0)}\Bigg[\sum_{i=1}^{I} \mathrm{e}^{-\int_{t(0)}^{t(i)}((1-\tau_F)r(s)+\ell)ds}\Bigg(\prod_{j=1}^{i}(1-\lambda^d_j-\lambda^p_j)\Bigg)\Bigg]M \\
I = 360 \qquad (8.2)
\end{aligned}
$$

where the first term represents the expected value of future payments, thus, the non-absorbing state. The second and third term represent the expected cash flows from prepayment and default respectively. For prepayment, this future value is simply the balance on the loan at that time. However, for default, the calculation is not so trivial. It represents the expected recovery on default.

In a technical sense, the cash flow from a default is also to be estimated. Since this value depends greatly on whether the loan was insured or not, this means I must make some assumptions on insurance. Generally, loans with a LTV greater than 80% will have insurance. It is important to realize that this LTV is origination LTV and not refreshed LTV.

I will use the "recovery of face value" (RFV) formulation known in the literature. This formulation is $W(i) = (\phi + \psi)A(i)$, where ϕ is the percent insured and ψ is the percent recovered in the absence of insurance. Following previous similar analysis, I assume the recovery rate $\psi = 1 - \omega$, where the loss rate ω is stochastic, but related to the original house value by $\omega = \kappa^e(U(i)/H(t))$. If I assume the loss is independent of the other stochastic state variables[91], then the expected recovery is $W^e(i) = min(1, \phi + \psi)A(i)$ where the min function is used to prevent the possibility of insurance and recovery exceeding unity.

In the event the default option is not exercised, both the mortgage payment and prepayment value, A(i) are easily calculated since they are completely specified in the origination data. The payment is calculated as $M' = L[\frac{c/12}{1-\frac{1}{(1+c/12)^{360}}}]$, with c being the contract rate and L being the loan amount. Part of this payment represents a taxable event to the lender and thus needs to be modified. Specifying τ_S and τ_F as the Federal and State tax rates respectively[1], the modified payment cash flow becomes $M = (1-\tau_F)(1-\tau_S))(1+c)U(i-1) + (M' - (1+c)U(i-1))$ where U(i)=$U(i) = M'[\frac{1-\frac{1}{(1+c/12)^{360-i}}}{c/12}]$ is the unpaid principal on the loan in period i. For prepayment, the cash flow is A(i)=U(i)+M.

8.2 introduces a new set of parameters to be estimated. Namely, the liquidity term since the market for mortgages is extremely sticky compared to that of the treasury market and the multiplicative and additive risk parameters. Some analysis simply restricts the multiplicative parameters to one[2], but I choose to estimate them explicitly.

8.2 IMPLEMENTATION

Calibration of the parameter vector is achieved by the multi-variable minimization function

$$min \sum_{h=1}^{H} (V_h(t(0))/L_h - (1-\delta_h))^2 \tag{8.3}$$

$$\Phi = (\mu_p, \nu_p, \mu_d, \nu_d, l, \kappa^e)$$

where H is the number of loans used in the calibration procedure. The entire universe of loans could not be used due to computational limitations. This calibration is achieved by use of Monte Carlo Integration. Explicitly, I use a simulated annealing technique that, over time, converges to a Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm.

[1] I set the Federal tax rate at 28% and the state tax rate at 4.32%. These amounts are taken from a report published by the Congressional Budget Office[1]

[2] This is the conditional diversification hypothesis referred to in chapter 2

Simulated annealing is a global optimization technique that overcomes a major flaw of generic algorithms that do not have a smooth state space surface. In other words, it helps keep the minimization function from getting trapped in a local minima. This method works because of a gradual ”cooling” of the optimization function. The cooling, which is how the procedure is named, is a function of time that gradually lowers the acceptance probability of a less-optimal solution. The procedure will, at first, allow both uphill and downhill movements in the optimization function. As the function “cools”, its tendency moves closer to choosing only downhill movements in the likelihood. Thus, at the end, it is normal optimization routine. In this case, the final routine is the BFGS algorithm.

8.2.1 SIMULATION SCHEME

As mentioned before, the stopping time of a fixed rate mortgage is $\tau = \tau_d \wedge \tau_p \equiv \min(\tau_p, \tau_d)$. I can model this by modeling a stopping time for both prepayment and for default. Each process is governed by its respective baseline that was estimated in chapter 7. These processes, in the form of 7.1 and 7.2 are described in the real world. By including the additive and multiplicative risk adjustments in the formula, we can get their representation in the risk-neutral world.

Hazard Diffusion Process in Real Probability Measure

$$d\lambda_0^p(t) = \kappa^p(\theta^p(t) - \lambda_0^p(t))dt + \sigma^p\sqrt{\lambda_0^p(t)}dW(t) \tag{8.4}$$

$$d\lambda_0^d(t) = \kappa^d(\theta^d(t) - \lambda_0^d(t))dt + \sigma^d\sqrt{\lambda_0^d(t)}dW(t) \tag{8.5}$$

Hazard Diffusion Process in Risk-Neutral Probability Measure

$$d\lambda_0^p(t) = \mu^p(\kappa^p\theta^p(t) - (\kappa + \nu^p)\lambda_0^p(t))dt + \sigma^p\sqrt{\lambda_0^p(t)}dW(t) \tag{8.6}$$

$$d\lambda_0^d(t) = \mu^d(\kappa^d\theta^d(t) - (\kappa + \nu^d)\lambda_0^d(t))dt + \sigma^d\sqrt{\lambda_0^d(t)}dW(t) \tag{8.7}$$

where μ^p, μ^d are multiplicative risk adjustment factors and ν^p, ν^d are additive risk adjustment factors.

With the baseline stochastic processes fully described, my attention turns toward simulation. While Euler discretization is very popular, I chose another method purposed by Feller and Cox[48].[3] They noticed that even though these processes are not explicitly solvable, the transition densities are known and, in fact, has a non-central chi-square distribution. Glasserman has shown that this simulation procedure can produce better than or equal results than the ordinary Euler discretization.

A noncentral chi-square random variable, χ^2, with ν degrees of freedom and noncentral parameter λ has distribution

$$P(\chi_\nu^{\prime 2}(\lambda) \leq y) = F_{\chi_\nu^{\prime 2}}(y) \equiv e^{-\lambda/2} \sum_{j=0}^{\infty} \frac{(\frac{1}{2}\lambda)^j/j!}{2^{(\nu/2)+j}\Gamma(\frac{\nu}{2}+j)} \int_0^y \; z^{(\nu/2)+j-1}e^{-z/2}dz \tag{8.8}$$

for $y > 0$. The transition densities of the termination processes can now be expressed in their transitional form:

$$\lambda(t) = \frac{\sigma^2(1-e^{-\alpha(t-u)})}{4\alpha}\chi_d^{\prime 2}(\frac{4\alpha e^{-\alpha(t-u)}}{\sigma^2(1-e^{-\alpha(t-u)})})\lambda(u), \; t > u \tag{8.9}$$

where

$$d = \frac{4\theta_t\kappa}{\sigma^2}. \tag{8.10}$$

[3]See Boyle et al. [12] for a thorough discussion of Monte Carlo methods in asset pricing. Although, an even more complete discussion can be found in Glasserman [48].

I must now draw from a noncentral chi-square distribution with d degrees of freedom and noncentrality parameter λ to simulate the termination processes and the interest rate processes.[4] The house process will be simulated following a Euler discretization since it differs in its transition density.

The interest rate diffusion processes were described in chapter 5. The treasury bonds and t-bills those processes were built on were available in the market and, thus, do not need any risk adjustments. Since they are also modeled in the CIR framework, they can be simulated using their transition densities. The interest rates will be used as covariates in the hazard model as well as in the Monte Carlo integration for discounting purposes.

Diffusion Processes for the Instantaneous Interest Rate

$$dy_\imath(t) = (\kappa_\imath(\theta_\imath - y_\imath(t)) - \nu_\imath y_\imath(t))dt + \sigma_\imath\sqrt{y_\imath(t)}dz_\imath^{\mathbb{Q}} \qquad \imath = 1, 2 \tag{8.11}$$

The house price index for each fip at the beginning of the observation period will serve as H(0) for that fip. It will then evolve correlated with interest rates affecting the hazards of each mortgage. Since it also is created from prices in the market, no risk adjustment is necessary.

Diffusion Processes for the Instantaneous Interest Rate

$$dH = (r_t - s)Hdt + \sigma_H dz_H \tag{8.12}$$

As mentioned earlier, the house price process is built under a different framework than the other processes and must be simulated differently. I will use a Euler discretization of the house price evolution. It is as follows:

$$H_{i+1} = H_i e^{[r_t - s - \sigma^2](t_{i+1} - t_i) + \sigma\sqrt{t_{i+1} - t_i}Z_{i+1}} \tag{8.13}$$

[4] All the uniform random variables used to generate other random variables were created using the Mersenne Twister.

With all the components fully described, the cash flows from each mortgage can be simulated by Monte Carlo integration and using 8.2. This is a multi-step process in which time is incremented one time-step.[5] At that time, all variables are recalculated and then the process is stepped forward once again. This continues for 30 time-steps until it is time for an observation window. At that point in time, an estimate for the number of loans defaulting and prepaying is constructed using the shifted baselines estimated in chapter 4, the interest rate processes from chapter 5, the house price process from chapter 6, and the termination processes estimated in chapter 7, Using the simulated annealing technique and the BFGS algorithm, estimation of the parameter vector can proceed.

8.3 PARAMETER CALIBRATION RESULTS

The minimization equation 8.3 caused a convergence problem. In fact, immediately after time t=0, there is a jump that will minimize equation 8.3 quite effectively. All of the mortgages in the sample were prepaying, either by way of the multiplicative parameter or the additive[6], therefore minimizing the equation to virtually zero. This resulted in an untenable situation. The circumvention of this problem was achieved by dividing the data set into two separate data sets.

The first data set needed to be one where the default hazard was effectively zero. I, thus, restricted it to loans with very low MTMLTV. This alone would not solve the convergence problem. I further restricted the data set to loans that paid positive points. With these two conditions, default can effectively be ignored and estimation of the prepayment vector is possible. The second data set consist of loans with relatively high MTMLTV values making

[5] dt here 1/30 of a month.

[6] This occured mostly in the multiplicative parameter, most likely due to how easy it is to trigger a prepayment with a slight change.

prepayment unlikely. With these limitations, both the prepayment and default parameters are able to be estimated with the particle filtering procedure described earlier.

Table 8.1 contains estimates of both sets of parameters used in the valuation process. The loss rate is only estimated in the default parameter set due to the segregation of default and prepayment. The valuation model performaed very well over the observation period. The multiplicative risk parameter for default is 1.452. While this is greater than 1, it is in agreeance with the diversification hypothesis.[7]. The prepayment multiplicative adjustment is slightly higher at 1.828. The additive risk adjustments are negative, but not enough to cause problems in the simulation procedure.[8] The RMSE for both prepayment and default are both under 1.5 %. These errors are small, but are in-sample errors.

8.4 PRICING PERFORMANCE

To gauge the performance of the valuation model and the estimated parameters several out-of-sample test was performed. To capture possible segment specific errors, I categorized sample loans by their origination LTV[9] and valued each category independently. The results are in table 8.2. The model is robust in that the RMSE are similar in magnatude to the in-sample test.

I also wanted to gauge the pricing formula on the individual options. I randomly picked a mortgage from the data set that was close to an "average" mortgage. I then systematically priced the mortgage with and without the different embedded options. Table 8.3 contains the results of this pricing as well as the decompostition of the cash flows. The sample mortgage without the possibility of default or prepayment is valued at $125,985. With only

[7]Most studies have restricted this risk parameter to 1 and not estimated it explicitly

[8]Negative risk adjustments can cause your simulation procedures to be mean-averting rather than mean re-verting.

[9]As opposed to their refreshed LTV that has been used throughout this analysis.

the possibility of default, it is valued at \$121,682. With only the possibility of prepayment, it is valued at \$124,966. With both options included, it is valued at \$123,423.

TABLE 8.1

Monte Carlo Calibration Results

Number of Loans in Each Sample: H=1,000

Number of Simulations: N=10,000

Propagation Method: Transition Density

Monthly Time Subintervals: $1/\Delta = 30$

	1995-2002 Sample
Variable	**Estimate**
μ_p	1.828
ν_p	-1.293
Liquidity (%)[10]	0.582
RMSE (%)	1.948
μ_d	1.452
ν_d	-0.986
Loss Rate (%)	21.472
Liquidity (%)	0.695
RMSE (%)	1.105

[10]The liquidity premium estimated at the first stage represents that for the loans with LTVs under 50%. It is not used for the following tests of pricing performance.

TABLE 8.2

Out-of-Sample Pricing Errors

Number of Loans in Each Sample: H=1,000

Number of Simulations: N=10,000

Propagation Method: Transition Density

Monthly Time Subintervals: $1/\Delta = 30$

	1995-2002 Sample
LTV Category (%)	**RMSE (%)**
$70^+ - 75$	1.697
$75^+ - 80$	1.367
$80^+ - 85$	1.534
$85^+ - 90$	1.864
$90^+ - 95$	1.468
$95^+ - 100$	1.406

TABLE 8.3

Monte Carlo Value Decomposition

	Loan Profile
	Sample
Original Loan Amount($)	122,863
Contract Rate(%)	7.875
Points	-0.155
Actual Value($)	123,053
Estimated Value($)	123,423
RMSE (%)	0.30
	Decomposition
Payments	19,312
Prepayment	102,480
Default	865
Total	122,657

8.5 CONCLUSIONS

Default and prepayment are events that determine the length of a mortgage's life. More importantly, they determine the resulting cash flows that will occur and, thus, the present value of cash flows to the lender. Duration analysis is a common statistical technique to analyze patterns of default and prepayment. However, in most studies, the baseline is not the focus of the work. In fact, in most papers the baseline is never even mentioned. This assumes that once the covariates of a particular mortgage are known, then the probability of default and prepayment are known. The major source of uncertainty in these models come from the unknown term structure.

Reduced-form modelling of mortgage termination recognizes that there are more sources of uncertainty other than the covariates and the term structure that influences them. The inclusion of the doubly stochastic nature of this model represents those sources of uncertainty. In a "doubly stochastic" model, not only the event of default is uncertain, but the probability of that event.

I unite several methods from mortgage termination literature by dividing mortgages into strata by time of origin and location of the underlying asset. I then treat each strata as a different draw from the underlying stochastic process. In order to capture the doubly stochastic nature of the model, I use a state-space structure to model the baseline hazard processes of prepayment and default. The stochastic variation in the baseline is meant to represent the idiosyncratic risk that influences the actual number of terminations.

While kalman filtering has been the traditional method of estimating a state-space model, it assumes that the model is linear and gaussian. This is, unfortunately not the case, for my hazard model since it is neither of those. Therefore, I use a particle filter approach, since it provides consistent estimates even in the non-linear, non-gaussian setting.

The interest rate process estimated is not a doubly stochastic process. To avoid assumptions about the measurement error and following Chen and Scott, I estimate the term structure as a two-factor extended CIR process. This enables me to use the exact maximum likelihood.

The house price process is modelled as a correlated stochastic process to one of the interest rate factors. This enables the mean-reverting trend of the house price process to be influenced by the current level of interest rates. This addition is a major addition in this paper in that it allows a full picture of the termination process and the factors that drive terminations.

The main goal of this paper is to ascertain the markets assessment of the risk of each loan. To capture this risk, I make use of the no-arbitrage principle. This states that the present value of all the loans future cash flows, priced according to the markets assessment of the riskiness of the loan, must be equal to the face value of the loan. This is commonly known as the famous "balancing equation". This will convert the actual processes into risk neutralized ones.

With two hazard processes, a two-state term structure, and a house price process, I have five state variables used in the Monte Carlo valuation and calibration stage. With interest rates and house prices included in the hazard covariates, I am explicitly allowing for dependency between the stochastic interest rates and the termination events, as well as house prices and the termination events. House prices and interest rates are modeled as correlated processes in the estimation stage.

A notable covariate that is missing from this analysis, although proxied for, is FICO scores. Having such a covariate is likely to be very informative about the likelihood of a default by the borrower and how the lender uses that information. Furthere analysis with more recent data and borrower specific information would be another major contribution to the literature.

Bibliography

[1] Effective federal tax rates: 1979-2001. Congressional Budget Office, April 2004.

[2] Odd Aalen. Non-parametric inference for a family of counting processes. *Annals of Statistics*, 6(4):701–726, 1978.

[3] Brent Ambrose and Anthony Sanders. Commercial mortgage-backed securities: Prepayment and default. *Journal of Real Estate Finance and Economics*, 26(2-3), 2003.

[4] P. K. Andersen and R. D. Gill. Cox's regression model for counting processes: A large sample study. *Annals of Statistics*, 10(4):1100–1120, Dec. 1982.

[5] S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp. A tutorial on particle filters for on-line non-linear/non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing*, 50(2):174–188, 2002.

[6] Alain Bélanger, Steven E. Shreve, and Dennis Wong. A general framework for pricing credit risk. Working Paper, April 2003.

[7] C. Berzuini, N. G. Best, W. R. Gilks, and C. Larizza. Dynamic conditional independence models and Markov chain Monte Carlo methods. *Journal of American Statistical Association*, 92:1403–1412, 1997.

[8] Tomasz R. Bielecki and Marek Rutkowski. *Credit Risk: Modelling, Valuation and Hedging*. Springer, 2002.

[9] Tomas Björk, Yuri Kabanov, and Wolfgang Runggaldier. Bond market structure in the presence of marked point processes. *Mathematical Finance*, 7(2):211–223, April 1997.

[10] Fischer Black and Myron Scholes. The pricing of options and corporate liabilities. *The Journal of Political Economy*, 81(3):637–654, May-June 1973.

[11] Robert R. Bliss. Testing term structure estimation methods. *Advances in Futures and Options Research*, 9:191–231, 1997.

[12] P. Boyle, M. Broadie, and P. Glasserman. Monte Carlo methods and security pricing. *Journal of Economic Dynamics and Control*, 21:1267–1321, 1997.

[13] Pierre Brémaud. *Point Processes and Queues: Martingale Dynamics*. Springer-Verlag, 1981.

[14] Michael J. Brennan and Eduardo S. Schwartz. Determinants of GNMA mortgage prices. *AREUEA Journal*, 13(3):209–228, 1985.

[15] N. E. Breslow. Covariance analysis of censored survival data. *Biometrics*, 30(1):89–99, Mar. 1974.

[16] Damiano Brigo and Fabio Mercurio. *Interest Rate Models-Theory and Practice*. Springer, 2001.

[17] Andrew L. Brunson, James B. Kau, and Donald C. Keenan. A fixed-rate mortgage valuation model in three state variables. *The Journal of Fixed Income*, 11(1):17–27, June 2001.

[18] Stephen A. Buser and Patrick H. Hendershott. Pricing default-free fixed-rate mortgages. *Housing Finance Review*, 3(4):405–429, 1984.

[19] Stephen A. Buser, Patrick H. Hendershott, and Anthony B. Sanders. Pricing life of loan caps on default-free adjustable-rate mortgages. *AREUEA Journal*, 13(3):248–260, 1985.

[20] Charles A. Calhoun and Yongheng Deng. A dynamic analysis of fixed- and adjustable-rate mortgage terminations. *Journal of Real Estate Finance and Economics*, 24:9–33, 2002.

[21] A. Colin Cameron and Pravin K. Trivedi. *Regression Analysis of Count Data*. Cambridge University Press, 1998.

[22] Tim S. Campbell and Kimball Dietrich. The determinants of default on insured residential mortgage loans. *The Journal of Finance*, 38(5):1569–1581, Dec. 1983.

[23] J. Carpenter, P. Clifford, and P. Fearnhead. Improved particle filter for non-linear problems. *IEE Proceedings on Radar and Sonar Navigation*, 1999.

[24] Amitava Chatterjee, Robert O. Edmister, and Gay B. Hatfield. An empirical investigation of alternative contingent claims models for pricing residential mortgages. *Journal of Real Estate Finance and Economics*, 17(2):139–162, 1998.

[25] Ren-Raw Chen and Loius Scott. Pricing interest rate options in a two-factor Cox-Ingersoll-Ross model of the term structure. *The Review of Financial Studies*, 5(4):613–636, 1992.

[26] Ren-Raw Chen and Loius Scott. Maximum likelihood estimation for a multifactor equilibrium model of the term structure of interest rates. *Journal of Fixed Income*, 3:14–31, December 1993.

[27] Ren-Raw Chen and Loius Scott. Interest rate options in multifactor Cox-Ingersoll-Ross models of the term structure. *The Journal of Derivatives*, 20:53–72, 1995.

[28] Ren-Raw Chen and Louis Scott. Multi-factor Cox-Ingersoll-Ross models of the term structure: Estimates and tests form a Kalman filter model. *Journal of Real Estate Finance and Economics*, 27(2):143–172, 2003.

[29] D. R. Cox. Partial likelihood. *Biometrika*, 62(2):269–276, Aug. 1975.

[30] John C. Cox, Jonathan E. Ingersoll, and Stephen A. Ross. An analysis of variable rate loan contracts. *Journal of Finance*, 40:293–308, 1980.

[31] John C. Cox, Jonathan E. Ingersoll, and Stephen A. Ross. A theory of term structure of the interest rates. *Econometrica*, 53(2):385–408, Mar. 1985.

[32] Donald F. Cunningham and Patrick H. Hendershott. Pricing FHA mortgage default insurance. *Housing Finance Review*, 3(4):383–392, 1984.

[33] Yongheng Deng, John M. Quigley, and Robert Van Order. Mortgage termination, heterogeneity and the exercise of mortgage options. *Econometrica*, 68(2):275–307, March 2000.

[34] A. Doucet, J. F. de Freitas, and N. Gordon. *Sequential Monte Carlo Methods in Practice*. Cambridge University Press, Cambridge, 2000.

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

[36] Joost Driessen. Is default event risk priced in corporate bonds? Working Paper, University of Amsterdam, March 2002.

[37] Gregory Duffee. Estimating the price of default risk. *Review of Financial Studies*, 12:197–226, 1999.

[38] Darrell Duffie. *Dynamic Asset Pricing Theory.* Princeton University Press, Princeton, New Jersey, 3rd edition, 2001.

[39] Darrell Duffie and Kenneth J. Singleton. An econometric model of the term structure of interest-rate swap yields. *The Journal of Finance*, 52(4):1287–3121, 1997.

[40] Darrell Duffie and Kenneth J. Singleton. Modelling term structures of defaultable bonds. *Review of Financial Studies*, 12(4):687–720, 1999.

[41] Kenneth B. Dunn and John J. McDonnell. A comparison of alternative models for pricing GNMA mortgage-backed securities. *The Journal of Finance*, 36(3):375–392, June 1981.

[42] Kenneth B. Dunn and John J. McDonnell. Valuation of GNMA mortgage-backed securities. *The Journal of Finance*, 36(3):599–616, June 1981.

[43] James F. Epperson, James B. Kau, Donald C. Keenan, and Walter J. Muller III. Pricing default risk in mortgages. *AREUEA Journal*, 13(3):152–167, 1985.

[44] James Follain, Jan Ondrich, and Gyan Sinha. Ruthless prepayment? Evidence from multifamily mortgages. *Journal of Urban Economics*, 41:78–101, 1997.

[45] Qiang Fu, Michael LaCour-Little, and Kerry D. Vandell. Commercial mortgage prepayments under heterogenous prepayment penalty structures. *Journal of Real Estate Research*, forthcoming, 2003.

[46] Kay Giesecke. Default compensator, incomplete information, and the term structure of credit spreads. Working Paper, August 2002.

[47] Kay Giesecke. Credit risk modelling and valuation: An introduction. Working Paper, Cornell University, January 2003.

[48] Paul Glasserman. *Monte Carlo Methods on Financial Engineering*, volume 53 of *Applications of Mathematics: Stochastic Modelling and Applied Probability.* New York: Springer-Verlag, 2003.

[49] N. Gordon, D. J. Salmond, and C. M. Ewing. Bayesian state estimation for tracking and guidance using the bootstrap filter. *AIAA Journal of Guidance, Control and Dynamics*, 18:1434–1443, 1995.

[50] N. J. Gordon, D. J. Salmond, and A. F. M. Smith. A novel approach to non-linear and non-Gaussian Bayesian state estimation. *IEE-Proceedings Part F*, 140:107–133, 1993.

[51] Vyacheslav Gorovoy and Vadim Linetsky. Intensity-based valuation of residential mortgages: An analytically tractable model. *Mathematical Finance*, 17(4):541–573, 2007.

[52] Jerry Green and John B. Shoven. The effects of interest rates on mortgage prepayments. *Journal of Money, Credit and Banking*, 18(1):41–59, Feb. 1986.

[53] Arden R. Hall. Valuing the mortgage borrower's prepayment option. *AREUEA Journal*, 13(3):229–247, 1985.

[54] Aaron Han and Jerry A. Hausman. Flexible parametric estimation of duration and competing risk models. *Journal of Applied Econometrics*, 5(1):1–28, 1990.

[55] Oskar Ragnar Harmon. A hazard rate analysis of single-family conventional mortgage loan delinquency payment patterns. *Housing Finance Review*, 8:291–303, 1989.

[56] M. Izard and A. Blake. *Computer Vision – ECCV'96*, chapter Contour Tracking by Stochastic Propagation of Conditional Density. New York: Springer, 1996.

[57] Robert A. Jarrow, David Lando, and Stuart M. Turnbull. A Markov model for the term structure of credit spreads. *Review of Financial Studies*, 10(2):481–523, 1997.

[58] John Kalbfleisch and Ross Prentice. *The Statistical Analysis of Failure Time Data*. John Wiley and Sons, second edition, 1980.

[59] E. L. Kaplan and Paul Meier. Non-parametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53(282):457–481, June 1958.

[60] James B. Kau and Donald C. Keenan. An overview of the option theoretic pricing of mortgages. *Journal of Housing Research*, 6(2):217–244, 1995.

[61] James B. Kau, Donald C. Keenan, Walter J. Muller III, and James F. Epperson. Rational pricing of adjustable rate mortgages. *AREUEA Journal*, 13:117–128, 1985.

[62] James B. Kau, Donald C. Keenan, Walter J. Muller III, and James F. Epperson. The valuation and securitization of commercial and multifamily mortgages. *Journal of Banking and Finance*, 11:1525–546, 1987.

[63] James B. Kau, Donald C. Keenan, Walter J. Muller III, and James F. Epperson. Pricing commercial mortgages and their mortgage-backed securities. *Journal of Real Estate Finance and Economics*, 3(4):333–356, 1990.

[64] James B. Kau, Donald C. Keenan, Walter J. Muller III, and James F. Epperson. The valuation and analysis of adjustable rate mortgages. *Management Science*, 36(12):1417–1431, 1990.

[65] James B. Kau, Donald C. Keenan, Walter J. Muller III, and James F. Epperson. A generalized valuation model for fixed-rate residential mortgages. *Journal of Money, Credit and Banking*, 24(3):279–299, Aug. 1992.

[66] James B. Kau, Donald C. Keenan, Walter J. Muller III, and James F. Epperson. Option theory and floating-rate securities with a comparison of adjustable and fixed-rate mortgages. *Journal of Business*, 66(4):595–618, 1993.

[67] James B. Kau, Donald C. Keenan, and Taewon Kim. Default probabilities of mortgages. *Journal of Urban Economics*, 35(3):278–296, Aug. 1994.

[68] James B. Kau, Donald C. Keenan, Constantine Lyubimov, and V. Carlos Slawson. Subprime mortgage default. working paper, 2009.

[69] James B. Kau, Donald C. Keenan, and Alexey A. Smurov. Reduced form mortgage valuation as an alternative to option-pricing models. *Journal of Real Estate Finance and Economics*, 33(3):183–196, 2006.

[70] John Krainer, Stephen F. Leroy, and Munpyung O. Mortgage default and mortgage valuation. working paper, 2009.

[71] David Lando. *Mathematics of Derivative Securities*, chapter Modelling Bonds and Derivatives with Default Risk, pages 369–393. Cambridge University Press, 1997.

[72] David Lando. On Cox processes and credit risky securities. *Review of Derivatives Research*, 2(1):99–120, 1998.

[73] Wai K. Leung and C. F. Sirmans. A lattice approach to fixed-rate mortgage pricing with default and prepayment options. *AREUEA Journal*, 18(1):91–104, 1990.

[74] D. Madan and H. Unal. Pricing the risk of default. *Review of Derivatives Research*, 2:121–160, 1998.

[75] Joe Mattey and Nancy Wallace. Housing prices and the (in)stability of mortgage prepayment models: Evidence from california. working paper, 1998-2005.

[76] Brian P. McCall. Unemployment insurance rules, joblessness, and part-time work. *Econometrica*, 64(3):647–682, May 1996.

[77] Brian P. McCall. The determinants of full-time versus part-time reemployment following job displacement. *Journal of Labor Economics*, 15(4):714–734, 1997.

[78] Robert C. Merton. On the pricing of corporate debt: The risk structure of interest rates. *The Journal of Finance*, 29(2):449–470, 1974.

[79] Bruce D. Meyer. Unemployment insurance and unemployment spells. *Econometrica*, 58(4):757–782, July 1990.

[80] Wiji Narendranathan and Mark B. Stewart. Modelling the probability of leaving unemployment: Competing risks models with flexible base-line hazards. *Applied Statistics*, 42(1):63–83, 1993.

[81] Andrey D. Pavlov. Competing risks of mortgage termination: Who refinances, who moves, and who defaults? *Journal of Real Estate Finance and Economics*, 23(2):185–210, 2001.

[82] N. D. Pearson and T. Sun. Exploiting the conditional density in estimating the term structure: An application to the Cox, Ingersoll and Ross model. *Journal of Finance*, 49:1279–1304, 1994.

[83] M. Pitt and N. Shephard. Filtering via simulation: Auxiliary particle filters. *Journal of the American Statistical Association*, 94:590–599, 1999.

[84] Michael K. Pitt. Smooth particle filters for likelihood evaluation and maximization. Working Paper, July 2002.

[85] Randell J. Pozdena and Ben Iben. Pricing mortgages: An options approach. *Economic Review*, 2:39–55, 1984.

[86] R. L. Prentice and L. A. Gloecker. Regression analysis of grouped survival data with application to breast cancer data. *Biometrics*, 34(1):57–67, March 1978.

[87] John M. Quigley and Robert Van Order. Efficiency on the mortgage market: the borrower's perspective. *AREUEA Journal*, 18(3):237–354, 1990.

[88] John M. Quigley and Robert Van Order. Explicit tests of contingent claims models of mortgage default. *Journal of Real Estate Finance and Economics*, 11(2):99–117, 1995.

[89] Scott F. Richard and Richard Roll. Prepayments on fixed-rate mortgage-backed securities. *Journal of Portfolio Management*, 15(3):73–82, Jan. 1989.

[90] Geert Ridder and Insan Tunali. Stratified partial likelihood estimation. *Journal of Econometrics*, 92:193–232, 1999.

[91] Philipp J. Schönbucher. *Credit Derivatives Pricing Models: Model, Pricing and Implementation*. John Wiley and Sons, 2003.

[92] Eduardo S. Schwartz and Walter N. Torous. Prepayment and the valuation of mortgage-backed securities. *The Journal of Finance*, 44(2):375–392, 1989.

[93] Eduardo S. Schwartz and Walter N. Torous. *Financial Markets and Financial Crises*, chapter Caps of Adjustable-Rate Mortgages: Valuation, Insurance, and Hedging. University of Chicago Press, 1991.

[94] Richard Stanton. Rational prepayment and value of mortgage-backed securities. *The Review of Financial Studies*, 8:677–708, 1995.

[95] Richard Stanton and Nancy Wallace. Mortgage choice: What is the point? *Real Estate Economics*, 26(2):173–205, 1998.

[96] Glenn T. Sueyoshi. Semiparametric proportional hazards estimation of competing risks models with time-varying covariates. *Journal of Econometrics*, 51:25–28, 1992.

[97] Glenn T. Sueyoshi. A class of binary response models for grouped duration data. *Journal of Applied Econometrics*, 10:411–431, 1995.

[98] Sheridan Titman and Walter Torous. Valuing commercial mortgages: An empirical investigation of the contingent-claims approach to pricing risky debt. *The Journal of Finance*, 44(2):345–373, June 1989.

[99] Kerry Vandell, Walter Barnes, David Hartzell, Dennis Kraft, and William Wendt. Commercial mortgage defaults: Proportional hazard estimation using individual loan histories. *Journal of American Real Estate and Urban Economics Association*, 20(4):451–480, 1993.