Institute of Business and Economic Research

Fisher Center for Real Estate and Urban Economics

PROGRAM ON HOUSING AND URBAN POLICY

WORKING PAPER SERIES

DISSERTATION NO. D02-001

MODELING RESIDENTIAL MORTGAGE TERMINATION AND SEVERITY USING LOAN LEVEL DATA

By

Ralph DeFranco

May 2002

UNIVERSITY OF CALIFORNIA, BERKELEY

Modeling Residential Mortgage Termination and Severity

Using Loan Level Data

by

Ralph DeFranco

A.B. (U.C. Berkeley) 1994

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Economics

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, BERKELEY

Committee in charge:

Professor John M. Quigley, Chair

Professor Richard Stanton

Professor Roger Craine

Spring 2002

The dissertation of Ralph Guy DeFranco is approved:

Chair Date

Date

Date

University of California, Berkeley

Spring 2002

Modeling Residential Mortgage Termination and Severity

Using Loan Level Data

By

Ralph Guy DeFranco

Abstract

Modeling Residential Mortgage Termination and Severity

Using Loan Level Data

by

Ralph Guy DeFranco

Doctor of Philosophy in Economics

University of California, Berkeley

Professor John Quigley, Chair

This dissertation consists of three essays on modeling residential mortgages. Chapter 1 presents and estimates a new model of loss given default using a new dataset of prime and subprime mortgages. The model combines option theory proxies with information on the loan contract and the cash flow position of the borrower. The results suggest that severity on subprime and adjustable rate mortgages are similar to losses on fixed rate prime loans, but that investor owned properties have significantly higher losses than owner occupied houses. The results also suggest systemic over-appraisals on refinanced loans.

Chapter 2 uses option pricing methodology to value the prepayment and default options associated with a residential mortgage, if house prices are mean reverting.

Numerical solutions compare the results from the mean reverting house price model to the results from a model where house prices follow a geometric Brownian motion process. The main contributions are: (1) the value of the implicit rent (service flow) is derived as a function of the house price process instead of assumed to be constant, as in prior research, (2) the mean reverting model has additional factors that may help forecast mortgage termination, and (3) the house price process is shown to have a significant effect on the value of a mortgage over a wide range of parameter values.

Chapter 3 presents a modeling framework for residential mortgages that has separate models for each loan payment status (Current, 30 Days Late, 60 Days Late, 90+ Days Late, in Foreclosure, in REO, or Paid Off). It is shown that several classes of traditional mortgage prepayment and default models are restricted forms of this model, and that the restrictions are rejected empirically.

Professor John Quigley

Dissertation Committee Chair

To my loving wife Margareta.

I wish to thank my friends and family for tolerating my neglect during these trying years. In particular, I wish to thank Roger Craine, Bob Edelstein, Arden Hall, Alan Neale, Richard Stanton, and seminar participants at the U.C. Berkeley Real Estate Seminar for their helpful comments and suggestions. Special thanks goes to John Quigley for his encouragement and support, and to Nancy Wallace for turning my attention from banking to real estate.

Table of Contents

List of Figures

List of Tables

Introduction

This dissertation consists of three essays analyzing questions relating to modeling residential mortgage termination. The first chapter presents a new way of modeling loss given default for mortgages. The second chapter extends option theory based mortgage valuation methods to the case of mean reverting house prices. The final chapter proposes an expanded transition model for mortgage termination, and presents statistical and empirical tests indicating that it is superior to traditional models. Chapters 1 and 3 present new models that are estimated empirically using data not previously utilized in academic research, while Chapter 2 is a option theory paper that utilizes simulations to produce results that are compared to results from a popular model. The major questions addressed that have not been examined in prior research are summarized in Table 1.

Chapter	**Question**
1	Can the various methods of estimating mortgage severity improved?
1	Are appraisals, on average, too high on refinanced loans?
1	Are severities on subprime loans similar to severities on prime loans?
2	What are the implications of mean reversion in house prices for mortgages?
3	Is it statistically and empirically meaningful to disaggregate a model for non-terminated loans into separate models based on the monthly payment status (such as 30 days late, 60 days late, 90 days late, etc.)?

Table 1: Primary Questions Addressed by this Dissertation.

Chapter 1 presents and estimates a new theoretic model of loss given default on residential mortgages. The model combines elements from option theory, features of the loan contract, and information on the cash flow position of the borrower. The model is estimated using WLS on a new data set of prime and subprime mortgages, which is believed to be the largest database of its kind.

Chapter 1 makes several contributions. First, it explores and expands the theoretic underpinnings of severity modeling. Second, it is unique in testing for systematic over-appraisals on refinanced loans. The results suggest a statistically, but not economically significant upward bias in appraisals on refinanced loans. Third, Chapter 1 reinforces the conclusions of earlier studies that found that the ruthless default option model is inadequate by itself for describing actual default behavior. Fourth, Chapter 1 suggests for the first time that losses given default on subprime and Adjustable Rate Mortgages (ARMs) are similar to losses on traditional prime loans, while investor owned properties and balloon loans have significantly higher losses if they default. Results also suggest that the most important determinants of losses are the Loan-to-Value (LTV) ratio, the size of loan, and the lien position.

Chapter 2 investigates the theoretic implications of mean reverting house prices. This chapter for the first time uses option pricing methodology to value the prepayment and default options associated with a residential mortgage when house prices are mean reverting. Numerical solutions compare the results of the model developed here to the Kau et al. (1995) model where house prices follow a geometric Brownian motion process. Optimal prepayment and default boundaries are contrasted between the two models. The main contributions are: (1) the service flow (i.e. value of living in a house) is shown to be mean reverting, even though it is often incorrectly assumed to be independent of the house price process, (2) the mean reverting model has additional factors, such as the rate of mean reversion, that may help forecast mortgage termination that were overlooked by prior research, and (3) the house price process is shown to have a significant effect on the value of a mortgage over a wide range of parameter values. Thus, this chapter presents a new solution to the puzzle that very few of the households with "in the money" default options actually default (Foster and Van Order (1984), Kau et al. (1993), Vandell and Thibodeau (1985)). The new solution is that the options may be far more value that previously estimated.

In Chapter 3, a seven state Markov transition model is proposed for modeling residential mortgages, that has separate models for each loan payment status (Current, 30 Days Late, 60 Days Late, 90+ Days Late, in Foreclosure, in REO, or Paid Off). It is shown that traditional mortgage prepayment and default models are restricted forms of this model, and that the restrictions imposed by traditional models are rejected empirically. In addition, out-of-sample forecasts from this transition model are shown to be far superior to forecasts from simpler models based on less payment states. The results point to improved methodologies for pricing loans, setting loss reserves, and setting capital for regulated entities.

This dissertation contributes to the large and growing literature on mortgage prepayment and default. The importance of this research stems from the fact that the inability to control credit risk is a key factor in many financial institution failures. The uncertain cash flows from the $5.2 trillion in outstanding residential mortgages represent a major source of credit risk for financial institutions[1]. Improving the accuracy of forecasting models is of interest to regulators as well as investors, due to the explicit and implicit government guarantees. All three chapters have theoretic and empirical methodological advancements that can be used in pricing models for mortgages and mortgage-backed securities. The results presented in Chapters 1 and 3 can be used immediately to price mortgage credit risk, while Chapter 2 suggests directions for future modeling research. The modeling issues addressed in this dissertation are topical because regulated firms are increasingly being allowed to set their own capital requirements using internal forecasting models similar to the models examined here (Berkowitz 1998).

One unifying element throughout is improvements to modeling methodologies that can be used by risk managers and regulators. Chapters 1 and 3 both shed light on the poorly understood subprime market, and use the same large datasets. There has been little research on Subprime losses, yet around $120 billion worth of subprime residential mortgages were originated in 2000 (approximately 12% of all mortgages.) Subprime lending is a topical subject because of the large expansion by banks into the

---

[1] The figure is from the Flow of Funds, published by the Federal Reserve. It is for the United States at the end of 2000.

subprime market[2]. This dissertation is organized into the following sections: Literature Review, Data Description, Chapters 1, 2, and 3, Bibliography and Appendixes.

---

[2]A few examples are Bank of America's purchase of Indy Mac, Washington Mutual's purchase of Long Beach Capital, and First Boston's purchase of the Money Store. First Boston subsequently closed the Money Store and wrote-off almost the entire 2.5 billion dollar investment, while Bank of America exited the subprime market with a substantial write down in the fall of 2001.

Literature Review

While estimating severities on defaulted loans is important for pricing mortgages, surprisingly little is published on this topic. The literature related to severity primarily consists of two categories: (1) articles that use loss data to investigate the implications of theoretical models; and (2) articles concerned with empirical modeling. Prominent examples of the former type of paper are studies that test the implications of option theory for mortgage termination. One such implication that is consistently supported empirically is that the equity position of the borrower is a dominant factor determining whether a distressed mortgage forecloses or prepays (Kendall 1995). Another option-theoretic implication, which is more empirically controversial, is whether simple option models *by themselves* are sufficient for modeling severity. This proposition is tested and rejected in this paper.

Since data on transaction costs are unavailable, the more interesting question of whether borrowers execute the default option optimally cannot be directly tested. Therefore, attention in the literature has primarily focused on the 'ruthless' version of the default option theory. Foster and Van Order (1984) define 'ruthless' default as defaulting immediately when the value of a property drops below the value of the mortgage. The ruthless model assumes that there are no transaction costs, that the borrower has the ability to make payments and can borrow immediately at the market rate to purchase an equivalent property. Lekkas, Quigley, and Van Order (1993) were the first to use loss data to test the ruthless (frictionless) option model of defaults. The ruthless default hypothesis predicts that for a fixed house value and interest rate, there is an optimal "in the money" point at which a borrower defaults, which is independent of region and initial Loan to Value ratios (LTV). Using a sample of Freddie Mac conforming loans, they reject the ruthless default hypothesis because severity varied by region and LTV. Likewise, Capone and Deng (1998) analyzed severity rates for a sample of single-family mortgages and found the influence of option valuation variables only matter at the margin. They concluded that option-pricing models cannot be used by themselves to generate severities for mortgage pricing models. These studies point to the need for expanding the set of information used to predict severity beyond what is considered by standard option theory.

Other research on default that expands beyond option pricing variables do not explicitly analyze severity. For example, Foster and Van Order (1984), using FHA data, estimated the market value of a mortgage by discounting the mortgage payments at the then-current market interest rate, assuming a prepayment date of 40 percent of the remaining term. They found that only 4.2 percent of the loans with market loan-to-value ratios in excess of 110 percent defaulted, presumably because of transaction costs. Like Foster and Van Order (1984), Springer and Waller (1993) attempted to link option theory and empirical default research by testing whether transaction costs are important. Springer and Waller (1993) found statistically, but not economically, significant evidence for the necessity of using non-option related variables, using a sample of 209 distressed loans in Texas. Capozza, Kazarian, and Thompson (1997) echoed the empirical importance of the role of transaction costs and trigger events.

They used a large, geographically diverse, sample of defaulted loans, but also had no information on severity.

In addition to the papers discussed above there are a few papers that focus on purely empirical methods for estimating losses. These papers suffer from a lack of theoretic underpinnings and small sample sizes. For example, Smith, Sanchez, Lawrence (1996) estimate severity using only three different buckets for loan size. Wilson (1995) estimated a more detailed empirical loss function using data from California from 1992 to 1995. They found that the primary drivers of loss were changes in home prices followed by the lender, LTV, property type, loan size and county. OFHEO (1999) used the first and second moments of the area house price distribution to estimate loss severities on agency portfolios[3]. Smith and Lawrence (1993) used data on manufactured homes from a single financial institution to construct a Markovian forecasting model, and a separate loss model. Their models provided estimates of the expected loan losses for an entire loan portfolio. The regressors were: regional dummies, indicators of being 30 or 60 days delinquent in the last year, log of loan age, original and estimated current LTV, initial interest-rate, maturity, borrower's age and occupation, average foreclosure time, number of months for right of redemption, an indicator of judicial foreclosure, and state-level data on unemployment, retail, income, and mobile-home prices. In addition to having

[3] OFHEO chose to break-up loss severity into three parts: current loan principal, transaction costs, and funding costs. This was done in order to account for the timing of the various income and expenses during the time in default. Since the various components of loss severity are not available in the data used in this study, this method was not pursued.

subprime data and a large dataset, the specification used here is based on stronger theoretic grounds.

Loss data has occasionally been used for other purposes than testing option theory and pure forecasting. This paper continues this tradition by investigating differences in severity across different loan markets, for example, by comparing severity on refinanced and new-purchase loans. Other examples of interesting uses for severity data include Quigley and Van Order (1991), who focus on the public policy concern of banks' large exposures to loan portfolios. Clauretie and Herzog (1990) study the effects of varying state foreclosure laws on losses. Van Order and Zorn (2000) used loss data to investigate the effects of Community Reinvestment Act of 1977. Their primary focus was whether the relatively low flow of loans into low-income areas is a market failure, or due to differences in risk.

The most relevant paper for Chapter 2 is Kau et al. (1995). Kau et al. (1995) focuses on the need to model the prepayment and default options simultaneously. The authors show the prepayment and default option values if house prices follow a random walk. Several studies examine optimal mortgage option execution using a similar framework to Kau et al. (1995), but they always assumed that house prices follow a geometric Brownian motion process. Early pioneers include Titman and Torous (1989), who applied the contingent claim method of Brennan and Schwartz (1980) to mortgages. They focused on commercial mortgages because they do not have a prepayment option. Their results suggest that commercial mortgage rates generated by a two state variable contingent-claims pricing model provide accurate estimates of both commercial mortgage rates, and the changes in the spread between treasury bonds and commercial rates. Another example from this theoretic literature is by Cunningham and Hendershott (1984), who combine a random walk process for the house price together with a deterministic term structure to analyze the value of default if prepayment is ruled out. Epperson et al. (1985) extend the investigation of this kind of contract by including a mean-reverting term structure.

The effect of mean reversion on option values has been studied in other contexts, but never in the case a residential mortgage. Dixit and Pindyck (1993) examine the value of a simple option when the underlying asset's value is mean reverting. Cauley and Pavlov (2001) have two stochastic processes, one for property values, and one for cash flows. Unlike both Cauley and Pavlov (2001) and Dixit and Pindyck (1993), whom restrict their models to only allow default, the model presented in Chapter 2 allows for both default and prepayment.

Most papers in the theoretic option pricing literature assume the no arbitrage condition. One exception is Kuo (1995), who examines the value of the default option under mean reverting house prices. He assumes that the log of the house price is the sum of region specific changes, which are assumed to be AR(1), and house specific errors which are modeled as having a persistent and a transient shock. Since Kuo's focus was on estimating house price indexes, this specification offers little help in estimating mortgage values for housing where data on multiple sales is not available.

The paper presented here has several major differences from Kuo (1995). The first difference is that I keep the no arbitrage condition. The second difference is that I do not estimate house price indexes, and instead focus on the practical implications for modeling prepayment and default.

Empirical studies of house price dynamics suggests that house prices are poorly approximated by a geometric Brownian motion process, and instead are more consistent with a mean reverting process (Englund and Ioannides (1997), Meese and Wallace (1997, 1998), England, Gordon, and Quigley (1999)). One branch of the mortgage literature investigates mean-reversion in house prices empirically, but typically does not examine the implications for the default option. For example, Englund and Ioannides (1997) used quarterly data from 15 OECD countries and found a highly significant first order autocorrelation coefficient of around 0.45. These results are consistent with England, Gordon, and Quigley (1999), who looked at virtually all housing transactions in Sweden over a twelve year period. They rejected the hypothesis that house prices follow a random walk, in favor of a model of first order serial correlation. Meese and Wallace (1997, 1998) take the additional step of empirically examining the effects of mean reversion on mortgage pool valuation. To do that, they mimic Stanton (1995) by using a Cox, Ingersoll, and Ross term structure process, a Poisson parameter for the frequency of refinancing decisions, and a beta distribution for transaction costs. Mean reversion is consistent with the error correction model's of Abraham and Hendreshott (1996) and Meese, Wallace (1997, 1998). Error correction models estimate the fundamental values to which house prices return based on construction costs, real income, employment, and interest rates or net migration, respectively. While the error correction models are more realistic, mean reversion was chosen here for tractability.

Data Description

The datasets used in all three chapters come from LoanPerformance's (formally know as Mortgage Information Corporation) database of 3,879,913 private-issue (i.e. not Government Sponsored Enterprises) securitized prime and subprime loans, from 1993 to June 2000. This proprietary dataset has not previously been used in academic research. The major caveats are that the only recession in the data is for California in the early 1990's. Care was taken in filtering the data and examining outliers. Additional details are given in each chapter.

Chapter 1: Unifying Models of Severity on Defaulted Mortgages

Introduction

Understanding the default risk inherent in residential mortgages has become an increasing source of concern in the mortgage and mortgage-backed securities market. Increased concern reflects in part the large recent expansion of subprime lending, which are loans are made to borrowers with poor credit histories. In 2000, for example, approximately 12% of all mortgages originated, representing around $120 billion in loans, were subprime. Even on prime loans, economic conditions and the next round of international bank regulations[4] have prompted renewed interest in projecting default risk at the loan level.

To understand fully mortgage risk, however, it is not sufficient simply to estimate the likelihood of default. It is equally important to estimate the *severity* of a mortgage -- the percentage of the unpaid principal balance that is lost in the event of default. Severity has a one-to-one mapping with dollar losses, and thus its estimate fully captures the expected loss on a mortgage conditional on default. Severity is the more common way of modeling losses because many of the cost components, such as lost interest and commissions, are related to the size of the loan. Such loss estimates are

[4] See the Board Of Governors Of The Federal Reserve System(1999) SR-18.

required to correctly price loans and derivatives, and to set economic capital and loss reserves. For instance, rating agencies require mortgage severity estimates in the process of rating mortgage-backed securities. Similarly, regulators are interested in setting regulatory capital requirements based on a financial institution's risk profile. Mortgage default risk is a concern to regulators because of the large exposures that many financial institution's hold in their portfolios[5]. Currently all residential mortgages are treated equally for setting regulatory capital, but with the next Basil agreement, banks may be allowed to set their own capital requirements based on estimates of the probability of default and loss given default.

It is common practice for financial institutions to forecast severity simply by assuming a constant severity on all defaulted loans[6]. This reflects in part the fact that the mortgage termination literature has focused almost exclusively on estimating the probability of default[7]. However, using a constant severity neglects the possibility that severity depends in systematic ways on characteristics of the loan and borrower, as well as on the legal and economic environment. Moreover, assuming a constant severity when ranking the relative riskiness of loans can potentially provide very

---

[5] One example is the regulator OFHEO (1999), which conducted a stress test of Freddie Mac and Fannie Mae's portfolios by estimates dollar losses on defaulted loans. However, OFHEO only used the house price index for estimating losses, even though the results of this paper suggest that many more variables affect losses.

[6] One example is a major West Coast Savings and Loan, which uses fixed severity rates of 26 percent for setting loss and capital reserves. Another example is the rating agency Fitch, where they use fixed severity rates of 10, 20, 30 and 40 percent.

[7] See Hendershott and Van Order (1987) and Kau and Keenan (1995) for reviews.

misleading results. As Wang (2001) has pointed out, many risks rankings actually used in the industry, such as FICO[8] scores and mortgage scores estimated by lenders, are indeed based solely on default. Considering how relative risk rankings change when heterogeneous expected losses are factored in represents the next logical step.

This paper explores how severity on residential mortgages can be estimated using information commonly tracked in loan servicing databases. The variables considered in this study are based on a careful review of factors that are commonly cited in the literature and among lending practitioners as potential determinates of mortgage severity. These include elements from option theory, the loan contract, the cash flow position of the borrower, and cost related variables. By combining predictions of loan severity in this study with a model that predicts the probability of default, the results can be used to more accurately project losses on residential mortgages. The results indicate that severity depends systematically on variables known at the time of loan origination, and that utilizing that information can provide a dramatic forecasting improvement over several common severity estimation methods.

In addition to providing a much clearer assessment of mortgage risk, the analysis allows us to untangle the various factors that influence losses. Estimates of how individual factors effect expected loan losses can be of significant interest in and of themselves. For example, by controlling for economic factors we can more accurately

[8] Fair Isaac and Co. (FICO) credit bureau risk scores are based on a borrower's credit history and range from 300 to 900. Freddie Mac and Fannie Mae normally reject borrowers with FICO scores below 620.

determine which borrower and loan contract characteristics influence losses. This information can potentially be used to improve mortgage contracts, and thus help to better manage and price default risk[9].

There are a couple of important applications of looking at severity estimates across different mortgage markets that are worth mentioning. First, comparing severity estimates from the prime, Alt-A[10], and subprime markets enable us to investigate if defaulted Alt-A or subprime loans are substantially more expensive to dispose of than prime loans. Given the recent rapid expansion of these mortgage markets[11], coupled with a relative lack of research on these markets, this issue is especially relevant today. The results suggest that subprime loans have only a slightly larger severity (1%) than prime loans, once other factors are taken into account, while Alt-A loans had a substantially lower severity (5.8%). Second, I obtain separate estimates for severity on newly purchased and refinanced homes, which enables examining the possibility of systematic over-appraisals on refinanced loans. Since there is no market transaction for a refinanced loan, the appraiser has some discretion in estimating the

[9] One of the largest banks already uses an empirical loss model for setting origination rates.

[10] Alt-A loans are loans that of higher quality than subprime loans, but do not conform to Freddie Mac and Fannie Mae's requirements in some way. One, example is failing to provide complete documentation on income.

[11] A few examples major recent entries into this market include Bank of America's acquisition of Indy Mac, Washington Mutual acquisition of Long Beach Capital, and First Boston acquisition of the Money Store. First Boston subsequently closed the Money Store and wrote-off almost the entire 2.5 billion dollar investment, while Bank of America exited the subprime market with a substantial write down in the fall of 2001.

house value. Some industry participants have expressed concern that appraisers may have incentive to make a loan appear more attractive by appraising a property so as to have a Loan-to-Value (LTV) ratio of 80% or less. This may result in more repeat business for the appraiser from mortgage brokers, which would result in increase fee income. Thus, appraisers may systematically provide overly optimistic appraisals on refinanced loans. This incentive problem does not exist for newly purchased homes, because the appraised value is typically the same as the sale price. I find that severity does differ among refinances and original purchases in a manner consistent with this hypothesis. Since the majority of outstanding mortgage loans are refinances, the implications for financial institutions and regulators could be quite important.

The severity model presented here represents a substantial theoretical advancement over existing severity models. The model unifies elements from option theory, the loan contract, the cash flow position of the borrower, and cost related variables. The model is shown to produce a dramatic forecasting improvement over several common severity estimation methods.

This chapter is organized into the following sections: section 2 describes the loss data, section 3 describes the derivation of the model, section 4 describes the severity model for first lien loans, section 5 describes the severity model for second liens, section 6 is on validating the model, and section 7 is the conclusion. The appendixes discuss Short Sales Rates, Data Filtering and the accounting assumptions used in testing the model.

Data Description

Prior work on losses has been done using smaller datasets with a narrower focus (Wilson 1995). Data used in existing studies come from a single firm, a single state, or from a dataset created by mortgage insurance companies, none of which can be expected to be representative of loans overall. This problem of coverage is avoided in this study by using a dataset covering mortgages over the entire country. The data come from a proprietary data set consisting of 1,927,235 loans underlying some 953 mortgage and asset backed securities[12]. Only securities that report all loan level losses are included in the dataset. Of the 953 securities, 48% of the loans are paid off or resolved, and around 28,000 loans had usable loss data (less than 1000 loans were filtered out due to missing or suspicious values). Recent years are more heavily represented. Ten percent of the loss data comes from prior to March 1995, and the earliest reported loss is from February 1992. The most recent data is from July 2000.

The definition of severity used here is the percentage of the unpaid principal balance that is lost:

$$\text{Severity} = \text{Loss} * 100 / \text{Unpaid Principal Balance} \tag{1.1}$$

[12] Dan Feshbach and Kyle Lundstedt of Mortgage Information Corporation graciously provided access to their database of non-agency, prime and subprime loans.

The source of losses on individual loans can be broken down using the following accounting formula:

Loss = Unpaid Principal Balance +Months*(Monthly Lost Interest) - (Current House Price)*(1 - Real Estate Commission- Fix-up Costs) +Unrecoverable Costs - Recoveries from Mortgage Insurance (1.2)

Months stands for the number of months of missed payments between the time when the borrower was last caught up on payments and the house was liquidated. *Unrecoverable Costs* are expenses related to the liquidation of the asset, advances for insurance premiums, property taxes, etc. The dataset used here only contains the aggregate dollar loss amount for each loan, and not the various components of loss[13]. It is not known if all of various components were reported by the companies that provided the data used in this research. To try to account for possible differences and definitions in the data, indicator variables for each mortgage servicing company were included in the regression. Table 2 shows the total number and value of all loans in the data set, as well as the average loss and severity by loan type for all first-lien loans.

[13] See Wilson (1995) for the relative size of the various components of severity, such as principle, interest, legal fees, etc.

Type of Loan	Total % of Loans	Total Value at Origination	# of Loans with Losses	Avg. Loss on Each Loan	Avg. Severity on Each Loan
Prime	44.92%	$232,863,049,010	12,486	$83,057	35%
Alt-A[14]	12.77%	$33,583,518,487	1,656	$45,046	28%
Subprime	40.23%	$62,189,639,210	14,790	$28,125	46%
Total	*100%*	$328,636,206,707	$28,932	$52,076	*41%*

Table 2: Total Values for the Entire Data Set, and Average Dollar Loss and Severity by Type of Loan.

The first two columns of Table 2 shows the entire universe of loans for which loss data would be reported, even though the vast majority of these loans did not actually default.

The average severity on is higher for subprime loans than for prime and Alt-A loans, even though the absolute loss is smaller. This is because subprime loans are smaller on average, so the denominator in severity (the outstanding balance) is smaller. Table 3 shows the total number of loans and percentage of the original balance lost for all paid off loans.

[14] Alt-A loans are loans that of higher quality than subprime loans, but do not conform to Freddie and Fannie's requirements in some way. One, example is failing to provide complete documentation on income.

Loan Type	# Paid Off Loans	Origination Value of Paid Off Loans	# Loans w. Losses	Total Losses	Loss/ Original Balance
Prime	567,625	$154,189,538,528	12,486	$1,037,045,810	0.67%
Alt-A	111,147	$17,030,965,941	1,656	$74,595,631	0.44%
Subprime	306,259	$24,440,062,249	14,790	$415,962,296	1.70%
Totals	*985,031*	*$196,707,114,588*	*$29,324*	*$1,541,708,896*	*0.78%*

Table 3: The Percentage of Original Balance Lost for All Paid off Loans.

Paid off loans include loans that defaulted. In order to estimate the percentage of the starting balance that was lost for each loan category (the final column in Table 3), only those loans that paid off or defaulted are counted. That is because loans that are still active may either pay off voluntarily or default. Below, Figure 1 shows the distribution of dollar losses for the various loan types.

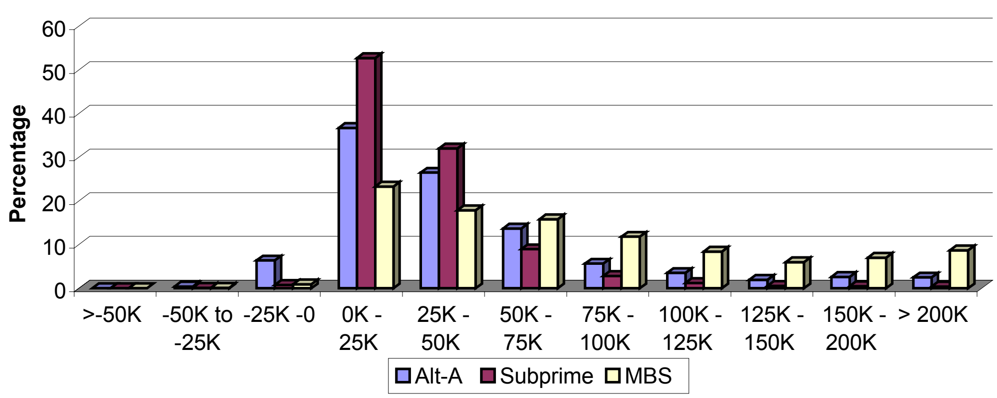

Figure 1: Distribution of Dollar Losses by Loan Type.

Figure 1 indicates that defaulted subprime loans usually have lower losses than defaulted prime loans. Negative losses represent a “gain-on-sale” where the lender actually make a profit on the disposed of loan[15]. Figure 2 shows that the severity on second lien mortgages are often close to 100 percent.

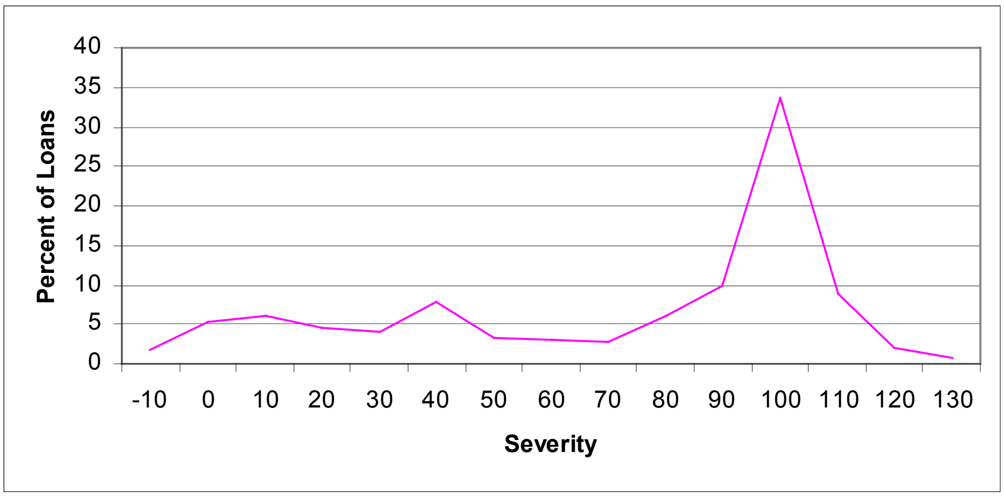

Figure 2: Distribution of Severity on Second Lien Loans.

Variable Descriptions

The following table introduces the variables used for estimating severity, and separates them into four broad categories.

[15] Most states require lenders to return any gains on a foreclosure to the borrower. Such occurrences are rare, since if there was sufficient equity in the house, the borrower is better off selling the house than having it foreclosed on.

Variable Categories	Source of Losses	Available Data
Option Theory	Economic factors	House price Index (HPI) means and volatility, interest rates
Loan Contract Characteristics	Original loan contract	Lien position, owner occupied or investor owned, mortgage insurance
Borrower's Cash Flow	Borrower's lack of upkeep	FICO score, original LTV, debt to income ratio, ARM vs. Fixed
Cost Variables	Servicer, transaction costs	State laws, servicer, loan age, bankruptcy, lost interest

Table 4: Factors Which Influence Losses on Defaulted Loans.

The first column of Table 4 shows the broad overall groupings for the variables used in the model. The second column highlights why these variables might influence losses. The third column shows the actual variables used in the severity model.

Losses on defaulted loans can be thought of as arising from four different categories of sources: financial option theory, loan contract characteristics, borrower's cash flow, and servicing cost variables. Each category is briefly summarized here, while more details are given in later sections. *Options theory* views default as a put option that gives the borrower the right to sell the house to the lender at the current house value. Standard option theory arguments suggest a mortgage is more valuable the lower the coupon rate relative to the current interest, and the higher the current house price. *Loan contract characteristics* that influence losses include the lien position, whether or not the property is owned by an investor, the mortgage coupon rate, whether a loan is a fixed-rate, balloon or adjustable-rate mortgage, etc. *Borrower's cash flow* related variables may proxy for a lack of upkeep by the borrower. *Cost variables* that influence losses include costs due to differing state eviction laws, the efficiency of the loan servicer, the loan age, if the borrower is in bankruptcy, and lost interest[16]. The one group of variables not used here are socioeconomic variables often found in default studies. While socioeconomic variables such as unemployment and divorce affect the probability of default, Clauretie and Herzog (1990) point out that these factors are not expected to influence loss given default.

The variables with the strongest foundation in theory are the *option theory variables*[17]. Rational borrowers will increase their wealth by defaulting when the balance of the mortgage exceeds the value of the house plus the value of the transaction and reputation costs. Similarly, by prepaying when market values exceed par, borrowers can increase their wealth by refinancing. Thus from the option theory point of view, the value of the mortgage depends on the house price, risk-free interest

[16] See Clauretie and Herzog (1990) for a brake-down of additional costs, which were not itemized in the data set used here.

[17] See Hendershott and Van Order (1987) and Kau and Keenan (1995) for reviews.

rates, the mortgage coupon rate, the outstanding mortgage balance, and the age of the loan[18].

Since the value to the borrower of defaulting is not directly observable, the proxy used here is the probability of negative equity. It is the probability that a house's value has depreciated sufficiently since the loan was originated to destroy the borrower's equity, and is calculated as in Deng, Quigley and Van Order (2000):

$$\text{Prob\_Neg\_eq} = \text{prob}(E < 0) = \Phi\left(\frac{\log(V) - \log(M)}{\sqrt{\omega}}\right) \tag{1.3}$$

Where E is borrower's equity, V is the present value of the remaining mortgage payments, M is the current market value of the property estimated using OFHEO’s (1999) Metropolitan Statistical Area (MSA) level house price index, Φ is a cumulative normal density function, and ω is the variance of the house price index. The default rate is expected to accelerate as house prices fall and the transaction costs become overwhelmed. Therefore, the probability of negative equity squared is also included.

Since the option to refinance has value, its value is also related to the probability of default. Current interest rates are important for default because of their inverse relationship to the value of future mortgage payments. As with the value of the default option, we do not directly observe the borrower's refinancing incentive.

[18] See DeFranco (2001) for additional variables that influence the mortgage option values when house prices follow a mean reverting process.

Therefore, the prepay incentive, or call option, proxy is the relative spread of interest rates (see Deng, Quigley and Van Order 1996.) Age has two competing effects on losses. The older the loan, the greater the potential for depreciation, but also the greater potential for built up equity.

Option based models provide important insights into the comparative statics of borrow behavior in a frictionless, perfect market where default and prepayment decisions are independent of decisions to move. Since these assumptions are rather stringent, it is not surprising that strictly option-based models have significant empirical shortcomings[19]. This suggests the need to include non-option related variables, such as the loan contract characteristics.

Major *loan contract characteristics* that may influence losses are: (1) the lien position, since second lien mortgages have higher severities; (2) whether or not the property is owned by an investor, since the debt burden of an investor increases significantly if a property is vacant or if renters cause a property to depreciate more rapidly; and (3) the mortgage coupon rate, since they higher coupon rate increases the amount of lost interest between the last payment and the month the house is sold. Adjustable Rate Mortgages (ARMs) and balloon loans typically have higher delinquencies, perhaps because these borrowers self-select these kinds of loans in

[19] Empirically, non-option theory variables have been found to be statistically significant and interest rates typically are not significant (Lekkas, Quigley, and Van Order (1993)).

order to achieve lower monthly payments compared to fully amortizing fixed rate loans. As in Deng, Quigley and Van Order (2000), the original Loan-to-Value (LTV) ratio is used as a proxy for information asymmetry[20]. As discussed in Kau, et al. (1992), high LTV loans should have shorter average times to default. The lower the LTV, the greater the borrower's incentive to protect their investment. The relationship may not be linear however, because many lenders require a lower LTV from borrowers with weaker credit histories. Therefore, the marginal effect of LTV is allowed to change for LTV values greater than 80%.

Borrower Cash flow variables consist of key borrower characteristics that may predict the likelihood that individuals will default due to the inability to meet their payments. Cash flow related variables may also proxy for a lack of upkeep by the borrower. It seems reasonable to assume, for example, that a homeowner struggling with their payments is much less likely to invest in home improvements and general upkeep. The debt to income ratio is one of the primary factors that loan originators consider in evaluating the ability of a borrower to meet their debt obligations.

The *cost variables* that influence losses include costs due to differing state eviction laws, the efficiency of the loan servicer, if the borrower is in bankruptcy, and lost interest. Mortgage servicers influence losses by minimizing expenses and the time

---

[20] Dunn and Spatt (1988) proposed interpreting mortgage contract terms as devices for dealing with asymmetric information inherent in mortgage lending.

needed to dispose of a property, by pursuing alternatives to the foreclosure process[21], and by optimally timing the sale of the property.

Since state laws and regulations affect the length of time required to evict residents, severity varies systematically by state. State foreclosure laws differ in three fundamental ways. First, foreclosure can be done with a judicial or non-judicial procedure. Judicial procedures typically take longer and require greater legal expenses. Second, states may differ with respect to the right of redemption, which allows the mortgagor to redeem the property in exchange for the delinquent payments and foreclosure expenses. In some states the mortgagor is allowed to remain in possession of the property during this period, the length of time of which varies greatly from state to state[22]. Finally, states may differ as to whether or not deficiency judgments are allowed, whereby attachment of the borrower' s personal assets occurs. Whether or not the lender actually uses this ability to attempt a deficiency judgment depends on the expected costs and gains. In theory the option to pursue deficiency judgments should result in lower losses at least some of the time.

The final cost related factors examined are mortgage insurance, short sales, and bankruptcy. If a loan carries private mortgage insurance (PMI), then the insurance company is obliged to pay all of the losses up to some agreed-upon limit, typically

[21] These are primarily short sales, forbearance, and loan modifications.

[22] Rights of redemption may have little practical significance. One servicer in the dataset reported 8 redemptions out of 1625 foreclosed loans.

6% to 30% of the current loan balance[23]. Since the coverage amount for the loans in the dataset is unknown, the regression uses dummy indicators for PMI and missing PMI. *Short sales* are when the house is sold before the foreclosure process is finished, for an amount that is "short" of the unpaid principal balance plus accrued interest. Short sales need to be negotiated between the borrower and lender, and thus are assumed to only occur when financially optimal for minimizing severity. If a short sale does not occur, then a loan enters REO at the end of the foreclosure process, and additional months of lost interest occurs. Therefore, the coefficient on the short sale indicator is expected to be negative (i.e. indicating lower severity.) If the borrower declares bankruptcy, this extends the length of time needed to foreclose. Since the lost interest is already accounted for when estimating the model, the bankruptcy indicator only shows the net of legal fees and judicial awards.

Econometric Specification

This section discusses the specification of the econometric model. The functional form for the model was created by first specifying each of the components of loss, which was done in equation (1.2). Since the loss components in that equation are unknown beforehand, this formula is only useful for suggesting the functional form for the regression. To tie the option and cash flow variables together with this loss equation, first convert the loss equation (1.2) to severity using equation(1.1). The

[23] FHA/VA insured loans are covered for the full amount of all losses. However, none on the loans were coded as FHA/VA insured loans.

resulting equation is:

$$\text{Severity}_i = \alpha + \delta \text{*Months}_i\text{*(Monthly Lost Interest}_i\text{)*100/Unpaid Principal Balance}_i$$
$$+ \gamma\text{*Est. House Value}_i\text{*100/Unpaid Principal Balance}_i + g(r,H_i,Y_i) + \beta\text{*}X_i \quad (1.4)$$

Where g(r,H,Y) are time-varying functions of option-related variables, such as the probability of negative equity and the relative spread, r stands for the relevant interest rates, Y is a vector of variables used to estimate the option values, and *i* stands for the individual loan. $\alpha + \beta^*X_i$ is a linear function of the option theory, loan contract, cash flow, and cost variables described in the data section as defined in detail in Appendix B, and takes the place of *Unrecoverable Costs* in equation (1.2). The unknown quantities in equation (1.2) *(Current House Price)*(1 - Real Estate Commission-Fix-up Costs)* was replaced with an Estimated House Value, which is the original house value adjusted by the change in that MSA's HPI. The coefficient on Estimated House Value, γ, represents the "haircut", or discount off the expected house price, which is due to sales commissions, fix-up costs, etc.

Estimation

Estimation of the statistical model was done separately for first and second lien loans. Each type is examined in turn.

First Lien Mortgages

For first lien mortgages, severity is modeled using Weighted Least Squares. Equation (1.4) is used to suggest the functional form of the regression. δ is constrained to be 1, since Months*(Lost Interest) are known in the dataset. $\alpha + \beta * X_i$ is a linear function of the option theory, cash flow, and cost variables listed in Appendix B. To correct for heteroskedasticity, estimation was done in several steps. First regression equation (1.4) was run, then the squared residuals were estimated as a function of the variables used in the original regression that were statistically significant in forecasting the squared residuals[24]. Second, the original regression was rerun using the estimated variances (from a second regression) as weights. The F-value for the regression indicates that the combined regressors are significant at the .0001 level. The adjusted R-squared is 0.33. The results are shown below in Table 5.

Variable	**Parameter Estimate**	**Standard Error**	**t Value**	**p Value**	**Standardized Estimate**
Intercept	58.70559	8.20162	7.16	<.0001	0
Balloon	3.82075	0.54346	7.03	<.0001	0.04966
PMI	-7.29293	0.68191	-10.69	<.0001	-0.11070
Missing_PMI	-2.70474	0.81929	-3.30	0.0010	-0.04846
PMI_amt_LT50	2.87948	1.74849	1.65	0.0996	0.01069

[24] The variables used for the estimated variance are the Judicial State dummy variable and a quadratic function of relative spread, age, and loan amount.

Variable	Parameter Estimate	Standard Error	t Value	p Value	Standardized Estimate
Est_House_Value	0.00004259	0.00000593	7.19	<.0001	0.04257
change_HPI	-74.65059	14.03159	-5.32	<.0001	-0.35301
change_HPI_sq	16.04674	5.98427	2.68	0.0073	0.16732
prob_neg_eq	9.19193	5.17029	1.78	0.0754	0.03971
prob_neg_eq_sq	-23.43238	8.53693	-2.74	0.0061	-0.04915
Relsprd	4.24828	1.07532	3.95	<.0001	0.13711
Relsprd_sq	0.09519	0.02044	4.66	<.0001	0.16015
Age	-0.03992	0.02173	-1.84	0.0662	-0.03793
Age_sq	0.00047791	0.00015231	3.14	0.0017	0.05874
Investor	14.49779	0.61314	23.65	<.0001	0.15187
Short_Sale	-5.33058	0.36088	-14.77	<.0001	-0.08983
Judicial_State	5.85302	0.58438	10.02	<.0001	0.09714
Original_CLTV	21.87255	2.53855	8.62	<.0001	0.08246
OrigcltvGT80	0.35625	0.76848	0.46	0.6430	0.00519
FICO_Between_300_550	-0.33794	0.76828	-0.44	0.6600	-0.00306
FICO_Between_550_620	1.94445	0.76670	2.54	0.0112	0.01794
FICO_Greater_than_620	1.74202	0.84453	2.06	0.0392	0.01420
Orig_amt0_50k	20.30220	0.87919	23.09	<.0001	0.27514
Orig_amt50_75k	5.88672	0.87119	6.76	<.0001	0.07568
Orig_amt75_100k	-1.13082	0.91052	-1.24	0.2143	-0.01212

Variable	Parameter Estimate	Standard Error	t Value	p Value	Standardized Estimate
Orig_amt100_200k	-4.96165	0.81282	-6.10	<.0001	-0.06844
Orig_amt200_300k	-9.72156	0.82270	-11.82	<.0001	-0.11973
Orig_amt300_600k	-7.58948	0.77447	-9.80	<.0001	-0.11510
ARM	0.00192	0.46467	0.00	0.9967	0.00003441
AltA	-5.87446	1.63158	-3.60	0.0003	-0.02389
Subprime	1.01775	0.92523	1.10	0.2713	0.01853
NonJudicial_State	3.59980	0.56942	6.32	<.0001	0.06463
Non_Recourse	-3.44997	0.49271	-7.00	<.0001	-0.06194
Debt_Ratio_Between_35_40	-0.95643	0.86190	-1.11	0.2672	-0.00739
Debt_Ratio_Between_40_45	-1.66368	0.86572	-1.92	0.0547	-0.01275
Debt_Ratio_Greater_than_45	-4.49674	0.92368	-4.87	<.0001	-0.03077
Refi	1.85005	0.43456	4.26	<.0001	0.03194
Bankruptcy	-1.35092	0.61880	-2.18	0.0290	-0.01331
Servicer88	0.79430	2.75218	0.29	0.7729	0.00184
Servicer90	15.52700	1.49993	10.35	<.0001	0.07858
Servicer93	5.78045	1.45917	3.96	<.0001	0.02974
Servicer94	19.53910	1.69210	11.55	<.0001	0.08569
Servicer109	9.57076	1.00546	9.52	<.0001	0.13503
Servicer110	4.85138	0.97944	4.95	<.0001	0.06743
Servicer113	13.71045	0.70898	19.34	<.0001	0.24291

Variable	Parameter Estimate	Standard Error	t Value	p Value	Standardized Estimate
Servicer114	8.96375	1.38060	6.49	<.0001	0.05798

Table 5: Regression Results.

The regression results suggest that once a loan defaults, controlling for the above variables:

1. Subprime loans have only a slightly larger severity (1%) than prime loans, while Alt-A loans had a substantially lower severity (5.8%). Since the coupon was already factored in, this is consistent with the proper pricing of risk, assuming the probability of default was also priced appropriately.
2. Original LTV is both statistically and economically significant. Thus, as with prior research, the ruthless (no transaction costs) default theory is inadequate by itself to describe observed behavior.
3. To test if that there is an upward bias in appraised home values on refinanced loans, refinanced loans with LTV's less than 80 percent are allowed to have different discounts than loans for new home purchases (the Refi variable). There is evidence of a minor upward bias on appraised values of refinanced loans suggested by the 1.8 percent higher severity.
4. Servicers appear to vary substantially in losses, even after controlling for the included factors[25].

[25] Since the servicer's coefficients should be interpreted with a great deal of caution, the actual names of the services are kept secret. A great many factors, such as the accounting methods used, can influence severity (Kyriacou and Westerback, 1999). Different coefficients probably reflect differing reporting standards more than differing ability.

5. Bankruptcy, FICO scores and debt to income ratios have the opposite sign from expectations and are of minor economic importance.
6. ARM loans didn't appear to be any different severity than fixed rate loans.
7. As expected, states requiring a judge to foreclose on the property had 2.2% higher severity than states that don't require a judge. However, the surprising finding of lower severity in states that do not allow recourse suggests that additional important state factors are not captured in the current model specification.
8. The original loan amount is very important, with high severity rates for loans under $50,000 and greater than $600,000.

The results are also interesting for suggesting which variables are not significant. The indicator for bankruptcy is not statistically significant. During model development, some variables were tried and rejected. Seasonal dummies, dummies for single family residences, type of documentation, loan terms of less than 15 years, prepayment penalties, and paper grade dummies were removed because they were neither economically nor statistically significant.

Measures of the effects that key variables can have on severity are of interest in and of themselves. The next three figures show the effects of three variables that turn out to be among the most important. All three figures correspond quite well with intuition. The figures are graphed holding all other loan variables constant at the median value, and are graphed over the entire range of values observed in the data.

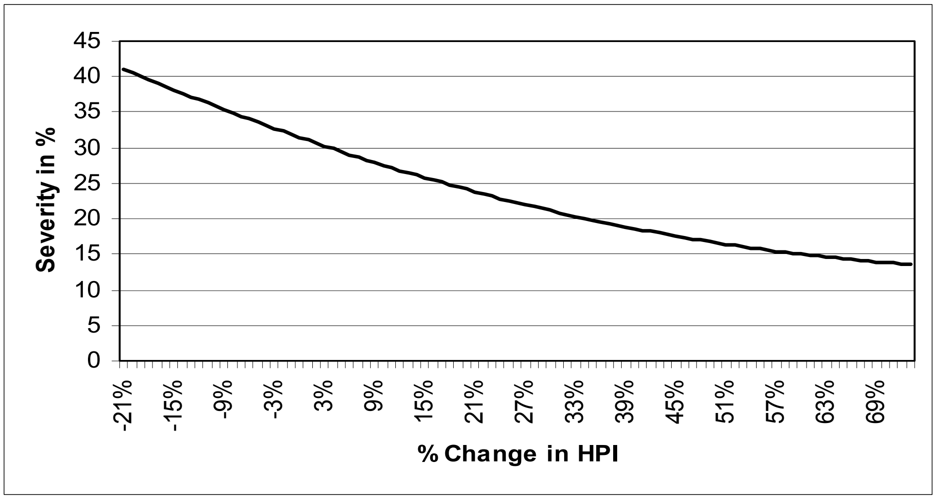

Figure 3: The Effect of Changes in Local House prices Indexes on Severity.

Figure 3 indicates that severity is lower for higher rates of house prices appreciation.

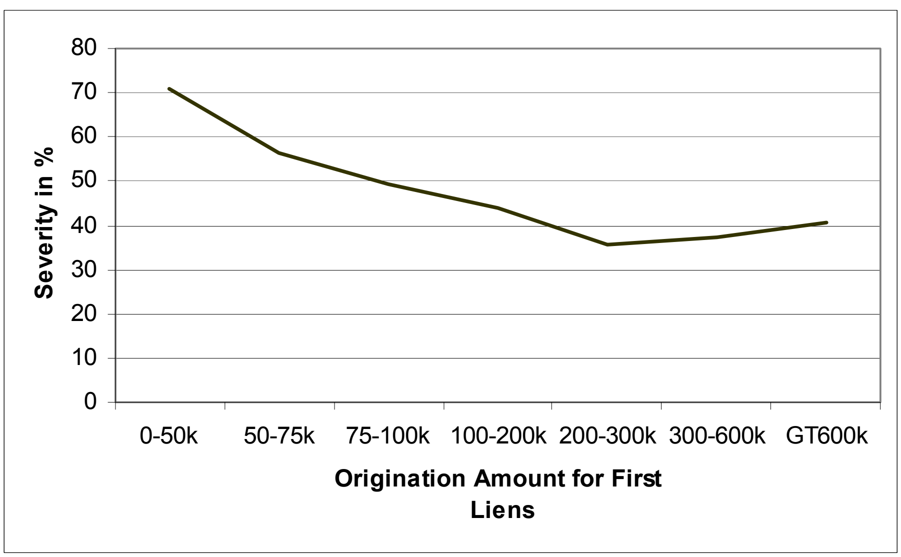

Figure 4: The Effect of Loan Size on Severity.

Figure 4 shows that the original loan amount is also very important, with high severity rates for loans with origination amounts under $50,000 and those with origination amounts greater than $600,000.

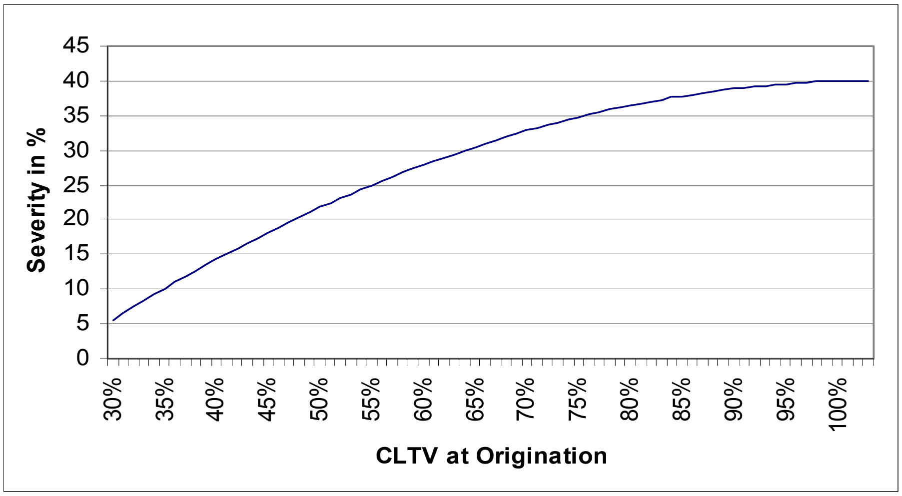

Figure 5: The Effect of the Combined Loan to Value Ratio at Origination on Severity.

Figure 5 indicates that the relationship between LTV at origination and severity is not 1 to 1. The reason that any losses are observed for low LTV loans is believed to be due to fraud or poor appraisals.

Table 6 shows the relative contribution of each category of variables as a group by showing the change in R^2 when each variable group is added to the other groups. In addition to the variable groups, lost interest explains 10% of severity.

Regressor Group	Percentage of Severity Explained
Option Theory Variables	1.8%
Loan Characteristics	2.6%
Cash Flow Variables	7.1%
Cost Related Variables	3.6%
Servicer Indicator Variables	5.5%

Table 6: Relative Effect of Variable Groupings.

Table 6 indicates that the most important information is contained in the variables related to the borrower's cash flow, while the least important are the option theory variables. Since the servicer specific effects were so large, they were separated out from the Cost Related Variables category for Table 6.

Second Lien Mortgages

A separate severity model was estimated using 561 second lien loans. These loans are estimated separately from first lien mortgages because of their substantially different nature. Besides being subordinate, second liens typically have high severity rates, low LTV and high interest rates.

Parameter estimates are estimated using SAS PROC REG with the fast backwards option, which iterates until only statistically significant variables (at the 10% level) remain. The specification was the same as for first lien loans, but only the variables shown in Table 7 are significant. The model is statistically significant (F Value = 32.51) and has an adjusted R-Squared of .37. No allowance was made from mortgage insurance since only 1 loan had mortgage insurance, 149 had no mortgage insurance, and the rest had no indication either way. The model results are presented below in Table 7:

Variable	**Parameter Estimate**	**Standard Error**	**Type II SS**	**F Value**	**Pr > F**
Intercept	275.51090	151.74774	2599.55810	2.53	0.1121
prob_neg_eq	113.90163	39.74570	8428.49446	8.21	0.0043
prob_neg_eq_sq	-87.02551	50.68848	3025.13744	2.95	0.0866
Judicial_State	0.15932	2.78988	3.34707	0.00	0.9545
Original_CLTV	-30.37382	5.64292	29735	28.97	<.0001
OrigcltvGT80	22.16914	4.41539	25872	25.21	<.0001
change_HPI	-277.93674	256.10321	1208.73853	1.18	0.2783
change_HPI_sq	110.56609	108.12633	1073.12798	1.05	0.3070
real_amt	-0.00079020	0.00018615	18494	18.02	<.0001
real_amt_sq	3.774724E-9	1.610556E-9	5637.53707	5.49	0.0194

Table 7: Results of Regression on Second Liens.

These results suggest a strong effect from changes in house prices.

Validating the Model

This sections tests the economic significance of the first and second lien severity models by comparing their forecasts to forecasts from more conventional methods. This allows us to assess the impact that considering information on the heterogeneity of mortgage characteristics can have on severity estimates. The conventional methods are an accounting equation[26] and assuming that losses are exactly equal to the median loss value in the dataset of $35,525. In order to make the test both out of time and out of sample, the severity model was first estimated without the most recent 1000 loans. Then losses on these 1000 loans were estimated using the severity model, the accounting equation, and the median. Two ways to measure the results are reported. The first method is the average of the absolute value of the differences between the actual losses and the predicted losses. The second method is the square root of the squared average differences[27]. Table 8 shows that the severity model's forecasts were between 22% to 35% more accurate than the forecasts of the other two methods.

[26] The accounting equation for first lien mortgages is the same as Equation 1.2, using the following assumptions: a 6% Real Estate Commission Rate, 10% Fix-up Costs, a 35% House Price Discount, $500 in Unrecoverable Costs, and 16 Months of Lost Interest, House Price =[Original Sales Price * HPI(at time of sale) / HPI(at time of origination)] , and Mortgage Insurance Payment = Max(0,Min [reasonable losses, Unpaid Principal Balance * MI Coverage]). For second lien mortgages, Loss = Unpaid Principal Balance + Lost Interest + Unrecoverable Cost - MAX(0, House Price - Real Estate Commission Amount - Fixup Cost Amount -1st lien loan UPB).

[27] By squaring errors, larger errors are given more weight than smaller errors.

Measurement	Statistical Model Error	Alternative 1: Accounting Error	Alternative 1: Accounting Improvement	Alternative 2: Median Error	Alternative 2: Median Improvement
Avg. Difference	$15,181	$19,677	23%	$20,207	25%
Sq. root of (SSE/N)	$21,037	$32,449	35%	$26,909	22%

Table 8: Comparing Various Loss Given Default Forecast Methods.

The first row indicates that the absolute value of the differences between the actual losses and the predicted losses were $15,181 from the statistical model and $19,677 from the accounting formula and $20,207 from the using the sample median. The improvements were also robust across servicers. The overall projections from the statistical model were not significantly biased upward or downwards. However, the accounting formula systematically underestimated errors by an average of $3208.

Conclusion

The severity model produced plausible coefficients that can be used by financial institutions and regulators for forecasting losses on defaulted loans. The results can be divided into two groups based on theoretical or empirical relevance. On the theoretic side, evidence against the 'ruthless' default hypothesis was found in that many non-option theoretic variables were significant. In addition, the inclusion of additional variables appears to greatly diminish the effect of the option theory variable original LTV. The most important theoretic implication is the discovery for the first time of evidence of a minor systematic upward bias in appraisals on refinanced loans. While this may be consistent with other theories, it does suggest further investigation of the moral hazard implications of the current industry structure is warranted. There are several empirical results of interest to regulators and market participants. The most important determinants of losses were found to be the variables related to costs and the servicers. Subprime, and ARM loans do not appear to result in larger loss severities than prime fixed rate loans, while Alt-A loans have a lower severity once the other variables in the model are controlled for. However, investor owned properties and balloon loans have significantly higher severities. Another empirical contribution to the literature is reporting for the first time how frequently loans that pay off from various loan payment states result in losses (see Appendix C).

Moving from a constant severity estimate to one based on loan and borrower characteristics represents a major advancement in risk management. However, the severity model developed here relies on information known at the point of origination and the change in interest rates and housing prices between the loan origination and the month of default. If a distribution of possible future interest rates and housing prices were somehow combined with this model, it could produce a distribution of future possible losses for each loan. That would represent a leap forward in risk measurement and analysis, and undoubtedly is the direction that sophisticated financial institutions will be moving in the future.

Chapter 2: Valuing Mortgages Using Mean Reverting House prices

Introduction

Uncertain cash flows from the $5.2 trillion in outstanding residential mortgages represent a major source of risk for large financial institutions[28]. Understanding the drivers of mortgage termination is very important to the development and specification of pricing models for mortgages and mortgage-backed securities. However, until now option theory based mortgage valuation models have assumed that house prices follow a geometric Brownian motion process[29]. This chapter extends existing models by allowing house prices to follow a geometric mean reverting process. The primary goal here is to investigate the implications of a mean reverting house price process on mortgage values, and to compare the results to those from a model assumes a geometric Brownian motion house price process in Kau et al. (1992, 1995).

This chapter bridges an important disconnect between the theoretic and empirical mortgage literature that has not previously been studied. All mortgage theory, such

[28] The figure is from the Flow of Funds, published by the Federal Reserve, and is for the United States at the end of 2000.

[29] See Kau and Keenan 1995 for a literature review.

as the seminal Kau et al. (1992,1995) papers, assume that house prices follow a random walk[30]. Empirical evidence, however, suggests that house prices are poorly approximated by a geometric Brownian motion process, and instead are more consistent with a mean reverting process (Englund and Ioannides (1997), Meese and Wallace (1997, 1998), England, Gordon, and Quigley (1999)). Microeconomic theory also provides strong support for the notion that the price of a good (in this case housing) in a competitive market is mean reverting, due to supply and demand responses to prices. For example, an increase in demand increases prices until a supply response drives prices back down to the production cost plus a normal economic return. Since this is such a fundamental story in economics, it is simply assumed as a starting point in this chapter.

By creating a theoretic model based on mean reversion, this chapter presents a model that is both more plausible and useful. The model presented here also suggests additional factors that may help forecast mortgage termination that have been overlooked in the literature. More specifically, the new model implies that the following aspects of local house prices matter for mortgage valuation: (1) deviations from the long run expected house values, and (2) the rate of mean reversion in the local housing market[31].

---

[30] For expositional purposes, this paper often refers to a geometric Brownian motion process as a random walk.

[31] Consistent with this theory, geographic region was shown empirically to matter to default by Lekkas, Quigley, and Van Order (1993).

In addition to creating a more realistic mortgage valuation model and suggesting additional factors that can be used for forecasting, this research may also have implications for the wider finance literature. This is done by illustrating that the service flow needs to be derived as a function of the mean of the underlying process. The traditional assumption that the service flow is a constant percentage of the value of the house turns out to be equivalent to assuming that the underlying process follows a random walk. In this context, the service flow is a value of living in the house, while in another context it might be a dividend. The assumption of a constant proportion service flow may be reasonable for assets that can be modeled with a random walk, but is shown to be a misspecification if the underlying asset is mean reverting, as may be the case with most real options[32]. Most mortgage models *assume* that the service flow is a constant percentage[33], even though it is shown here to be mean reverting, and thus not constant.

This chapter draws heavily upon financial options theory to explain borrower prepayment and default behavior. The value of a mortgage equals the value of the remaining payments plus the value of the prepayment and default options. Default is viewed as a put option that gives the borrower the right to sell the house to the lender

[32] While some papers in the real options literature assume a mean reverting process, the majority of the literature assumes that the underlying process follows a random walk. See Trigeorgis (1996) for a review.

[33] Examples include Hendershott and Van Order (1987), who dismiss service flows by simply stating that it is difficult to specify, and Downing (1998).

at a price equal to the outstanding balance plus any accrued interest. Rational borrowers will exercise options when they can increase their wealth. Absent either transaction or reputation costs, wealth can be increased by defaulting when the market value of the mortgage exceeds the value of the house[34]. Similarly, by prepaying when value of the mortgage exceeds par, borrowers can increase their wealth by refinancing. The problem of determining when to exercise an option requires specifying the underlying state variables, and deducing the rule for exercising those claims that maximizes the borrower' s wealth. Since prepayment involves consideration of the term structure and default value of the house, the key state variables are house prices and interest rates. The solution method is to derive analytically a partial differential equation (PDE) that describes how the option values evolve over time, and then to find a numeric solution to the PDE. The same process is followed for two models: one where house prices follow a geometric Brownian motion process, and one where house prices follow a mean reverting process. Apart from the house price process, both models use the same continuous time framework and assumptions, such as no arbitrage and a Cox, Ingersoll, and Ross (1985) interest rate process.

The reason mean reversion influences the value of a mortgage is that the value of living in a house (hereafter called the service flow) turns out to also be mean reverting. That is because the required rate of return on an investment in a house is

---

[34] This is not to say that transaction costs don't matter, or shouldn't be modeled. They are only omitted to get clean theoretic results and to be consistent with Kau et al. (1992).

composed of the service flow and the expected capital gains, which is mean reverting by assumption in the model developed here. The required rate of return on the investment in a house is exogenously determined by the equilibrium rate of return on similar investments. Since the required rate of return is externally determined independently of the house price process, the service flow is *by definition* mean reverting[35]. Thus, the parameters of the mean reverting process still influence a mortgage's value through the service flow.

The default option is more valuable under the mean reverting model than the random walk model if house prices are substantially above the long-term trend, but less valuable if house prices are substantially below the long-term trend. Intuition for why the option values differ across the two models comes from thinking about what happens if the house value declines below the value of the outstanding balance. If house prices follow a random walk, then the house value is just as likely to continue to sink as to recover. In the mean reverting model, the option value depends on whether or not the current house value is above or below the long-term trend house value. If the house value is below its long-term expected value, then the value of the house is expected to increase in the mean reverting model, thus making the option less valuable. Likewise, if the house value is above the long-term expected value, then the value of the house is expected to decline in the mean reverting model, thus making the default option more valuable.

[35] The same insight is described in Dixit and Pindyck (1993) in the context of a real option.

This papers is divided into the following sections: section 2 describes the specification of the geometric Brownian motion model, section 3 describes the mean reverting house price model, section 4 describes the derivation of the model dynamics, section 5 explains the numeric solution methodology, section 6 describes a simulation results, and section 7 summarizes the research and discusses future areas of research. The appendix describes the derivation of the PDEs.

The Geometric Brownian Motion Model

The problem of determining when to exercise the option requires specifying the underlying state variables. In this case, the state variables are the risk-free interest rate and the rate of housing appreciation. The set-up comes from Kau et al. (1992), and assumes that the processes describing the environment relevant for a mortgage are:

$$dr = \gamma(\theta - r)dt + \sigma_{\mathrm{r}}\sqrt{r}dz_r \qquad (2.1)$$

$$\frac{dH}{H} = (\alpha - s)dt + \sigma_H dz_H \qquad (2.2)$$

$$dz_r dz_H = \rho dt. \qquad (2.3)$$

Where

H = the house price

r = the risk-free interest rate

θ = the long term mean of the interest rate

γ = the rate of mean reversion in interest rates

ρ = the correlation of the disturbances to the house price and the interest rate

s = the *constant* rate of service flow, or value of implicit rent, from the house

σ = the instantaneous standard deviation for the relevant processes

Equation (2.1) indicates that the interest rate is expected to change at any time t at the rate $\gamma(\theta\text{-}r)$. In other words, on average, the interest rate r converges towards the value θ at the rate γ. The actual changes in house prices and interest rates differ from their expectations because of disturbances to the underlying processes, which are assumed to be normally distributed, and serially uncorrelated. The return to owning a house consists of the price appreciation and the service flow, s from using the house. In this particular model, the service flow is assumed to be proportional to the value of the house.

The primary goal is to compare the implications of mean reversion in house price process, to the implications of the Brownian motion model in Kau et al. (1995). Therefore, attention is restricted to financially motivated prepayment and default only and complete financial markets and rational agents are assumed. Thus, default only occurs at the end of each month, when payment is due. Prepayment can occur at anytime, for low enough interest rates and high enough house values (the default option is worth less, the higher the current house value). For simplicity, transactions costs and mortgage insurance are assumed away and only fixed rate fully amortizing contracts are examined. Since the focus here is only on how the mean reversion parameters influence the option values, those interested in the effects of points, mortgage insurance, and changes in the term structure should consult Kau et al. (1995).

The Mean Reverting House Prices Model

The mean reverting model uses the following setup:

$$dr = \gamma(\theta - r)dt + \sigma_r \sqrt{r} dz_r \qquad (2.4)$$

$$\frac{dH}{H} = \eta(\overline{H} - H)dt + \sigma_H dz_H \qquad (2.5)$$

$$dz_r dz_H = \rho dt. \qquad (2.6)$$

Where the variables not defined earlier are the following:

η = the speed of mean reversion in house prices

$\overline{H}$ = the long run mean of house prices[36]

One difference from the geometric Brownian motion model is that the return on housing reverts at speed η to its long run average $\overline{H}$. The other difference

---

[36] $\overline{H}$ is assumed to be constant for simplicity, but can be generalized to be a function of economic variables.

is that the service flow is a mean reverting function instead of a constant. This is because the required return on the investment in the house, μ, is composed of the service flow, *s,* and the expected capital gains:

$$\mu = s + \eta(\overline{H} - H) \tag{2.7}$$

Rearranging, this gives the only difference between the geometric Brownian motion and mean reversion PDEs (shown in the next section):

$$s = \mu - \eta(\overline{H} - H) \tag{2.8}$$

According to CAPM, the required return on the investment in the house, μ, should only reflect the asset's non-diversifiable risk. The key difference is that the service flow, *s*, is now a function of the house value, *H*, whereas in the geometric Brownian motion model it is constant. Unfortunately, this replaces the unknown parameter *s* with the unknown parameters $\eta, \overline{H}$ and μ. The parameters can be empirically estimated, but this chapter instead gives simulation results for a wide range of parameter values.

Derivation Of Model Dynamics

In this section the PDE that describes how the option values evolve over time is derived. This PDE provides insights into the differences between the models, and is used to generate the numeric solutions.

The Local Expectations Hypothesis[37] is sufficient to derive an equilibrium condition for the value of a mortgage, *M* (details are given in Appendix D). Two different second-order partial differential equations are derived, one assuming a random walk house price process and one assuming a mean reverting house price process. The value of any derivative asset *X(H,r,t)* under the random walk set-up has the following fundamental PDE:

$$\begin{aligned} &Pay+\frac{1}{2}H^2\sigma_H^2\frac{\partial^2 x}{\partial H^2}+(r-\mathrm{s})\mathrm{H}\frac{\partial x}{\partial H}+\frac{1}{2}r\sigma_r^2\frac{\partial^2 x}{\partial r^2}+ \\ &\gamma\ (\theta-r)\frac{\partial x}{\partial r}+\rho H\sqrt{r}\sigma_H\sigma_r\frac{\partial^2 x}{\partial r\partial H}+\frac{\partial x}{\partial t}=rX \end{aligned} \tag{2.9}$$

[37] The local expectations hypothesis is that the expected change in the value of a pure discount bond is simply equal to the instantaneous interest rate. It essentially serves to prevent there being any risk premia in the term structure. See Cox, Ingersoll, and Ross (1981). This condition follows from no arbitrage.

Where *Pay* is the monthly mortgage payment[38]. The mean reverting house price model has the following fundamental PDE:

$$Pay+\frac{1}{2}H^2\sigma_H^2\frac{\partial^2 x}{\partial H^2}+(r-\mu+\eta(\overline{H}\;-\;H))H\frac{\partial x}{\partial H}+\frac{1}{2}r\sigma_r^2\frac{\partial^2 x}{\partial r^2}+ \\ \gamma(\theta-r)\frac{\partial x}{\partial r}+\rho H\sqrt{r}\sigma_H\sigma_r\frac{\partial^2 x}{\partial r\partial H}+\frac{\partial x}{\partial t}=rX \tag{2.10}$$

The only difference between the two model's PDEs comes from the fact that in the mean reverting model:

$$\mathrm{s}=\mu-\eta(\overline{H}-H) \tag{2.11}$$

The mean reverting model suggests additional factors that may help forecast mortgage terminations that have been overlooked in the literature. More specifically, the new model implies that the following aspects of local house prices matter for mortgage valuation: (1) deviations from the long run expected house values, (2) the rate of mean reversion in the local housing market, and (3) the required rate of return on housing.

---

[38] Pay = $\left[\dfrac{(c/12)(1+c/12)^{n}}{(1+c/12)^{n}-1}\right]*LoanAmount$ where c is the coupon rate on the mortgage.

Numeric Solution Methodology

This section describes the numerical solution method used to do the comparisons of the two models. The solution method is to start at a point in time where the details of the mortgage contract completely determine the option value. In this case, this is at the time of the final month on the contract. A discrete time version of the PDE that describes how the mortgage value evolve over time is used to calculate values at each point in time back to the time of origination. This is done at discrete intervals in a 3-D grid of house prices, interest rates and time. The following definitions are used in describing the terminal conditions:

M[H, r, t]   = A(r,t) - O(H,r,t) = the value of the mortgage

D(H,r,t) = default option

C(H,r,t) = prepayment option

O(H,r,t) = C(H,r,t) +D(H,r,t) = the joint prepayment and default option

A(r,t)   = value of the remaining promised mortgage payments

UPB   = unpaid principal balance

n   = term of the loan in months

Pay   = monthly mortgage payment

${}^{-}t$ = instant before a payment is due

t^{+} = instant after a payment is made

At the time the final payment is due the borrower will either default if the value of the house is below the value of the final payment or payoff the loan[39]:

$$M[H, r, 0] = \min[\textit{Final Pay}, H]$$

The default decision right before each payment is due is to default if the house is worth less than the mortgage:

$$M[H, r, {}^{-}t] = \min(M[H, r, t^{+}] + Pay, H)$$

The boundary conditions at each point in time are that for infinitely high house prices, the value of the default option goes 0, and that the mortgage becomes worthless if the house value falls to zero, or if interest rates go to infinity:

$M[\infty, r, t]$ = Value of an option free mortgage (or $\partial M[\infty, r, t]/\partial H = 0$),

$M[0, r, t] = 0$,

$M[H, \infty, t] = 0$,

The lower boundary conditions are:

$C(0,r,t) = 0$

$D(0,r,t) = A(r)$

---

[39] The transactions costs can easily be added here, but are assumed to be 0 without loss of generality.

Prepay occurs at any time if:

M[*H, r, t*] >UPB + accrued interest

Even though it is well known that the Black-Scholes formula is independent of the mean drift, this does not mean that the option price is independent of the drift process. This apparent paradox was resolved by Lo and Wang (1995), who showed how a mean reverting process has a different estimate of the variance than a geometric Brownian motion process. Thus, if expected returns are time varying, this predictability must be taken into account in estimating the variance[40]. They show that an option's value increases as the rate of mean reversion increases. Unlike in Lo and Wang (1995), mean reversion matters here because the service flow is mean reverting.

Simulation Results

This section compares the optimal prepayment and default boundaries from the two models and investigate how sensitive the models are to the parameter values. It is shown that the options are usually, but not always substantially more valuable in the mean-reverting model. To use the models, estimates of the parameters of the mortgage value processes are required. Table 1 presents the base values of the

[40] For an excellent treatment on this topic, see pages 369-377 of Campbell, Lo, and MacKinley (1997).

parameters used in the numerical solutions. These values are assumed to hold unless otherwise stated.

Symbol	Description	Value	Sign of Effect on Value
H	Starting house value	$100,000	
$\overline{H}$	Long-run mean house value	$100,000	
	Loan amount	$90,000	
	Term in years	30	
	Years remaining	30	-
σ_H^2	House process volatility	0.15	+
r	Starting interest rate	0.07	
θ	Mean interest rate	0.07	
η	Rate of house price reversion	0.0003	-
μ	Required rate of return on similar investments	0.08	+
σ_H^2	Interest rate process volatility	0.15	+
ρ	Correlation of the house price and interest rate	0.02	+
γ	Interest rate speed of reversion	0.1	
S_{rw}	Fixed service flow in random walk model	0.085	-
c	Coupon, i.e. the mortgage interest rate	0.075	

Table 9: Base Simulation Values[41].

The last column shows what an increase in a parameter value does to the value of the prepayment and default options, and thus to the value of a contract to the borrower. The sign of some parameter effects can be easily understood from recalling that the default option is akin to a put option and the prepayment option is akin to a call

[41] See Kau, et al. (1995) for justifications of the default parameter values. Titman and Torous (1989) estimate the correlation coefficient to be between -.01 and -.03.

option. For example, the larger the variance, and the longer the time until expiration, the greater the option value. Values listed in the following tables are for the options. The value of a mortgage to the mortgage holder is the present discounted value of the payments minus the option values. Since Kau, et al. (1995) show how the option values change for changes in parameters in the random walk model, the focus here is only on how the mean reversion parameters influence the option values.

Figure 6compares how the two models estimate the value of the default option based on the current house value.

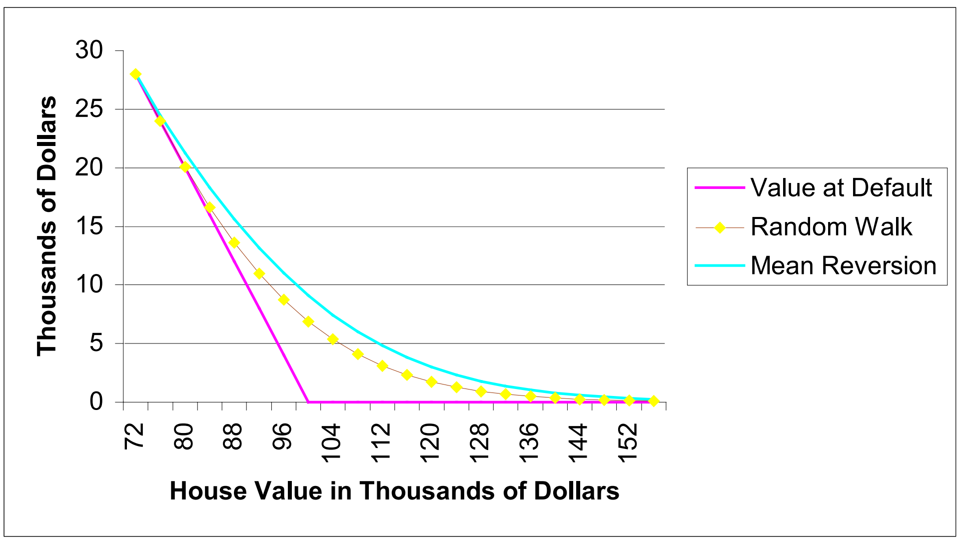

Figure 6: Comparing the Value of the Default Options for the Two Models.

Figure 6 shows graphically that the default option will be executed at higher house values in the random walk model. At one year to expiry the optimal default point for the random walk model is when the house value is 80% of the outstanding loan balance, while for the mean reversion model, it is where the house value is 72% of the outstanding loan balance. Table 10 shows that the options decrease in value as the years remaining decreases in both models.

Years To Go	**Option Value Random Walk**	**Option Value Mean Reversion**	**% Difference**
30	$9,121	$8,025	-12.0%
29	$8,567	$7,495	-12.5%
28	$7,990	$6,946	-13.1%
27	$7,392	$6,381	-13.7%
26	$6,774	$5,803	-14.3%
25	$6,141	$5,216	-15.1%
24	$5,497	$4,623	-15.9%
23	$4,848	$4,032	-16.8%
20	$2,951	$2,349	-20.4%

Table 10: Value of Prepayment and Default Options by Time to Maturity.

Table 10 shows that for the based parameter values, the options were worth less in the mean reverting model. The next table shows the value of the default option for the two models for different current Loan-to-Value (LTV) ratios.

LTV	**Option Value Random Walk**	**Option Value Mean Reversion**	**% Difference**
0.6	$900	$545	-39.5%
0.7	$2,027	$1,427	-29.6%
0.8	$3,720	$2,915	-21.6%
0.85	$4,833	$3,954	-18.2%
0.9	$6,141	$5,216	-15.1%
0.95	$7,664	$6,720	-12.3%
1.01	$9,789	$8,864	-9.5%
1.1	$13,651	$12,834	-6.0%
1.25	$22,011	$21,529	-2.2%

Table 11: Value of Prepayment and Default Options for Differing LTV Ratios.

Note that for lower LTVs, the default option value is actually worth less for the mean reverting model than for the random walk model. This results from the fact that in the mean-reverting model, the probability of house prices dropping so far as to wipe out all of the equity in the house is much less than in the random walk model, provided that the house value is substantially larger than the outstanding balance.

Tables 4 and 5 show the value of the default option in the mean reversion model for various values of the rate of house price revision and house price volatility.

Speed of House Price Reversion	Option Value
0	**$7,458.80**
0.0001	**$7,362.13**
0.0002	**$7,271.29**
0.0003	**$7,185.09**
0.0004	**$7,102.72**
0.0005	**$7,024.57**
0.0006	**$6,950.32**
0.0007	**$6,878.67**
0.008	**$4,414.53**

Size of House Price Volatility	Option Value
0.01	**$0.00**
0.05	**$0.12**
0.1	**$459.60**
0.151	**$2,205.10**
0.2	**$4,623.01**
0.25	**$7,366.01**
0.3	**$10,104.19**
0.35	**$12,668.80**
0.4	**$14,991.30**

Table 12 and Table 13: Prepayment and Default Option Values For Various Parameter Values in the Mean Reverting Model.

If the speed of mean reversion differs across the regions, it is the local rate of revision that matters. Comparisons are based on ranges around values estimated in the empirical literature[42]. Table 14 shows how sensitive the mean reverting model is to the required rate of return.

[42] Abraham and Hendreshott (1996) estimated an annual rate of mean reversion of -.151 for 30 cities from 1978 to 1992. Meese and Wallace (1997) found a much higher rate of adjustment for Paris of 30 percent per month.

Required Rate of Return	Option Value in Mean Reversion Model
0.08	**$5,421.20**
0.09	**$4,598.71**
0.1	**$6,533.49**
0.11	**$8,612.82**
0.12	**$10,759.88**
0.13	**$12,922.87**
0.14	**$15,069.23**
0.15	**$17,176.71**
0.16	**$19,232.50**
0.17	**$21,228.26**

Table 14: Prepayment and Default Option Values For Various Required Rates of Return in the Mean Reverting Model.

Conclusion

A new mortgage prepayment and default model is presented that extends mortgage option theory to cases where the house price process is mean-reverting. This chapter compares theoretic option values for a geometric Brownian motion and a mean reverting house price model for various parameter values. Existing mortgage valuation models assume that house prices follow a geometric Brownian motion process, despise strong empirical evidence in favor of mean reversion. The key difference between two models is that the service flow, or rental value, is constant if you assume a random walk, while it is a mean reverting function if house prices are mean reverting.

This chapter makes several contributions. It bridges an important disconnect between the theoretic and empirical mortgage valuation literature that has not previously been studied. The model presented here also suggests additional factors that may help forecast mortgage termination: deviations from the trend house prices, the rate of mean reversion in the local housing market, and the required rate of return on investing in the particular house under consideration. This research suggests that the value of the default option on a house may be substantially different than previously estimated, once mean reversion is taken into account. By showing that mean reversion influences the value of a derivative asset through the service flow, this chapter calls into question the real options theory literature that assumes a constant service flow.

It is hoped that this work will inspire research on several fronts. By showing that mean reversion matters, perhaps additional work will be done on estimating housing price indexes assuming a mean reverting process. Future extensions could include empirical tests of the additional factors relating to mortgage termination suggested by the mean reversion model. Future work could also focus on valuing mortgages when the "no arbitrage" condition fails to hold, and adding mortgage insurance to the mean reversion model. Another possible extension is to redo the model in discrete time[43]. Lastly, further research on the house price process is warranted. The true house price process is undoubtedly far more complex than the one modeled here. For example, house prices may behave more like an ARMA process, exhibiting positive short-term correlation and negative long-term correlation, or only exhibit mean reversion above

---

43 While prepayment decisions are assumed to be made continuously, this may not be the case in reality. Empirical estimates of the prepayment model in Stanton (1995) suggests that the average time between successive prepayment decisions is more than a year.

and below some threshold. Option values produced by the methodology outlined in this paper may be more accurate if a more realistic process is specified in place of the mean house value $\overline{H}$.

Chapter 3: Modeling Subprime Mortgage Prepayment and Default Using the Monthly Payment Status

Introduction

This paper presents a major extension to traditional empirical loan level mortgage modeling and compares it to more traditional modeling methods. The primary goal of both the three state and seven state models is to estimate the probabilities of prepayment and default. One reason the extended model is to allow for the information on the current monthly payment status to affect the probabilities of prepayment and default. An additional advantage of the seven state model is that it forecasts the probability of serious delinquency (90 and F). The percentage of loans that are seriously delinquency often has an impact on the way cash flows are divided up among the tranches of a mortgage backed security.

This paper focuses exclusively on the under researched subprime mortgage sector [44]. There are two reasons for this. First, subprime loans are far riskier then traditional mortgages and yet have expanded rapidly in market share. Second, since subprime loans are more likely to be delinquent, the gain from modeling additional payment states is likely to be greater for these loans.

[44] Subprime loans are defined as loans to borrowers with poor credit histories.

The extended model empirically outperforms traditional Markov mortgage transition models, which only allow a loan to be in one of two or three states: Active, Defaulted or Paid Off[45]. The extension proposed here is to allow a loan to transition to one of seven payment states: Current, 30 Days Late, 60 Days Late, 90+ Days Late, Foreclosure, REO (*Real Estate Owned*, which is a category for the time between when the loan originator takes legal ownership of a house and when they sell it), and Paid Off. The model uses data on loan features, the economic environment, and borrower characteristics to calculate the probability of a loan transitioning from one status to another. The probabilities are recalculated each month using the new loan age, payment status, and market interest rates. The extended model is compared to two types of traditional mortgage models: (1) models that limit a loan to be in one of two states (Active and Defaulted or Active and Paid Off), and (2) models that limit a loan to be in one of three states (Active, Defaulted or Paid Off). The models are compared with goodness of fit measures, statistical tests and by out-of-sample forecasts. The restrictions implied by the three-state model are rejected statistically. To do the empirical comparisons, four different statistical models were created using the same functional form and data, varying only in the number of states modeled.

The importance of the improved modeling method proposed here stems from the enormous size of the primary and secondary mortgage markets. The inability to properly measure credit risk embedded within these mortgages has been a major

---

[45] One example is the model described by Smyth, Sanchez, Lawrence (1996), which is used by a "large Californian institution." Another is the model presented in Deng, Quigley and Van Order (2000).

factor in bank failures[46]. Therefore, the primary goal of this study is to offer an improved modeling methodology for predicting prepayments and defaults on residential mortgages.

Models that treat prepayment and default separately are still commonly employed in financial institutions. This is due to the fact that different departments within financial institutions have different responsibilities. Prepayments are a concern of those charged with hedging interest rate risk, while defaults are a concern of those charged with pricing and setting capital reserves. The need for a better understanding of the differences between combined and separate models of risk was pointed out by the Federal Reserve System Task Force On Internal Credit Risk Models (May 1998).

The seven state model proposed here bases monthly payment states on the number of days since the last payment[47]. Table 15 summarizes the states used in the expanded model, and introduces some notation.

[46] Federal Reserve System Task Force On Internal Credit Risk Models (May 1998).

[47] These definitions correspond to those used by the Mortgage Banker Association.

Payment Status	**Notation**	**Definition**
Current	C	Loan payments are caught up.
30 Days Late	30	Loan is 30 days delinquent.
60 Days Late	60	Loan is 60 days delinquent.
90+ Days Late	90	Loan is 90 or more days delinquent.
Foreclosure	F	Loan is in the foreclosure process.
REO	REO	Lender has taken over the property at the end of the foreclosure process. The loan remains in REO until it is sold.
Paid Off	PO	The loan is fully paid off. This is a terminal, or absorbing, state.

Table 15: Payment States in the Seven State Model.

The vast majority of prepayment and default models can be thought of as two or three state models. Table 16 defines the payment states and notation used in traditional two and three state models.

Payment Status	Notation	Definition
Active	A	Loan account is still on the books. $A \in \{C, 30, 60, 90, F\}$
Default	D	Loan ends due to borrower default. In this paper, this is equivalent to a loan entering REO[48], since a loan cannot become current after that.
Paid Off	PO	The loan is fully paid off. This is a terminal, or absorbing, state.

Table 16: Payment States in the Two and Three State Models.

The seven state model can be thought of an expanded version of a combined prepayment and default (or three state model) model. Several papers in the literature emphasize the need to jointly model prepayment and default (Kau et al. (1992, 1995), Kau and Keenan (1996) and Titman and Torous (1989)). Quoting from Kau et al. (1992), "Since prepayment and default substitute for one another, contracts with only one of the default or prepayment provisions lead the borrower to behave differently than when both are present. This substitution effect means that one cannot accurately value either the individual provisions or their interaction without both options being present." This is because the probability of prepayments or default is a function of

[48] This paper does not examine the possibility of short sales, which is when there are losses from loans that do not enter REO.

the extent to which the other option is in the money. Deng, Quigley, and Van Order (2000) and Lundstedt (1999) echo the importance of joint prepayment and default modeling. They do this by demonstrating the statistical significance of the joint options by finding that the variables that proxy for the prepayment option are significant for forecasting defaults, and that the proxies for the default option are significant in forecasting prepayments.

The only paper the author is aware of that expands the number of loan states beyond three is Smith, Sanchez, Lawrence (1996). They constructed a quarterly Markovian forecasting model with four states: Current to 60 Days Late, 90+ Days Late, Payoff, and Default. The seven state transition model presented here extends this framework by moving from quarterly to monthly intervals, and by going to a finer level of loan payment status detail.

In comparison with these previous studies, this paper makes a number of contributions to the mortgage termination literature. This paper is unique in focusing on the theoretic, statistical and empirical differences due to the number of payment states modeled. This paper is also unique in proposing and testing an expansion of the number of mortgage states modeled from three to seven. The one drawback to the extended seven state model is that a huge number of loans are needed to accurately model each transition. While it may seem obvious that taking advantage of the current payment status will produce superior forecasts, there is no guarantee that in finite samples the greater precision will not be swamped by the greater noise inherent in using smaller samples to estimate each transition. Hence the need for careful theoretic and empirical examinations of this topic.

Also discussed are the theoretic failings of traditional prepay only models. Since defaults are treated as prepayments in a prepayment only model, this diminishes the ability to use loan characteristics such as the Loan-to-Value ratio (LTV), and the borrower's FICO credit score. Since the influence of variables such as current LTV should have ***opposite*** effects on the probability of prepayment and default, by estimating a single model you essentially average these two effects, greatly diminishing the forecasting value of the loan, borrower and economic characteristics. This may cause some useful variables to be dropped from the model, since their statistical significance may be weakened to the point where they appear to have no effect. This is particularly true for subprime loans, where the probability of default is non-trivial. Thus the two state model will give incorrect forecasts when applied to any data set which has a different ratio of prepayments to defaults then the ratio of the data from which the model was estimated.

This paper also identifies a major drawback of three state models, relative to seven state model. The three state models force the probability of making a transition to be the same across many different payment statuses (for example, the probability of going from Current to Payoff is forced to be the same as the probability of going from 60 Days Late to Payoff.) This restriction is tested and rejected in this paper. In other words, the three state model loses or ignores the information contained in the monthly loan payment status. traditional three state models is that forecast error is introduced if the proportion of loans in each payment status (Current, 30 Days Late, etc.) is different between the data used to estimate the parameters and the data to be forecast.

This chapter is organized as follows: section 2 describes the data, section 3 describes the mortgage models, section 4 is on the regression methodology, section 5 describes the statistical comparison and testing of the models, and section 6 concludes. Appendix F describes the data used in the empirical comparisons, and the Appendix E, describes the regressors used in the statistical models.

Data Description

The data comes from LoanPerformace's[49] Securities database, which is unique in size and accuracy. Of the dataset's 3,879,913 non-agency, securitized prime and subprime loans 1.1 million subprime loans were used in this study. Only first lien loans were used. The data is from January of 1990 to June 2000. Loans were filtered and crosschecked for accuracy. Therefore, to estimate the transition models only a subset of loans were used due to missing or incomplete data. Loans with missing data or suspicious values for the following fields were excluded from this study: combined LTV, lien type, interest rate, origination date, origination amount, and term. Of the remaining approximately one million loans, all transitions were used in the estimation

[49] Dan Feshbach graciously provided access to the data.

process, with the exception of current to current transition where 1 out of 100 sampling was done.

To illustrate how likely various transitions are for subprime loans, Table 17 shows the empirical transition probabilities from the data used in the parameter estimation. Each cell in these tables has the percentage of the loans in the specified FROM state in one month that went to the specified TO state in the following month. The sum across each row is 100%. The tables illustrates which transitions were most prevalent, as well as which transitions were so rare, if not impossible, that they were ignored in the parameter estimation process (these are indicated as 0.0%).

FROM	**TO**						
	Current	**30**	**60**	**90**	**Foreclosure**	**REO**	**Paid**
Current	94.1%	3.6%	0.0%	0.0%	0.0%	0.0%	2.2%
Late 30	35.2%	47.2%	13.3%	0.0%	0.6%	0.0%	3.8%
Late 60	19.7%	21.0%	21.0%	26.6%	7.8%	0.0%	4.0%
Late 90+	7.1%	2.3%	2.7%	69.9%	15.2%	0.6%	2.3%
Foreclosure	5.3%	0.6%	0.1%	4.3%	83.1%	4.1%	2.5%
REO	0.0%	0.0%	0.0%	0.0%	0.0%	12.1%	87.9%
Paid	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	100%

Table 17: Empirical Transition Probabilities for Subprime Fixed Rate, First Lien Loans.

The table above helps illustrates why the seven state model makes more intuitive sense than a two or three state model. Note that the probabilities of payoff and default depend on the current payment status. For example the probability of paying off next month when a loan is 60 days late is nearly twice as high as when a loan is current. Likewise, the probability of defaulting (entering REO) in the next month is 0 when a loan is current, but is 4.1 percent when a loan is in foreclosure.

Mortgage Models

This section details the formal structure of the models examined in this paper. These models are summarized in Table 18.

Model Name	**Number of Loan States**	**Possible Loan Payment States**
Prepayment Only	2	Active, Paid Off
Default Only	2	Active, Defaulted
Combined Prepayment and Default	3	Active, Paid Off, Defaulted
Seven State or Expanded Model	7	Current, 30 Days Late, 60 Days Late, 90+ Days Late, Foreclosure, REO, Paid Off

Table 18: Mortgage Models Examined in This Paper, in Order of Complexity.

Most mortgage models can be thought of as Markov transition models. To describe these models formally, we need define to some notation.

$q_t^{i\text{-}j}$ = the probability of a loan at time t going from state i to state j,

where i, j $\in$ {A (active), D (default), PO (paid off)} for the three state model and

i, j $\in$ {C (current), 30 (30 days late), 60 (60 days late), 90 (90 or more days late), F (foreclosure), REO, PO (paid off)} in the seven state model.

Default is defined differently by different researchers, but is used here to indicate when a loan is unrecoverable, (i.e. when it enters REO). For the three state model, the transition probabilities can be put into the following matrix form:

From\To	**Active**	**Default**	**Paid Off**
Active	$q_t^{A\text{-}A}$	$q_t^{A\text{-}D}$	$q_t^{A\text{-}PO}$
Default	0	1	0
Paid Off	0	0	1

Table 19: Transition Probabilities in the Three State Model.

The monthly payment status a loan transitions *from* is listed on the left-hand side of the matrix. The monthly payment status a loan transitions *to* is listed along the top of the matrix. Note that once a loan defaults or pays off its probability of staying in that payment status is one. For the seven state model, the monthly mortgage payment status transition probabilities take the following form:

From\To	Current	30	60	90	Foreclosure	REO	Paid Off
Current	$q_t^{C\text{-}C}$	$q_t^{C\text{-}30}$	0	0	0	0	$q_t^{C\text{-}PO}$
Late 30	$q_t^{30\text{-}C}$	$q_t^{30\text{-}30}$	$q_t^{30\text{-}60}$	0	0	0	$q_t^{30\text{-}PO}$
Late 60	$q_t^{60\text{-}C}$	$q_t^{60\text{-}30}$	$q_t^{60\text{-}60}$	$q_t^{60\text{-}90}$	0	0	$q_t^{60\text{-}PO}$
Late 90+	$q_t^{90\text{-}C}$	$q_t^{90\text{-}30}$	$q_t^{90\text{-}60}$	$q_t^{90\text{-}90}$	$q_t^{90\text{-}F}$	0	$q_t^{90\text{-}PO}$
Foreclosure	$q_t^{F\text{-}C}$	$q_t^{F\text{-}30}$	$q_t^{F\text{-}60}$	$q_t^{F\text{-}90}$	$q_t^{F\text{-}F}$	$q_t^{F\text{-}REO}$	0
REO	0	0	0	0	0	$q_t^{REO\text{-}REO}$	$q_t^{REO\text{-}PO}$
Paid Off	0	0	0	0	0	0	1

Table 20: Transition Probabilities in the Seven State Transition Model.

Note that some transitions, such as 30 to 90 Days Late are not possible. The probability that a mortgage would be in a particular state is the sum of all the probabilities of going to that state times the probability of being in that state. For example, the probability that the mortgage would be Current next period is:

$$p_{t+1}^{C} = p_t^{C} * q_t^{C\text{-}C} + p_t^{30} * q_t^{30\text{-}C} + p_t^{60} * q_t^{60\text{-}C} + p_t^{90} * q_t^{90\text{-}C} + p_t^{F} * q_t^{F\text{-}C}$$

The three state model loses most of the payment status detail of the seven state model by reducing the process to forecasting the probability of a loan defaulting ($q_t^{A\text{-}D}$) or paying off ($q_t^{A\text{-}PO}$).

The three-state model compresses several states into a single transition. The three-state model maps the monthly transitions in the following way:

FROM	TO						
	Current	**30**	**60**	**90**	**Foreclosure**	**REO**	**Paid**
Current						ACTIVE	ACTIVE
Late 30		ACTIVE				TO	TO
Late 60		TO				DEFAULT	PAID-
Late 90+		ACTIVE					OFF
Foreclosure							

Table 21: Mapping of Transitions from the Seven State Model to the Three State Model.

From the above table, we can see that loans that make any of the transitions back and forth between current and foreclosure and a particular month are simply treated as if they remained in the single category called Active. The problem with this mapping is that forecasting error is introduced if the *proportions* of loans in each payment status in the data to be forecast are different than those used to estimated the parameters. That is because the effect each loan characteristic on the probability of making a transition differs based on the starting monthly payment status. This interferes with the accurate estimation of the model parameters. The size of the forecasting error in the three state model is expected to increase as the difference in the proportions of loans in each payment status increases. If the three state model were only used to forecast loans from the month of origination, then the error may be negligible. However, for loans that are even a few months old the forecasting value of knowing if the loan is 60 or 90 days late, for example, may be quite substantial.

Statistical Estimation Methodology

This section describes the statistical estimation of the different models. Each model uses loan, economic, and borrower characteristics to calculate a unique probability of a loan transitioning from one payment status to another. Logit was used to estimate four models: the seven state model, a prepayment only model, a default only model, and a joint model that has three states: Active, Paid Off, or Defaulted. Historical performance information for close to one million loans were used to the probability that a loan will transition from one state to another. Since default can only occur on payment due dates, (i.e., is a discrete time process) logit was used to estimate the model parameters. The effects of these variables are estimated using Maximum Likelihood Multinomial Logit[50]. Applying a logit estimator to an event history is equivalent to a particularly simple type of discrete hazard model. The unit of observation is not the survival time of an individual mortgage, but rather the monthly payment status of an individual mortgage. Thus, for estimating the coefficients each month of loan data is treated as a separate data point.

---

[50] Logit is equivalent to a discrete time Cox proportional hazard model, and yet offers several advantages. First, it allows for explicit estimates for the coefficient on the age of the loan variable.

Separate models for ARM and fixed rate loans were estimated. The affect of each variable is independently estimated for each transition. The loan, borrower and economic variables used in the model are those commonly found in prepayment and default models[51]. These variables include the age of the loan, the estimated current loan to value ratio (LTV), the borrowers FICO score at the time of origination, the relative spread of current interest rates over the mortgage rate, and indicators for various loan characteristics such as if the borrower provided full documentation or not. Complete variable definitions and the functional forms used are in Appendix F.

Statistical Comparison of The Models

This section compares the regression statistics from the various models and tests the restrictions implied by the two and three state models. By comparing the separate prepayment and default models with the combined transition model, we find that separate prepayment and default models produce inferior forecasts. This is done by comparing the forecasts from the various models to the actual prepayments and defaults for 30 securitized pools of subprime loans. Since the two and three state models are nested within the seven state model, any increased forecasting ability from the seven state model is purely due to the additional information contained in the monthly payment status.

Second, it is computationally feasible, given the large number of tied survival times (Allison 1995). The estimation methodology is similar to that used by OFHEO (1999).

[51] Similar models are Quigley and Van Order (1995), by OFHEO (1999), and Lundstedt (1999).

A common measure of the forecasting ability of logit regressions is a c-statistic, which is the area under a receiver operating curve (ROC). C-statistic values range from .5 to 1, where .5 suggests that the model is no better than random, and 1 is a perfect fit. The next table shows the c-statistics from the regressions for the seven and three state models.

From\To	30	60	90	Foreclosure	REO	Paid Off
Current	0.647	0	0	0	0	0.714
Late 30	0.597	0.589	0	0.724	0	0.682
Late 60	0.601	0.673	0.598	0.641	0	0.66
Late 90+	0.567	0.651	0.644	0.624	0.809	0.656
Foreclosure	0.626	0.625	0.674	0.628	0.712	0.628

Table 22: C-statistics for ARM Subprime Loans in the Seven State Model.

Larger c-statistics are associated with the transitions are easier to predict, such as going from 90+ to Foreclosure. There are no c-statistics for going to current, since in multinomial logit one outcome is always derived from the other probabilities. This ensures that the probabilities for the payment statuses that a loan can transition to add up to 1. The above table can be compared with the following table of c-statistics for the three state model:

From\To	Default	Paid Off
Active	0.585	0.624

Table 23: C-statistics for ARM Subprime Loans in the Three State Model.

Since the C-statistics from the seven state model are in general larger that the ones for the three state model, this suggests that the loss of information from compressing down the number of states has a substantial effect on the model's fit. For example, the transitions in the seven state model related to default, such as 90 Days Or More Late to REO, have much larger c-statistics than the Active to Default transition in the three state model. These tables are only for adjustable-rate mortgages (ARM**).** The results for fixed rate loans were similar.

Statistical Tests of the Restrictions Implied by the Three State Model

The focus of this section is to test the restrictions placed on the data by the 2 and 3 state models. Since the typical goal of mortgage forecasting models is to predict prepayment and default, the restrictions pertaining to prepayment and default are the only restrictions considered. The three state model implies that the probability of paying off is the same, regardless of the current payment status of the loan. In other words, the three state model forces the coefficients to be the same for all of the transitions to Paid Off in the seven state model (current to Paid Off, 30 Days Late to

Paid Off, 60 Days Late to Paid Off, 90+ Days Late to Paid Off, and Foreclosure to Paid Off). The plausibility of this restriction can be assessed graphically in Figure 7.

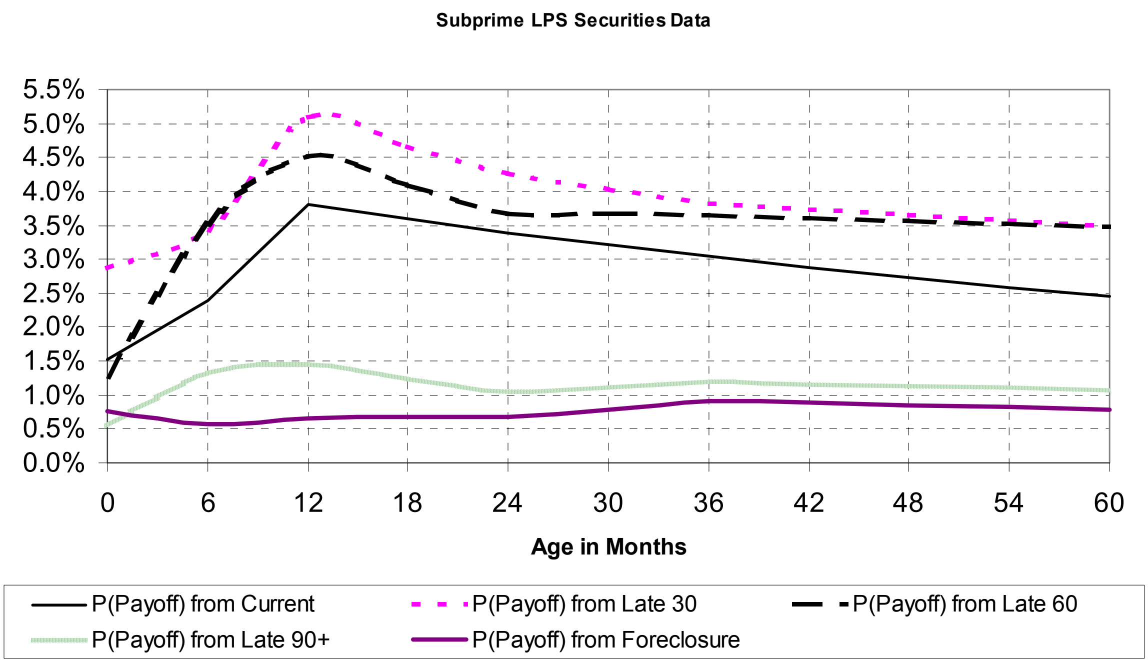

Figure 7: The Probability of a Subprime Loan Paying off Each Month, for Each Current Payment Status.

Figure 7 shows the probabilities of a loan paying off, P(Payoff), each month for each loan payment status from the unrestricted seven state model, holding other characteristics fixed[52]. The lines in the graph are all drawn for the same typical loan, where the only characteristic that was varied was the age of loan. From the graph, it is clear that the probabilities of payoff differ dramatically depending on the current payment status. However, the three state model forces all of these lines to the equal. Thus, the probability of payoff in the three state model is a weighted average of the

[52] Since the model depends on loan characteristics, the probability of prepayment as a function of age will be different for each loan.

lines shown in the graph. The weighting depends on the proportion of loans in each payment status in the original data set used to estimate the model, which has a high probability of being different from the proportion of loans in the different payment statuses in the data to be forecast.

More formally, the restriction that the probability of payoff is the same no matter what the current payment status is tested as follows. First, note that the seven state model is estimated with five different multinomial logit equations, one for each current loan status (i.e. one logit for each row of Table 24).

Logit Regression	**From Loan Status**	**To Loan Status**		
		Active	**Default**	**Paid Off**
1	**Current**	$q_t^{C\text{-}A}$	$q_t^{C\text{-}D}$	$q_t^{C\text{-}PO}$
2	**Late 30**	$q_t^{C\text{-}A}$	$q_t^{30\text{-}D}$	$q_t^{30\text{-}PO}$
3	**Late 60**	$q_t^{60\text{-}A}$	$q_t^{60\text{-}D}$	$q_t^{60\text{-}PO}$
4	**Late 90+**	$q_t^{90\text{-}A}$	$q_t^{90\text{-}D}$	$q_t^{90\text{-}PO}$
5	**Foreclosure**	$q_t^{F\text{-}A}$	$q_t^{F\text{-}D}$	$q_t^{F\text{-}PO}$

Table 24: Separate Regressions Run for Testing Restrictions Implied by The Three State Model.

Testing were conducted by imposing the restriction that the coefficients in the 5 logits are the same as the coefficients from the three state model. The three state model forces this restriction in the estimation process. The test of the restriction for prepayment consists of five separate tests. The results, as shown in the next table, strongly reject each restriction.

Logit Equation	**Transistion**	**DF**	**Wald Chi-Square**	**p Value**	**Count of Non-Paid Off Loans**	**Count Paid Off Loans**
1	Current to Paid Off	42	870.6712	<.0001	16118400	52739
2	30 Days Late to Paid Off	42	2124.2546	<.0001	160092	7063
3	60 Days Late to Paid Off	42	441.0718	<.0001	45406	1807
4	90+ Days Late to Paid Off	42	471.7252	<.0001	58381	1458
5	Foreclosure to Paid Off	42	611.2765	<.0001	63545	2148

Table 25: Restriction Test Results for ARM loans to Payoff.

The DF column shows the degrees of freedom for the tests. Since there are 42 regressors in the model, forcing them to all to be the same as in the three state model results in 42 degrees of freedom. The Wald Chi-Square was chosen for convenience. The p values indicate that if the restrictions are correct the probability of observing the values estimated in the seven state model are less than .0001 percent. Thus, we can reject the restriction at a high level of statistical confidence. Likewise, similar restrictions and results obtain for the Active to Active and Active to Default transitions.

Empirical Comparison of The Various Model's Forecasts

While the last section indicated that the three state model involves restrictions that are easily rejected, this section focuses on whether the differences in the forecasts of the differing models are economically meaningful. Thus, we compare the prepayment and default forecasts of the various models to the actual performance of a set of 30 actual Asset Backed Securities[53]. The goal is to test if differences between the models have any economic significance. The strategy is to see what the different models would have predicted for the 30 securities as far back in time as data is available, and then compare those predictions to actual outcomes. The securities tested represent the largest set of security data available from LoanPerformance that had both historical loan performance information and the requisite loan characteristics required for input into the models.

Figure 8 shows the forecasts of the cumulative percentage of loans that prepaid between the earliest and the last months (October of 2000) for which data was available. The forecasts of the various models are plotted with the actual cumulative prepayment rates for each of the 30 test securities. In each of the charts, the actual rates (the historically observed quantities) are denoted by round markers.

[53] The Asset Backed Securities used here are mortgage-backed securities based on subprime loans.

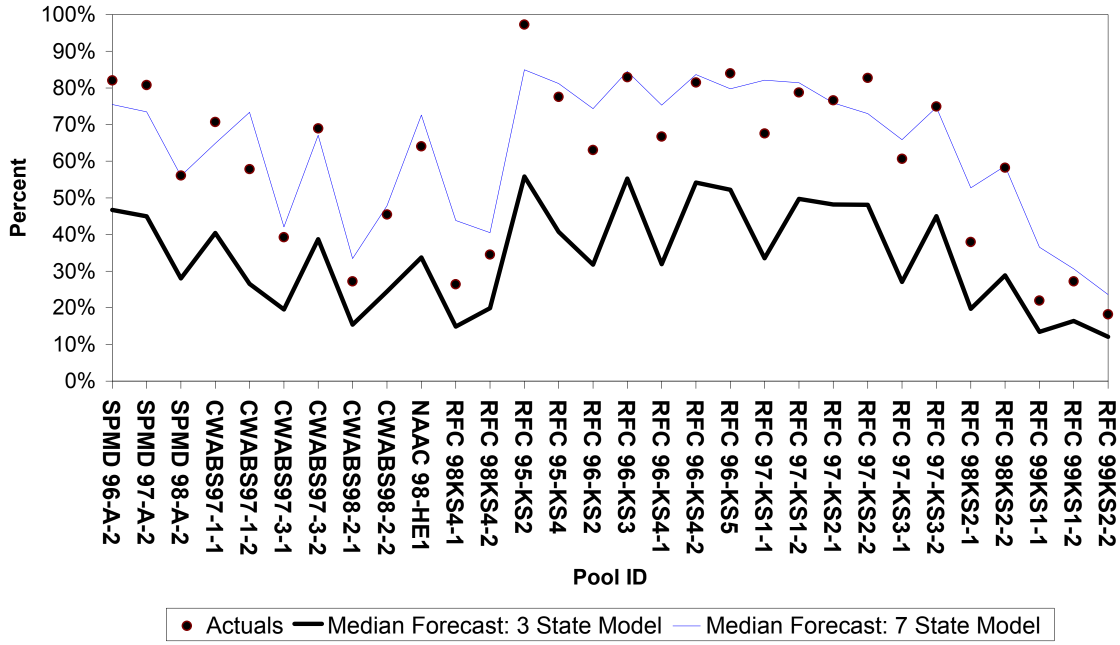

Figure 8: Model Forecasts vs. Actual Cumulative Percentage of Loans That Paid Off.

We can see clearly in Figure 8 that the seven state model produced more accurate forecasts than the three state model. The three state model under predicted prepayments on every ABS deal tested. The seven state model appears to be far more accurate, and did not appear to systematically over or under forecast prepayments. The poor performance of the three state model is undoubtedly related to the fact that fewer variables were included in the model. This is because fewer variables were statistically significant in the two and three state models.

The following graph compares the default forecasts from the two modeling methods. It shows forecasts of the cumulative percentage of loans that defaulted from the various models, plotted alongside the actual default rates.

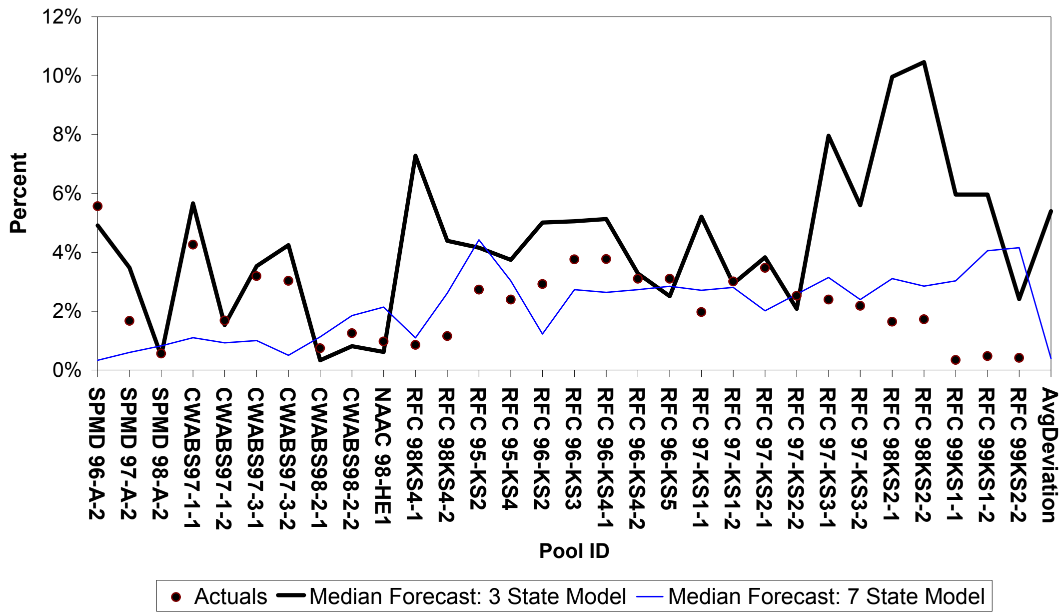

Figure 9: Cumulative REO for the Models vs. Actuals.

Figure 9 shows that the three state model produced far more volatile forecasts. The seven state model produced errors that were, on average, three times smaller than the errors produced by the three state model.

Conclusion

A seven state mortgage payment status transition model was presented and its forecasts were compared to those from similar two and three state models. In order to make the comparison fair, the models are estimated using the same functional form and data. The seven state model adds more realistic detail without changing the specification of how variables enter into the model. The results indicate that that the expanded model proposed here statistically and empirically dominates models built using existing methods. However, the seven state model requires a larger sample size to estimate, since it requires sufficient numbers of loans for each possible transition. The seven state model takes advantage of the extra information contained in a current payment status to produce superior forecasts compared to a three state model in a sample of 30 securities.

One contribution was detailing the theoretic advantages of expanding the number of loan states from the two or three traditionally used to the seven proposed here. One such observation is that for two and three state models, forecasting error is introduced if the proportion of loans in each monthly payment status in the data to be forecast are different than the proportions used to estimate the parameters.

One possible extension is to repeat the methodology in this study for loan types besides subprime, such as conforming and jumbo. It would also be interesting to investigate further if it is optimal to combine or break out various payment states in ways not studied here.

Bibliography

ABRAHAM AND HENDERSHOTT: (1996): "Bubbles in Metropolitan Housing Markets," *Journal of Housing Research*, V7, 2: 191-207

AMBROSE, B., BUTTIMER, R., AND C. CAPONE (1997): "Pricing Mortgage Default and Foreclosure Delay," *Journal of Money, Credit, and Banking*, Volume 29, No. 3 August 1997

AMBROSE, B., AND C. CAPONE (1998): "Modeling the Conditional Probability of Foreclosure in the Context of Single-family Mortgage Default Resolutions," *Real Estate Economics,* V26, 3: 391-429.

AMBROSE, B., AND C. CAPONE (1996): "Costs-Benefit Analysis of Single-Family Foreclosure Alternatives," *Journal of Real Estate Finance and Economics,* 13, 105 - 120.

AMBROSE, B., C. CAPONE, AND Y. DENG (2000): "The Optimal Put Exercise: An Empirical Examination of Conditions for Mortgage Foreclosure," Forthcoming in *Journal of Real Estate Finance and* Economics, 23 (2)

ALLISON, P. (1995): *Survival Analysis Using the SAS System A Practical Guide*, The SAS Institute Inc. Cary, N.C.

BERKOWITZ, J. (1998): "Evaluating the Forecast of Risk Models," Federal Reserve Board, unpublished work

BRENNAN, M. AND E. SCHWARTZ (1980): Analyzing Convertible Bonds, *Journal of Financial and Qualitative Analysis*, 15, 907-929

CAMPBELL, J., A. LO, AND A. MACKINLEY (1997): *The Econometrics of Financial Markets*, Princeton University press, Princeton NJ

CALHOUN, C. A. (1996): "OFHEO House Price Indexes: Technical Description," Washington, D. C., Office of Federal Housing Enforcement Oversight, April 1996.

CAPONE, C., *Providing Alternatives to Mortgage Foreclosure: A Report to Congress. Washington*, D.C.: U.S. Department of Housing and Urban Development, Aug. (HUD-1611-PDR).

CAPONE, C., AND Y. DENG (1998): "Loss Severity and Optimal Put Exercise: An Examination of Negative Equity in Mortgage Foreclosure," Washington, D. C., Office of Federal Housing Enforcement Oversight, January, 1998.

CAPOZZA D., D. KAZARIAN, AND T. THOMPSON (1997): "Mortgage Default in Local Markets," *Real Estate Economics*, 1997

CAULEY, S. AND A. PAVLOV (2001): "Non-Recourse Loans and the Value of Income Producing Real Estate," Unpublished Working Paper

CLAURETIE, T., AND T. HERZOG (1990): "The Effect of State Foreclosure Laws on Loan Losses: Evidence from the Mortgage Insurance Industry", *Journal of Money, Credit and Banking,* Volume 22, Issue 2 May 1990

COX, J., J. INGERSOLL, AND S. ROSS (1981): "A Re-examination of Traditional Hypothesis about the Term Structure of Interest Rates," *Journal of Finance*, 36, 769-799

COX, J., INGERSOLL, J. AND S. ROSS (1985): "Option Pricing: a Simplified Approach," *Journal of Financial Economics*, Sept.

CUNNINGHAM, D. AND C. CAPONE (1990): “The Relative Termination Experience of Adjustable to Fixed-rate Mortgages, " *The Journal of Finance*, December 1990

CUNNINGHAM, D. AND P. HENDERSHOTT (1984): "Pricing FHA Mortgage Default Insurance," *Housing Finance Review,* 3 (4) 373-392

DENG, Y., J. QUIGLEY AND R. VAN ORDER (1996): "Mortgage Defaults and Low down Payment Loans: the Costs of Public Subsidy," *Regional Science and Urban Economics*, 26 (3-4), to 6 3-285

DENG, Y., J. QUIGLEY AND R. VAN ORDER (2000): "Mortgage Terminations, Heterogeneity, and the Exercise of Mortgage Options," *Econometrica,* 2000

DIXIT, A., PINDYCK, R. (1993): *Investment under Uncertainty*, Princeton University Press, Princeton NJ

DOWNING, C. (1998): "Topics in the Home Mortgage Industry," Dissertation, U. C. Berkeley

EPPERSON, J., J. KAU, D. KENNAN, AND W. MULLER (1985): "Pricing Default Risk in Mortgages," *Journal of AREUEA*, 13 (3), 261-272

FEDERAL RESERVE SYSTEM TASK FORCE ON INTERNAL CREDIT RISK MODELS (1998): “Credit Risk Models at Major U.S. Banking Instituti*ons: Current State of the Ar*t and Implications for Assessments of Capital Adequacy,” May 1998

BOARD OF GOVERNORS OF THE FEDERAL RESERVE SYSTEM(1999): SR-18 Assessing Capital Adequacy in Relation to Risk at Large Banking Organizations and Others with Complex Risk Profiles

FOSTER, C. AND R. VAN ORDER (1984): "An Option Based Model of Mortgage Default," Housing Finance Review 3 (Oct. 1984), 351-372

FREDDIE MAC (2000): “EarlyIndicator: 4.0 User’s Guide” Freddie Mac web site

HEATH, D., R. A. JARROW, AND A. MORTON (1991): "Contingent Claim Evaluation with a Random Evolution of Interest Rates," *Review of Futures Markets*, 9 (1) 54-76. pp 77-82.

HEATH, D., R. A. JARROW, AND A. MORTON (1992): "Bond pricing and the Term Structure of Interest Rates: A New Methodology For Contingent Claims Evaluation," *Econometrica*, 60 (1) 77-105.

GARDENER, M. AND D. MILLS (1989): Evaluating The Likelihood of Default on Delinquent Loans," *Financial Management*

HALL, A. (2000): “Controlling for Burnout in Prepayment Model estimation,” Fair Isaac

HENDERSHOTT, P., AND R. VAN ORDER (1987): “Pricing Mortgages: An Interpretation of the Models and Results,” *Journal of Financial Services Research*, 1, 77-111

INSIDE MORTGAGE FINANCE PUBLICATIONS (2000): “Inside MBS & ABS,” Sept 15, 2000

JONES, L. (1993): Deficiency Judgments and the Exercise of the Default Option in Home Mortgage Loans," *Journal of Law and Economics*, Volume 36

KAU, J., D. KEENAN, W. MULLER, AND J. EPPERSON (1992): "A Generalized Evaluation Model for Fixed-Rate Residential Mortgages," *Journal of Money, Credit and Banking*, Volume 24, No. 3, August 1992

KAU, J., D. KEENAN, W. MULLER, AND J. EPPERSON (1995): "The Valuation at Origination of Fixed-Rate Mortgages with Default and Prepayment," *Journal of Real Estate Finance and Economics*, 11: 5-36

KAU, J., AND D. KEENAN (1995): “An Overview of the Option-Theoretic Pricing of Mortgages,” *Journal of Housing Research*, 6(2), 217-244

KAU, J., D. KEENAN, AND T. KIM (1993): “Transaction Costs, Suboptimal Termination and Default Probabilities,” *Journal of the American Real Estate and Urban Economics Association* v21,3: pp 247-263

KRAMER, B., AND I. S. MANSUKHAN (2000): "Analyzing Loan-level Default Performance in the Jumbo MBS Sector," Bear Stearns and Co. Residential Credit Research 2000

KUO, C. (1996): "Housing Price Dynamics and the Valuation of Mortgage Default Options," *Journal of Housing Economics*, 5, 18-40

KYRIACOU, K., AND D. WESTERBACK, (1999): "Clearing the Mists: Inside the Elusive Concept of Loan Loss Severity," Structured Finance Special Report, Moody's Investors Service, Aug. 27, 1999

LACOUR-LITTLE, M. (1997): “Mortgage Prepayments: Do Borrower Characteristics Matter?” Citicorp. Mortgage Inc. St. Louis, MO

LEKKAS, V., QUIGLEY, J., AND R. VAN ORDER (1993): "Loan Loss Severity and Optimal Mortgage Default" *Journal of the American Real Estate and Urban Economics Association* v21,4: pp 353-371

LO, A. AND J. WANG (1995): "Implementing Option Pricing Models When Asset Returns Are Predictable," Journal of Finance, 50, 87-129

LUNDSTEDT, K. (1999): "The influence of Non-Option-Related Variables Upon Corporate Default and Residential Mortgage Termination," Dissertation, U. C. Berkeley

MEESE AND WALLACE (1997): "Dwelling Price Dynamics in Paris, France," Unpublished paper

MULHERIN, J., AND W. MULLER (1987): "Volatile Interest Rates and the Divergence of Incentives in the Mortgage Contracts," *Journal of Law, Economics, and Organization* Spring 1987

NEFTCI, S. (1996): *An Introduction to the Mathematics of Financial Derivatives*, Academic Press, San Diego, CA

OFHEO (1999): Risk-Based Capital Regulation: *Second Notice of Proposed Rule Making II* Office of Federal Housing Enforcement Oversight page 321-360

QUIGLEY, J., AND R. VAN ORDER (1991): "Defaults on mortgage obligations in capital requirements for U.S. savings institutions," *Journal of Public Economics*, 44, pp. 353-369

QUIGLEY, J., AND R. VAN ORDER (1995): "Explicit Tests of Contingent Claims Models of Mortgage Default," *Journal of Real Estate Finance and Economics*, Volume 11

SIRMANS, C. (1985) *Real Estate Finance*, McGraw-Hill

SMITH, D. AND E. LAWRENCE (1995): "Forecasting Losses on a Liquidating Long-term Loan Portfolio," *Journal of Banking and Finance,* 19 p. 959-985

SMYTH, L, SANCHEZ, S., AND E. LAWRENCE (1996): "Comprehensive Model for Managing Credit Risk on Home Mortgage Portfolios," Decision Sciences, Volume 27 No. 2 p. 291-317

SPRINGER, T., AND N. WALLER (1993): "Termination of Distressed Residential Mortgages: an Empirical Analysis," *Journal of Real Estate Finance and Economics*, Volume 7:1 (1) 43-53 July 1993

SPRINGER, T., AND N. WALLER (1993): "Lender Forbearance: Evidence From Mortgage Delinquency Patterns," *Journal Of The American Real Estate And Urban Economics Association* V21, 1:pp.27-46

STANTON, R. (1995): "Rational Prepayment and Evaluation of Mortgage-

Backed Securities," *The Review of Financial Studies*, Volume 8, No. 3

STANTON, R. AND N. WALLACE (1998) "Mortgage Choice: What is the Point?," *Real Estate Economics*, Volume 26, No. 2, 1998

TITMAN, S. AND W. TOROUS (1989): "Valuing Commercial Mortgages: an Empirical Investigation of the Contingent-Claims Approach to Pricing Risky Debt," *Journal of Finance*, 44 (2): 345-73

TRIGEORGIS, L. (1996): *Real Options*, MIT Press, Cambridge, MA

VANDELL, K. (1995): "How Ruthless Is Mortgage Default? A Review and Synthesis of the Evidence," *Journal of Housing Research*, Volume 6, Issue 2

VANDELL, K. AND T. THIBODEAU (1985): "Estimation of Mortgage Defaults Using Disaggregate Loan History Data," Journal of the American Real Estate and Urban Economics Association, 13 (3): 292-316

WANG, W. (2001): "Chapter 6 Going beyond Credit Score: the Development of a Loss Model for the Home Equity Business," *Handbook of Credit Scoring*, Elizabeth Mays Editor, Glenn Lake Publishing Co., Ltd., Chicago

WILMOTT, P. (1998): *Derivatives*, John Wiley& Sons Ltd., West Sussex, England

WILSON, D. (1995): “Residential Loss Severity in California: 1992 to 1995,” *Journal of Fixed Income*, December, 1995

YANG, T, H. BUIST AND I. MEGBOLUGBE (1999): An Analysis of the Ex Ante Probabilities of Mortgage Prepayment and Default, *Real Estate Economics* 26, 4 (1999): 651–676.

Appendix A: Subprime Pools Tested in Chapter 3

This Appendix describes the Subprime pools used to test the models. The distribution of various loan characteristics are shown below.

Pool	Ave Age at Closing	# Loans	Total Value	%FRM	Avg FICO	Ave Balance	WAC	Avg CLTV	Mean Year
CWABS 97-01-1	1.5	1880	$183,309,851	0.0	610	$97,505	9.4	70.6	1996
CWABS 97-01-2	1.6	2655	$106,402,039	100	675	$40,076	10.6	81.7	1996
CWABS 97-03-1	2.3	1300	$87,189,776	100	605	$67,069	10.4	66.3	1996
CWABS 97-03-2	1.2	1840	$189,099,397	0.0	600	$102,771	9.4	71.5	1996
CWABS 98-02-1	1.7	1212	$95,066,913	99.0	601	$78,438	9.8	69.6	1997
CWABS 98-02-2	2.2	1606	$180,657,166	0.0	600	$112,489	9.4	75.1	1997
NAAC 1998-HE1	4.6	1977	$220,268,331	19.1	606	$111,415	9.6	75.7	1997
RFC 1995-KS2	5.6	939	$105,746,713	0.0	.	$112,616	10.4	69.7	1994
RFC 1995-KS4	6.4	990	$104,153,010	57.6	.	$105,205	10.3	73.8	1994
RFC 1996-KS2	6.1	973	$100,912,126	75.5	623	$103,712	10.2	71.1	1995
RFC 1996-KS3	4.1	1107	$143,965,783	0.0	617	$130,050	9.7	74.9	1995
RFC 1996-KS4-1	4.6	966	$95,154,009	85.1	620	$98,503	10.8	72.6	1995
RFC 1996-KS4-2	4.3	1699	$223,461,241	0.0	622	$131,525	9.9	76.6	1995
RFC 1996-KS5	3.2	1524	$200,241,890	0.0	621	$131,392	9.8	78.3	1995
RFC 1997-KS1-1	5.7	1432	$122,234,448	67.9	618	$85,359	11.3	68.2	1996

Pool	Ave Age at Closing	# Loans	Total Value	%FRM	Avg FICO	Ave Balance	WAC	Avg CLTV	Mean Year
RFC 1997-KS1-2	3.5	1204	$152,067,963	0.0	619	$126,302	9.6	77.3	1996
RFC 1997-KS2-1	3.5	1962	$250,438,021	0.0	610	$127,644	9.8	77.6	1996
RFC 1997-KS2-2	4.5	916	$115,372,803	0.0	612	$125,953	9.9	78.1	1996
RFC 1997-KS3-1	4.8	2464	$200,641,225	69.5	614	$81,429	11.3	66.8	1996
RFC 1997-KS3-2	3.4	2034	$250,433,817	0.0	613	$123,124	10.0	78.8	1996
RFC 1998-KS2-1	4.1	5272	$403,056,978	76.5	608	$76,452	10.5	70.2	1997
RFC 1998-KS2-2	4.0	3849	$445,949,950	0.0	601	$115,861	9.9	80.1	1997
RFC 1998-KS4-1	4.2	4957	$351,181,901	77.5	605	$70,846	10.4	72.0	1997
RFC 1998-KS4-2	4.0	4034	$475,885,816	4.5	596	$117,969	10.0	79.2	1997
RFC 1999-KS1-1	4.2	8253	$651,891,067	82.2	603	$78,988	10.5	70.4	1998
RFC 1999-KS1-2	4.4	6365	$651,293,011	0.0	590	$102,324	10.2	78.8	1998
RFC 1999-KS2-2	3.7	5438	$575,870,164	0.0	587	$105,897	10.2	79.0	1998
SPMD 96-A-2	3.4	613	$82,647,987	0.3	.	$134,825	10.1	74.0	1995
SPMD 97-A-2	10.8	665	$87,758,062	0.0	647	$131,967	9.9	71.9	1996
SPMD 98-A-2	5.4	3951	$486,945,633	0.0	601	$123,246	9.1	74.3	1997
Totals	4.0	74077	$7,339,297,091	34.0	606	$99,077	10.1	74.6	1997

Table 26: Description of Pools Tested in Chapter 3.

Appendix B: Regressors Used in Estimating All Statistical Models in Chapter 1

The next table shows the variables used in the models. Note that many variables use linear splines. Dummies for missing values are used to avoid biased coefficients. These dummies may pick up servicer or market specific effects. Since severity is the object of interest, a positive coefficient indicates larger losses. If the expected sign is unclear, then the Expected Sign field is left blank.

Variable Name	**Definition**	**E(Sign)**
	Option Theory Variables	
Age	Age of the loan in months	
Age_sq	Age squared	
Relsprd	Relative spread. For Fixed: (Spread at time of default)/Freddie 30yr rate . For ARM: (Spread at default)/ 1 Year ARM rate [54]	Positive
Relsprd_sq	Relative spread squared	
Prob_Neg_eq	Probability of negative equity	Positive
Prob_Neg_eq_sq	Probability of negative equity squared	
change_HPI	Change in house price index	
change_HPI_sq	Change in house price index squared	
	Cost Related Variables	

[54] Interest rate data comes from the Federal Reserve web site.

ServicerX	Indicators for servicers	Mixed
Right_of_Redemption	Number of months that a State allows the borrower to take back a property[55]	Positive
Non_Recourse	State doesn't allow of deficiency judgments[56]	Positive
NonJudicial_State	State uses non-judicial proceedings: AL, AK, CA, DC, GA, ID, MD, MA, MS, MT, NC, NH, NV, OR, RI, TN, TX, VA, WV, WY	Negative
Judicial_State[57]	State uses judicial proceedings: CT, DE, FL, IL, IN, KS, KY, LA, ME, ND, NJ, NM, NY, OH, OK, PA, PR, SC, VI, VT	Positive
Bankruptcy	Indicates borrower has filed for bankruptcy	Positive
Short Sale	Indicates loan sold before reaching REO	Negative
	Cash Flow and Loan Characteristic Variables	
Orig_CLTV	Combined LTV at origination	0
OrigcltvGT80	Combined LTV at origination if > 80%	
ARM	Indicates Adjustable Rate Mortgage loans	Positive
Subprime	Indicates Subprime loans	Positive

[55] The data comes from Freddie Mac's Servicing Manual (2000). No attempt was made to account for legal changes during the time of this study.

[56] States used for this were CA, MN, MT, NC, and WA, as reported by Lin and White (2000).

[57] Several states allow for both judicial and non-judicial procedures, so these states are the used as the base case.

Alt-A	Indicates loan was not quite prime	Positive
Balloon	Indicates Balloon loans	Positive
Investor	Investor owned property	Negative
Original_Margin	Spread at origination. For Fixed: Spread at origination over Freddie 30yr. For ARM: Spread at origination over 1 Year ARM	Positive
FICO_Between300_550	Credit scores at origination was between 300 and 550	Positive
FICO_Between550_620	Credit scores at origination was between 550 and 620	
FICO_Greaterthan_620	Credit scores at origination was greater than 620	Negative
Refi	Indicates borrower refinanced loan, and an LTV <=80%	Positive
Debt_Ratio_35_40	Indicates debt to income ratio at origination was between 35% and 40%	Positive
Debt_Ratio_40_45	Indicates debt to income ratio at origination was between 40% and 45%	Positive
Debt_Ratio_GT45	Indicates debt to income ratio at origination was greater than 45%	Positive
PMI	Indicator that loan has private mortgage insurance	Negative
Missing_PMI	Indicates that mortgage insurance status was unknown	Negative

Orig_amtX_Y	Original loan amount is between X and Y X ={0, 50, 75, 100, 200, 300}, Y= {50, 75, 100, 200, 300, 600} (in thousands)	

Table 27: Variables Used in the Statistical Severity Model.

Appendix C: Short Sales Rates

In addition to loan characteristics, the data includes the monthly loan payment status: Current, Late 30 Days, Late 60 Days, Late 90 Days or More, Foreclosure, REO, or Paid Off. By examining the final monthly loan payment status, we can find percentage of loans that were in a particular payment status that ended in a short sale. Short sale rates are estimated by looking at all loans for all pools that report loss data. The following table shows for each final payment status, the number of all loans that ended from that payment status, the number of loans with losses, and the short sale rate.

Final Payment Status	**All Loans That Paid Off or Reported Losses**	**Loans with Losses**	**Short Sale Rate**
30 Days Late	62329	295	0.8%
60 Days Late	17426	314	3%
90 Days Late	13937	3749	27%
Foreclosure	16240	5091	31%

Table 28: Last Payment Status of Paid Off Loans, and the Short Sale Rates.

Appendix D: Derivation of the PDE in Chapter 2

This appendix shows the derivation of the PDE used to calculate the option values. First, assume that you can create a portfolio, *P*, consisting of the Mortgage and

$-\frac{\partial M}{\partial H}$ units of the house the mortgage is written on[58] (or some traded asset that is perfectly correlated with it). Then the change in a portfolio is given by the following (time subscripts are suppressed for simplicity):

$$dP = dM - \frac{\partial M}{\partial H} dH \qquad (3.1)$$

Using Ito's Lemma to get an equation for the instantaneous return on a derivative asset, in this case a mortgage, M(H,r,t) that depends on the underlying housing and interest processes in equations (2.4) -(2.6) yields the following equation:

[58] See Neftci (1996) p. 240, for an excellent treatment on this topic. This assumption is rather unrealistic, but it underlies all theoretic option pricing work done on mortgages that the author is aware of.

$$dM = \frac{1}{2}H^2\sigma_H^2\frac{\partial^2 M}{\partial H^2}dt + (\eta(\overline{H} \; - \; H)Hdt + \mathrm{H}\sigma_H \mathrm{dz_H})\frac{\partial M}{\partial H} + \frac{1}{2}r\sigma_r^2\frac{\partial^2 M}{\partial r^2}dt +$$
$$(\gamma(\theta - r)\mathrm{dt} + \sqrt{r}\sigma_r dz_r)\frac{\partial M}{\partial r} + \rho H\sqrt{r}\sigma_H\sigma_r\frac{\partial^2 M}{\partial r\partial H}dt + \frac{\partial M}{\partial t}dt$$

(3.2)

Plugging this into equation (3.1) gives:

$$dP = \frac{1}{2}H^2\sigma_H^2\frac{\partial^2 M}{\partial H^2}dt + (\eta(\overline{H} \; - \; H)Hdt + \mathrm{H}\sigma_H \mathrm{dz_H})\frac{\partial M}{\partial H} + \frac{1}{2}r\sigma_r^2\frac{\partial^2 M}{\partial r^2}dt +$$
$$(\gamma(\theta - r)\mathrm{dt} + \sqrt{r}\sigma_r dz_r)\frac{\partial M}{\partial r} + \rho H\sqrt{r}\sigma_H\sigma_r\frac{\partial^2 M}{\partial r\partial H}dt + \frac{\partial M}{\partial t}dt$$
$$-\frac{\partial M}{\partial H}(\eta(\overline{H} \; - \; H)Hdt + \mathrm{H}\sigma_H \mathrm{dz_H})$$

(3.3)

The weight $-\frac{\partial M}{\partial H}$ on H in the portfolio was chosen so that the terms involving dz_H cancel out. By the pure expectations hypothesis $E[dz_r] = 0$, so the return on the portfolio is *completely predictable*:

$$dP = \frac{1}{2}H^2\sigma_H^2\frac{\partial^2 M}{\partial H^2}dt + \frac{1}{2}r\sigma_r^2\frac{\partial^2 M}{\partial r^2}dt + \gamma(\theta - r)\frac{\partial M}{\partial r}dt + \rho H\sqrt{r}\sigma_H\sigma_r\frac{\partial^2 M}{\partial r\partial H}dt + \frac{\partial M}{\partial t}dt$$

By no arbitrage, the return on this portfolio must equal the instantaneous risk-free rate of interest adjusted for the fact that the mortgage does not receive the service flow:

$$dP = rPdt - sdt = rdM_t dt \text{ - } r\frac{\partial M}{\partial H}dH_t dt - sdt$$

Equating the two equations above, substituting in $s = \mu - \eta(\overline{H} - H)$, and "canceling" the *dt*s gives the final PDE for the mean reverting model[59]:

$$\frac{1}{2}H^2\sigma_H^2\frac{\partial^2 M}{\partial H^2} + (r - \mu + \eta(\overline{H} - H))H\frac{\partial M}{\partial H}$$
$$+\frac{1}{2}r\sigma_r^2\frac{\partial^2 M}{\partial r^2} + \gamma\ (\theta - r)\frac{\partial M}{\partial r} + \rho H\sqrt{r}\sigma_H\sigma_r\frac{\partial^2 M}{\partial r\partial H} + \frac{\partial M}{\partial t} = rM$$

(3.4)

The mean capital gain on the house value drops out of the equation, but the mean reverting service flow does not. One way to understand the result is to realize that even if we assumed that the value of the underlying asset can be hedged, the service flow cannot be hedged. The PDE for the Brownian motion process is derived in a similar fashion.

[59] This is a generalization of the mean reversion process studied in Dixit and Pindyck (1994) page 162.

Appendix E: Regressors used in Estimating Statistical Models in Chapter 3

The next table shows the variables used in the models. Dummy variables end with the letter D. Note that many variables use linear splines. Dummies for missing values are used to avoid biased coefficients. These dummies may pick up servicer or market specific effects.

Variable Name	**Definition**	**Expected Sign to Payoff**	**Expected Sign to Default**
Age	Age in months. Coefficient is 'slope' for first 5 months.	+	+
Age6	Coefficient is change in the 'slope' between 6 and 12 months.	-	
Age12	Coefficient is change in the 'slope' between 12 and 24 months.	-	
Age24	Coefficient is change in the 'slope' between 24 and 36 months.	-	
Age36	Coefficient is change in the 'slope' for 36+ months.	-	
arm2frm	(10 year Treasury Note-1 year Treasury Note)/100 (ARM only)	-	+

BalloonD	Dummy for Balloon loans.		
currcltv	Current Combined LTV at Origination < .5		
currcltv50	Change in slope for .75 > Current Combined LTV > 0.50		
currcltv75	Change in slope for .80.> Current Combined LTV > 0.75		
currcltv80	Change in slope for .90 > Current Combined LTV > 0.80		
currcltv90	Change in slope for Current Combined LTV > 0.90		
Fico_act	Slope on fico <550	+	-
FICO_CLTV	FICO*Current CLTV interaction term		
Fico_actQ1	Coefficient is change in the 'slope' between 550 and 620	+	
Fico_actQ2	Coefficient is change in the 'slope' between 620 and 710	+	
Fico_actQ3	Coefficient is change in the 'slope' above 710	+	
FifteenD	Dummy = 1 if term < 15 Years	-	-
first_reset24D	15 < first_rate <= 27 (ARM only)		
first_reset36D	27 < first_rate (ARM only)		
FulldocD	Full Documentation Dummy	+	-
lnsz	Slope on (Size of original loan)/(average loan		

	for MSA that year) if <.65		
loansizeQ1	.65 < relative loan size < .92		
loansizeQ2	.92 < relative loan size < 1.28		
loansizeQ3	1.28 < relative loan size		
InvestorD	Investor owned property	-	+
m_docD	Dummy for Missing Observations for 'document'	0	0
M_fico	Dummy for Missing Observations for FICOs	+	-
m_pp_penaltyD	Prepayment Penalty Dummy for Missing Observations	?	?
m_prop_typeD	Dummy for Missing Observations for 'prop_type'	0	0
m_purposeD	Dummy for Missing Observations for 'Purpose'	0	0
m_occupancy	Dummy for Missing Observations for 'occupancy'	0	0
months_to_reset	Months until reset (ARM only)		
prep_penaltyD	Prepayment Penalty Dummy	+	0
ReficashD	Refinance Cash Out Dummy	-	+
Not_SfrD	Not a Single Family Residence Dummy	0	0

Relsprd2	Relative spread. For Fixed: Spread at origination over Freddie 30 year. For ARM: Spread at origination over 1 Year ARM[60]		
Relsprd2C1	Coefficient is change in the 'slope' between .18 and .28		
Relsprd2C2	Coefficient is change in the 'slope' between .28 and .48		
Relsprd2C3	Coefficient is change in the 'slope' between .48 and .54		
Relsprd2C4	Coefficient is change in the 'slope' between .54 and .68		
Relsprd2C5	Coefficient is change in the 'slope' > .68		

Table 29: Regressors used in Estimating All Statistical Models.

[60] Interest rate data comes from the Federal Reserve web site.