

ABSTRACT

Title of Dissertation: ESSAYS ON INFORMATION ASYMMETRY IN THE
 U.S. RESIDENTIAL MORTGAGE MARKET:
 INCENTIVES AND ESTIMATIONS

Menglin Cao, Doctor of Philosophy, 2005

Dissertation directed by: Professor Roger Betancourt
 Department of Economics

This dissertation focuses on a phenomenon called appraisal bias in the residential mortgage market that stemmed from information asymmetry. It is composed of two essays, one theoretical and one empirical. The theoretical essay analyzes the existence of appraisal bias in a dynamic game of incomplete information framework and solves for the perfect Bayesian equilibria. It establishes how adding semi-verifiability condition to a cheap-talk game helps construct non-babbling equilibria in an asymmetric information environment. The empirical essay quantifies appraisal bias at individual loan level and measures its effect on mortgage terminations. It tests the extent to which option theory explains default and prepayment behavior in residential mortgage market. It treats default and prepayment hazards as dependent competing risks and jointly estimates mortgage terminations in a competing risk proportional hazard model framework and controls for unobserved heterogeneity using Heckman-Singer nonparametric method. It replaces the inaccurate approximated likelihood function that has been applied so far on competing

risks analysis with an exact likelihood function. Armed with repeat sale transaction data, this paper is the first to analyze the effect of appraisal bias on mortgage terminations. It concludes that appraisal bias is important in determining mortgage terminations and needs to be controlled for to correctly estimate termination hazards.

ESSAYS ON INFORMATION ASYMMETRY IN THE U.S. RESIDENTIAL
MORTGAGE MARKET: INCENTIVES AND ESTIMATIONS

By

Menglin Cao

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2005

Advisory Committee:

Professor Roger Betancourt, Chair
Professor Peter Cramton
Professor Ingmar Prucha
Professor Brian Ratchford
Professor John Rust

© Copyright by

Menglin Cao

2005

Dedication

This dissertation is dedicated to my parents, Datong Cao and Jinghui Hong, for always believing in me, inspiring me, and encouraging me to reach higher to achieve my goals. To my husband, whose support, love, and patience make everything I do possible.

Acknowledgements

I wish to express sincere appreciation to Dr. Eric Rosenblatt for his assistance in making this manuscript possible and to Dr. Mark Y. An, who provides me with constant guidance on econometric techniques. Special thanks are due to my committee chair, Professor Roger Betancourt, who encourages me with great patience and whose critical eye and enlightened mentoring are instrumental and inspiring. I am also grateful to the members of the committee council for their valuable input.

Table of Contents

Dedication	ii
Acknowledgements	iii
Table of Contents	iv
List of Tables.....	vi
List of Figures.....	vii
Chapter 1: Introduction to Dissertation	1
Chapter 2: How Mortgage Appraisers Produce Biased Appraisals by Inflating Property Value – A Theoretical Illustration.....	7
2.1 Introduction	7
2.2 Residential Mortgage Market.....	8
Figure 1. Market Participants and Their Relationship in the Primary and Secondary Mortgage Market	10
2.3 Literature Review	22
2.4 Model	28
2.4.1 Market Structure Revisit.....	29
2.4.2 Setup	29
2.4.3 Game Structure	32
2.4.4 Hybrid Equilibrium.....	41
2.4.5 Pure Strategy Equilibria.....	52
2.4.5.1 Pooling Equilibrium on \hat{H}	53
2.4.5.2 Separating Equilibrium Where H Reports \hat{H} and L Reports \hat{L}	59
2.5 Conclusion	61
Chapter 3: An Empirical Analysis of Mortgage Termination With Appraisal Bias – Maximum Likelihood Estimation of A Proportional Hazard Competing Risks Model With Grouped Duration Data	63
3.1 Introduction	63
3.2 Literature Review on Mortgage Termination	70
3.2.1 Early Literature Review.....	70
3.2.2 Economic Theory and Option Theory Review	73
3.2.3 Reviews of Econometric Methodology.....	86
3.3 Data.....	103
3.3.1 Dataset.....	103
3.3.2 Appraisal Bias	105
3.3.3 Explanatory Variables	118
3.3.4 Summary Statistics of Explanatory Variables.....	121
Table I. Descriptive Statistics of Time Invariant Variables by Loan Termination Type	122

Table II. Descriptive Statistics for Time Variant Variables By Loan Duration	123
Table III. Descriptive Statistics for Time Variant Variables By Calendar Time.....	124
3.4 Model	124
3.5 Estimation Results	136
3.5.1 Non-Ruthless Specifications.....	136
Table IV. Maximum Likelihood Estimation of Competing Risks Proportional Hazard Model – Model 1: Non-Ruthless Model With Controls for Appraisal Bias	138
3.5.2 Unobserved Heterogeneity	149
Table V. Maximum Likelihood Estimation of Competing Risks Proportional Hazard Model – Model 2: Non-Ruthless Model With Unobserved Heterogeneity and Controls for Appraisal Bias	153
3.5.3 Predicted Hazards	159
Figure 2. Actual vs. Predicted Hazards.....	160
Panel A. Prepayment.....	160
Panel B. Default	160
Figure 3. Prepayment Hazard and Market Interest Rate.....	161
3.6 Conclusion	162
Appendix A: Specification of Call and Put Option Variables	167
Appendix B: The Weighted Repeated Sales House Price Index.....	169
Appendix C: Other Specification Estimations.....	172
Bibliography	178

List of Tables

Table I. Descriptive Statistics of Time Invariant Variables by Loan Termination Type	122
Table II. Descriptive Statistics for Time Variant Variables By Loan Duration	123
Table III. Descriptive Statistics for Time Variant Variables By Calendar Time.....	124
Table IV. Maximum Likelihood Estimation of Competing Risks Proportional Hazard Model – Model 1: Non-Ruthless Model With Controls for Appraisal Bias	138
Table V. Maximum Likelihood Estimation of Competing Risks Proportional Hazard Model – Model 2: Non-Ruthless Model With Unobserved Heterogeneity and Controls for Appraisal Bias	153

List of Figures

Figure 1. Market Participants and Their Relationship in the Primary and Secondary Mortgage Market	10
Figure 2. Actual vs. Predicted Hazards.....	160
Panel A. Prepayment.....	160
Panel B. Default	160
Figure 3. Prepayment Hazard and Market Interest Rate.....	161

Chapter 1: Introduction to Dissertation

My dissertation focuses on a phenomenon called appraisal bias in the residential mortgage market that stemmed from information asymmetry. It refers to practices conducted by real estate appraisers and mortgage lenders who intentionally inflate collateral values in order to maximize their profit at the expense of mortgage investors.

My dissertation is composed of two essays, one theoretical and one empirical. The theoretical essay analyzes the existence of appraisal bias in a dynamic game of incomplete information framework and solves for the perfect Bayesian equilibria of this game. It establishes how adding semi-verifiability to a cheap-talk framework helps construct non-babbling equilibria in an asymmetric information environment.

The empirical essay quantifies appraisal bias at individual loan level and measures its effect on mortgage terminations. It jointly estimates mortgage terminations in a competing risk proportional hazard model framework and controls for unobserved heterogeneity in a nonparametric approach. It concludes that appraisal bias is important in determining mortgage terminations and needs to be controlled for to correctly estimate termination hazards.

Essay I: How Mortgage Appraisers Produce Biased Appraisals by Inflating Property Value – A Theoretical Illustration

The U.S. residential mortgage market consists of two markets: a primary market and a secondary market. The primary market is one where new mortgages are created through credit extension from mortgage lenders to borrowers. The secondary market is one where investors and mortgage lenders buy and sell existing and prospective loans. In order to reduce mortgage default risks faced to investors, most mortgages sold to the secondary mortgage market are guaranteed against default risks by government-sponsored enterprises (GSEs) such as Fannie Mae and Freddie Mac.

The structure of the secondary market limits the roles of GSEs by prohibiting them to participate in the primary market. They need to rely on information provided by mortgage lenders to evaluate the credit risks of mortgages and charge guarantee fees accordingly.

The structure of the U.S. secondary mortgage market creates an asymmetric information environment between GSEs and mortgage lenders. Information asymmetry provides incentive for one party to maximize his own utility at the expense of the other. In the secondary mortgage market, this incentive is realized through the practice of appraisal bias where lenders report inflated collateral values to GSEs in order to reduce the guarantee fees.

This paper sets up a dynamic signaling game with incomplete information structure to analyze the existence of appraisal bias. The game is played between mortgage lender and GSEs. In this game, mortgage lenders send out signals in the form of appraisal prices to GSEs about the credit quality of their loans and GSEs react to these signals in

conjunction with the termination outcomes of mortgages. In contrast to Spence (1976)'s signaling game, where the sender's signal is exogenously costly, the signals sent by mortgage lenders are costless and do not depend on lenders' types. In addition, they are nonbinding and non-verifiable. Hence, these messages are not informative to separate the types. These types of games are referred to as "Cheap Talk" games.

This paper modifies the "Cheap Talk" structure by adding an extra condition, namely semi-verifiability to make the signals informative. It establishes how adding semi-verifiability to a "Cheap Talk" game structure helps construct a non-babbling Bayesian equilibrium in an asymmetric information environment. It proves that appraisal bias can exist in the secondary mortgage market and provides equilibrium conditions for the hybrid and pure strategy equilibria that bring it into play.

It proves that there exists a hybrid equilibrium where it is optimal for mortgage lenders to randomize between truthfully reporting the property value and defrauding given GSEs' action. The existence of the hybrid equilibrium confirms empirical evidence that appraisal bias exists in some but not all mortgage cases. It also provides a scope of parameter conditions such that a pooling equilibrium where all lenders choose to behave honestly/dishonestly exists, as well as conditions for a separating equilibrium to exist where lenders behave differently according to the types of mortgages they receive.

This paper contributes to the literature by providing semi-verifiability conditions to reverse the uninformative signals in the "Cheap Talk" game specification. It is the first to

theoretically prove the existence of appraisal bias in the residential mortgage market and solves for the equilibrium strategies between secondary market participants.

Essay II: An Empirical Analysis of Mortgage Termination With Appraisal Bias -
Maximum likelihood Estimation of a Proportional Hazard Competing Risks Model with
Grouped Duration Data

Following the theoretical establishment on the existence of appraisal bias, this paper empirically tests the effect of appraisal bias on mortgage terminations in a competing-risk proportional hazard duration framework with controls for unobserved heterogeneity.

Armed with repeat sales transaction data, this paper is the first to measure appraisal bias at individual loan level and quantifies its effect in mortgage termination studies.

Following literature on mortgage terminations, this paper tests the extent to which financial option theory explains default and prepayment behaviors. By running different model specifications, this paper concludes that a non-ruthless model specification where mortgage terminations depend on borrower, property, and macroeconomic characteristics in addition to financial option related variables outperforms the ruthless model where only financial option related variables are deemed relevant. It also demonstrates that a non-ruthless model where appraisal bias is controlled for outperforms one that does not. Throughout model specifications, default and prepayment hazards are treated as dependent competing risks and are estimated jointly using maximum likelihood estimation method. In addition, this paper accounts for unobserved heterogeneity by

Heckman-Singer's nonparametric method and estimates it simultaneously with the parameters and baseline hazards for default and prepayment.

The findings suggest that high probability of negative equity increases the default hazard and reduces the prepayment hazard; large gap between the market value and the face value of the mortgage increases the prepayment hazard. In addition, default hazard is significantly influenced by appraisal bias. A high appraisal bias leads to a high likelihood of default and the effect increases in a decreasing fashion. Borrower's willingness to exercise financial options is also triggered or hindered by trigger events and asymmetric information. For example, default hazard monotonically increases as borrower credit score deteriorates; it increases with borrowers' debt-to-income ratio, origination loan-to-value ratio and state level quarterly unemployment rate; and decreases with mortgage loan amount. Mortgage loan amount is also found to be positively influencing prepayment default.

The model also suggests that there exist unobserved heterogeneity in mortgage terminations. The heterogeneity can be attributed to difference in borrowers' attitude toward prepayment and default, it can also be attributed to unmeasured house-specific factors such as unexpected depreciation or appreciation of property values, as well as borrower tastes or abilities. Among all plausible interpretations, I assume that unobserved heterogeneity is picking up differences in borrowers' financial awareness toward prepayment and default throughout this paper. Under the assumption that the heterogeneity is picking up the difference in borrowers' attitude, the results suggest that

80 percent of people have a high tendency toward prepayment risk, whereas 20 percent have low tendency toward prepayment; there does not seem to be any difference toward default.

This paper concludes that origination appraisal bias is significant in predicting default probabilities and must be controlled in order to achieve optimal model performance. The higher the appraisal bias, the bigger the default hazard is. Unobserved heterogeneity exists in both default and prepayment before appraisal bias is controlled, indicating existence of omitted variables. The heterogeneity disappears in the default space after controlling for appraisal bias, suggesting once again the importance of accounting for appraisal bias in mortgage termination studies.

The paper is the first to measure origination appraisal bias on loan level and examine its effect on mortgage terminations; it demonstrates the importance of appraisal bias in default hazard and establishes the necessity to control for it in mortgage termination studies; it is the first to replace the widely used yet inappropriate approximation likelihood function by an exact likelihood function to freely estimate the weights for the two hazards; it applies the Heckman-Singer semiparametric heterogeneity distribution to allow for independence between default and prepayment heterogeneity.

Chapter 2: How Mortgage Appraisers Produce Biased Appraisals by Inflating Property Value – A Theoretical Illustration

2.1 Introduction

It is well recognized that information asymmetry between sellers and buyers can lead to market inefficiency, thanks to the seminal work lead by Akerlof (1970). Ever since then, information economics gradually stepped out of the shadow and became a research area that draws great attentions from economists. In addition to recognizing this problem, researchers who work on information economics attempt to develop mechanisms to achieve efficient outcomes by reducing information asymmetry in the market (Akerlof 1970; Spence 1974; Stiglitz 1976, 1977).

Although information asymmetry exists extensively in the market, economic parties are always able to develop instruments to mitigate it. For example, product warranties, certified mechanical inspections, skills certifications are all instruments that can reduce information asymmetry, to name a few. This paper analyzes a phenomenon that stems from information asymmetry structure in the residential mortgage market that had been well maintained historically but not any more, thanks to drastic changes in the market environment. This phenomenon is referred to as appraisal bias, it refers to actions taken by mortgage lenders who intentionally inflate collateral property values to under report default risks in order to maximize their profit at the expense of mortgage investors.

This paper is organized as follows: section 2 introduces residential mortgage market, its composition and participants. It also explains in detail the incentive and existence of appraisal bias under the current market environment. Section 3 reviews literatures on

market efficiency, information asymmetry and various mechanisms that market participants introduced to mitigate the problem. It follows with a review on a particular game introduced by Crawford and Sobel (1982) as cheap talk. Cheap talk has the unique property that it is costless to declare players' types, and consequently a babbling Bayesian equilibrium shall be reached. This game structure can be applied to mortgage market and helps explain the existence of appraisal bias. In section 4, I shall prove how the introduction of semi-verifiability into the appraisal bias cheap talk game structure can help construct a non-babbling Bayesian equilibrium in an asymmetric information environment. Section 5 concludes the paper and provides suggestions for future improvements.

2.2 Residential Mortgage Market

What is residential mortgage market? Residential mortgage market is the origination, sale, and servicing of mortgage loans secured by residential real estate (MBA of America). The residential real estate industry is one of the largest sectors of the economy, constituting 14 percent of U.S. total GDP in 2003 (Bureau of Economic Analysis: 2003 NIPA Table for GDP). It is also a significant contributor to the U.S. economy, providing millions of job opportunities and generating hundreds of billions of dollars of economic output each year. Real estate is also an important source of wealth building for homeowners, with home equity serving as the largest share of household wealth. This is a market that has enjoyed a strong growth in the past years, with an average growth rate

twice as much as that of the GDP¹. In dollar terms, U.S. residential single-family mortgage originations totaled \$2.6 trillion in 2002, 29 percent higher than that in 2001. U.S. residential MDO in 2003 reached a stunning \$7.8 trillion. This is one of the markets that keep the U.S. economy moving forward.

The U.S mortgage market consists of primary and secondary markets. The primary mortgage market is where new loans are created. Borrowers who seek mortgage credit to finance real estate apply for mortgages from mortgage lenders who provide long-term funds with fixed and variable rates of interest. This process includes origination, processing, underwriting, and closing of the loan. The secondary mortgage market is where investors and mortgage lenders buy and sell existing and prospective loans as investment tools. The secondary mortgage market provides liquidity to allow mortgage originators to meet immediate needs for capital and enables investors to invest in mortgages easily. It assists the flow of capital from cash surplus areas where available capital exceeds credit demands to areas with cash shortage. By balancing capital distribution, geographical differences in interest rates disappear, making rates competitive nationally. To summarize, the primary market involves an extension of credit to borrower, and the secondary market markets a sale of that credit instrument.

¹ During the last decade when U.S. GDP sustained an average 3.5 percent growth rate, the U.S. mortgage debt outstanding (MDO) yielded a seven percent annual growth rate. Between 2001 and 2002, MDO's growth rate further escalated to 11.3 percent and 12.0 percent, making 2001 –2002 the first two consecutive years of double-digit residential mortgage debt outstanding growth since 1988-1989 (Fannie Mae 2002 Annual Report, U.S. Census Bureau 2002).

There are many participants in the mortgage market. They include borrowers, lenders, government agencies, government-sponsored enterprises (GSEs), private agencies and investors.

Mortgage market participants and the relationship among them can be depicted by the following flow chart.

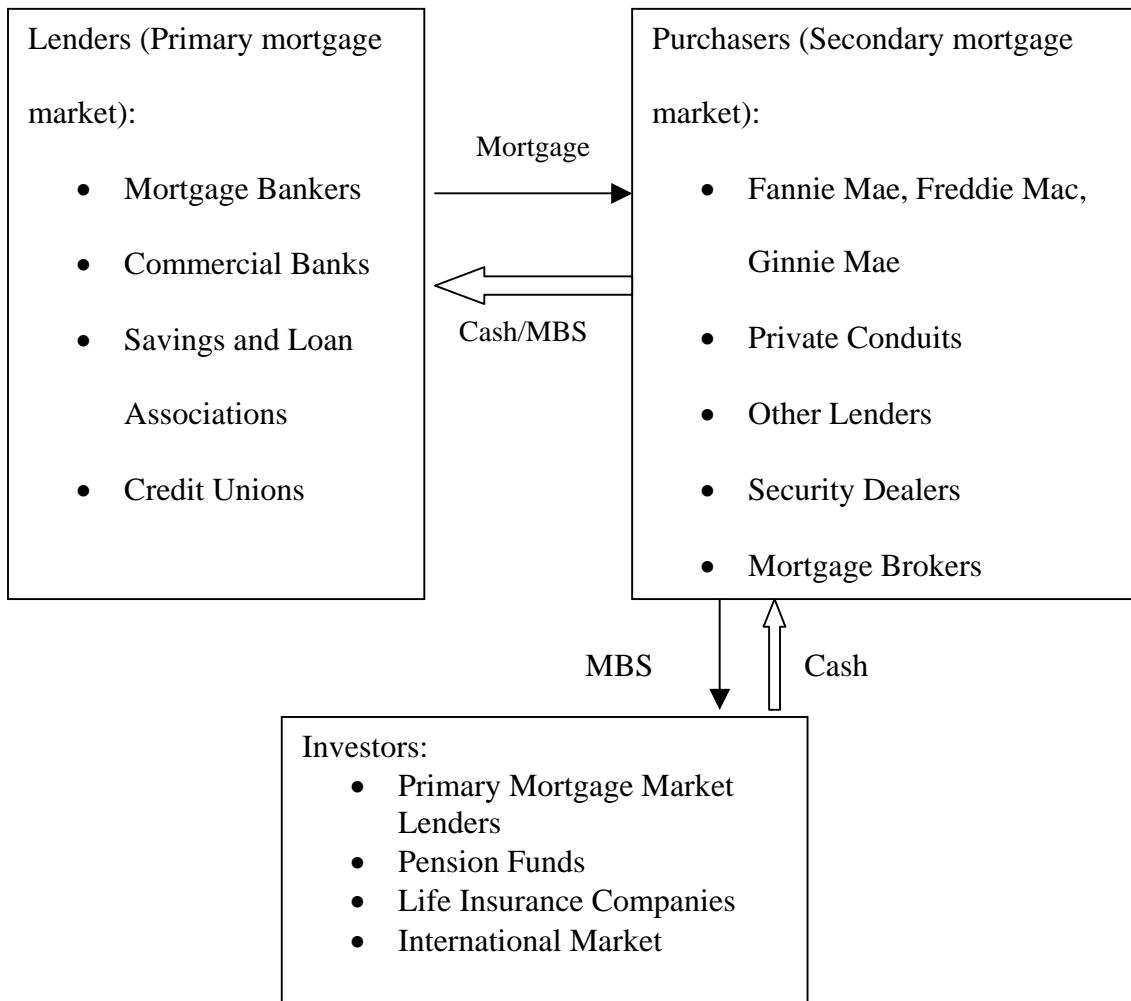


Figure 1. Market Participants and Their Relationship in the Primary and Secondary Mortgage Market

Among the above participants, the relationship between mortgage lenders and GSEs is of particular interest to this paper. Mortgage lenders include mortgage bankers, commercial banks, mutual savings banks, savings and loan associations, credit unions and others that are capable of lending money to borrowers to finance mortgages. These institutions may hold the mortgages they originated in their portfolio for investment purposes, they may also sell those mortgages into the secondary mortgage market for liquidity concerns. In order to reduce mortgage default risks faced to investors, most mortgages sold to the secondary mortgage market are guaranteed against default risks by GSEs.

There are two GSEs in the secondary market: Federal National Mortgage Association (Fannie Mae) and the Federal Home Loan Mortgage Corporation (Freddie Mac). Fannie Mae and Freddie Mac are identified as GSEs because the U.S. federal government was involved in the creation of them. In 1938 and 1970, the U.S. Congress established Fannie Mae and Freddie Mac respectively to expand the flow of mortgage money by creating a secondary market. Currently, both of them are private corporations, although they still obtain a credit line at the U.S. Treasury. The importance of the GSE classification is that the debts of the GSEs are perceived in the marketplace as “United States Agency” securities and thus are sold at lower rates, even though neither is a part of the U. S. federal government.

Fannie Mae and Freddie Mac purchase mortgages from mortgage lenders and issue mortgage-backed securities (MBS) to sell to the investors. In principle, a lender delivers a package of mortgages to either of the two GSEs. Fannie Mae or Freddie Mac issues

MBSs to the lender, guarantees timely payment of principal and interest to the investors, and assumes the credit risk on the loans. In return for the credit guarantee, the lender is charged a guaranty fee. He then sells the guaranteed MBS to investors through Wall Street dealers.

Fannie Mae and Freddie Mac issue the biggest volumes of mortgage-backed securities in the secondary mortgage market. Jointly they hold on their balance sheet nearly \$300 billion home mortgages, plus an additional \$1.2 trillion of MBS's, compared with a total \$6.6 trillion of home mortgages outstanding in the United States (FM Watch, 2003).

Fannie Mae, Freddie Mac, and the government owned agency Ginnie Mae together account for 58 percent of residential MBS as of September 2004 in the U.S. market (ETF Connect). These MBSs are popular among lenders because standardization has made loan sales efficient and effective. Investors also favor these MBS products because repayment of the securities is backed by the full faith and credit of government-sponsored enterprises. As a result, mortgage money needed to finance residential and commercial properties is available in every geographic region of the United States.

There are several steps in a typical mortgage transaction involving selling loans to GSEs. It starts with a mortgage borrower who applies for a loan from a mortgage lender. The lender then underwrites the loan upon receiving the application and makes an offer to the borrower. The offer is often in the form of a combination of an interest rate, loan term and monthly payment. It is determined based on pricing schedules set up by GSEs in their automated underwriting tools. The borrower can then reject or accept the offer. The

underwriting process needs to comply with Fannie Mae or Freddie Mac's standard for delivery. Once accepted by Fannie Mae or Freddie Mac, the loan package is sold and delivered to its intended purchaser. The lender generates revenue by enjoying an origination fee from the borrower and a servicing fee from the investor for future loan administration. In addition, the lender will be free of repayment obligations should the loan default in the future. Fannie Mae or Freddie Mac charges guarantee fee and enjoys net interest income from the interest spread. They assume default risk and guarantee timely payment of the mortgage principal and interest even if the borrower fails to fulfill his mortgage obligations. In essence, the lender passes the mortgage and associated default risk to Fannie Mae or Freddie Mac, depending on who agrees to guarantee the mortgage.

This paper focuses on the unique relationship between mortgage lenders and GSEs, the information asymmetry structure between them and the consequence of information asymmetry.

The structure of the secondary mortgage market makes information asymmetric between mortgage lenders and GSEs, and consequently provides existence for an incentive problem. According to charter rules, GSEs are not allowed to participate in the primary mortgage market; they need to rely on solely the information provided by mortgage lenders at loan origination to evaluate the risks associated with each application and determine its pricing schedules and purchasing decisions². Mortgage lenders, on the other

² However, GSEs can conduct costly post foreclosure reviews after default has occurred, but successful fraud detection is not guaranteed.

hand, know more about the borrower and property characteristics than GSEs do. They have incentive to behave in ways to maximize their profit at the expense of the GSEs provided that their actions are not detectable.

For example, a profit maximizing lender would manipulate the mortgage application information in order to achieve favorable outcome if there is no consequences of their actions. Two types of application information that affect pricing decisions are required to submit to GSE's in addition to mortgage parameters. One is borrower specific information; the other is property specific information. Borrower information includes a borrower's financial conditions, his credit ratings, his debts and other financial obligations, etc. Property information includes current property value. Between the two types of information, borrower specific information is provided and verified independently by third parties such as credit bureaus, the borrower's employer, bank and IRS. On the other hand, the mortgage lender is the only source that provides information regarding the collateral property value when property purchase price does not exist, for example, in refinance transactions³. Since current property value at mortgage origination significantly influences the price being offered, mortgage lenders have incentive to manipulate property values in order to maximize their profit. The fact that fraud on loans delivered to Fannie Mae or Freddie Mac cannot be costlessly detected further strengthens that incentive.

³ In purchase transactions where both purchase price and the appraisal value of the property exist, the current property value is determined by the minimum of the property purchase transaction price and the appraisal price. In refinance transactions where no purchase price exists, the current property value is determined solely based on the lender-provided appraisal price.

Real estate mortgage market has realized long ago the information asymmetry and incentives embedded in this market. In order to reduce the asymmetric information and incentives, it requires an objective third party to verify the truthfulness of the collateral value information. This job is done entirely by the real estate appraisal industry. Real estate appraisers are the channels chosen by mortgage lenders and borrowers to deliver information from one party to the other. Their purposes are to protect real estate sellers and buyers from frauds as well as mortgage lenders from credit loss.

The real estate appraisal industry certifies the value of properties that are about to be sold or refinanced into the mortgage market. They provide an estimate of the property value by researching comparable sales in the same area and by inspecting the property condition. Their estimate helps mortgage lenders determine if the property they are about to lend money for are worth the mortgage amount. In refinance transactions where no purchase price is available on the property, their estimate is the one and only information that lenders and GSEs rely on for pricing decisions. In theory, real estate appraisers should maintain independence of mortgage borrowers and mortgage lenders in order to present objective and unbiased opinion on the value of the property. However, recent changes in mortgage market structures and intensive competition among appraisers provide incentives to real estate appraisers not to fulfill their duties objectively.

The residential mortgage market experienced significant changes during the past decades. During 1970s and 1980s when home ownership was low, the residential mortgage market

was dominated with low loan-to-value (LTV)⁴ ratio purchase money mortgages.

Purchase money mortgages, or purchase transactions are ones whose purpose is to purchase a house, either new or existing. It involves an ownership change of hand. In purchase money mortgages, there exist purchase transaction prices that are jointly determined by the property sellers and the property buyers. In purchase money mortgage transactions, both lenders and secondary market mortgage buyers observe the purchase transaction prices coupled with the appraisal prices. Secondary market participants such as GSEs rely on the combination of the two sources of information on property value to determine appropriate pricing schedules. The existence of an actual purchase transaction price deters the appraisers to deviate much from it.

Apart from a large concentration of purchase money mortgages, the mortgage market in early period also enjoyed a pool of high quality borrowers with low credit risks. In those days, mortgage loans were offered exclusively to high quality borrowers who could afford to put down large down payments, resulting in a small credit risk for the mortgage lenders. In addition, the incentive to defraud did not exist because mortgage lenders typically held the mortgages they originated in their portfolio instead of selling them to GSEs; it was in their best interest to report the accurate collateral value to fully understand the credit risks. As a result, those market characteristics significantly relaxed the market dependence from real estate appraisers. Appraisers simply were asked to verify the reasonability of the purchase price on the property. Rarely were appraisers asked to determine if the borrower would have enough equity in the transaction so that

⁴ Loan-to-value ratio (LTV) measures the ratio of loan amount to equity; the higher the LTV is, the higher the credit risk is.

the lender would be protected in case of default or should mortgage insurance be required. Most lenders simply counted on rising housing prices, making such decisions irrelevant.

Today, every aspect of the old environment is gone. For example with the significant reduction in long term interest rates in the 1990s, refinancing and home equity loans have come to dominate the residential mortgage market, making purchase money mortgage less than one quarter of total mortgage transactions. Refinance mortgages are loans whose purpose is to payoff an existing mortgage and refinance into a new mortgage; it does not involve a change of ownership. Unlike in purchase money transaction where an actual sale price is available, there does not exist transaction price on the refinance property. Hence appraisal values are the sole source for GSEs to determine pricings, making appraisers' role unprecedented important in the mortgage market.

Changes in the market also encompass an increasing share of high LTV loans with less than 20 percent down payments in the market. Those loans are relatively riskier in terms of credit loss than their safe counterparts of the old days. However, due to their affordability feature, they continue to be on the rise and account for more than a quarter of total mortgage products since 1995. In addition, more than half of all loans are originated with the intension to be sold as MBSs (Fannie Mae Annual Report 2003). Lastly, with the rising home ownership and increasing demand for housing, mortgage lenders began to expand to borrowers with sub-prime credit history and financial conditions, resulting in a higher consumer delinquency rates on mortgage products. All

these new market trends call for stringent evaluations and pricing decisions in order to reduce credit loss. As a result, real estate appraiser's estimates on collateral values are more and more heavily relied upon.

However, the mortgage market fails to adjust the market structure to prevent inefficient outcomes under the new environment.

Previously, real estate appraisers worked closely with mortgage lenders in order to reduce the search cost for appraisers and to increase efficiency. Consequently, mortgage lenders rely on their in-house or designated appraisers to conduct collateral evaluations. Although harmless in the old environment, this relationship reduces the independent role of real estate appraisers and causes them to share the same objectives as mortgage lenders. As a result, appraisers behave in ways that maximize mortgage lenders' profit at the expense of maintaining independence. Pressure from mortgage lenders becomes an important reason for appraisers to compromise their principle. According to an article on the Denver Post, "The people who certify what homes are worth when they are sold or refinanced say they are being pressured to inflate their numbers to ensure that the lending deals go through" (The Denver Post, July 20, 2003). Appraisers acknowledge they are under pervasive client pressure and have a tendency to give in when exerted by important clients (Kinnard, Lenk and Worzala 1997). Pressures from lenders are often in the forms of threat of withholding fee payments or reduction in future assignments (Levy and Schuck 1999).

Apart from pressures exerted by mortgage lenders, competition in the appraisal industry accelerates the trend of unethical practices. On average, an appraiser charges \$300 to \$600 per case; it is an unstable source of income for real estate appraisers. Even in the refinance boom year of 1994, there were only 100 transactions per appraiser on average, generating an average income of \$25,000 to \$30,000 for appraisers. Their average incomes declined by more than one-third over the next two years after the refinance slowed down. Consequently, real estate appraisers have tendency to succumb to the market trends.

The common fraudulent practice is to arbitrarily inflate the property value in the absence of purchase price in refinance transactions. Instead of conducting comparable research and in-house research, the appraiser finds out the size of the underlying mortgage that the homeowner wants to refinance or how big a credit line he wants to establish. Often appraisers are given this information by the mortgage borrower or the mortgage lender. Or the appraiser finds out what interest rate the homeowner wants to qualify for. In other cases, appraisers inflate property value to help borrower avoid mortgage insurance, which could be as high as half of the mortgage payment. This systematic fraudulent action is referred to as appraisal bias.

Below is an example of how appraisal bias works in favor of mortgage borrowers in a refinance transaction. Suppose a homeowner chose an 80 percent LTV loan for a property worth \$100,000 in 1990. Let's assume his property appreciated to \$110,000 in 2000 and he chose to refinance and cash in some of his home equity appreciation in a declining

interest rate environment. Suppose further that he had paid down \$10,000 of the mortgage payment and was left with \$70,000 as outstanding unpaid principal balance in 2000. Thanks to his equity appreciation, this borrower could apply for up to \$88,000 in mortgage amount and still keep his LTV equal to 20 percent. This allowed him to have \$18,000 in cash after paying off his current mortgage balance of \$70,000. Appraisal bias could escalate this borrower's ability to take more cash out of his mortgage yet keep the LTV intact.

Mortgage lenders have strong incentive to maintain the existence of appraisal bias because it brings substantial monetary values to them. From a revenue perspective, holding everything else constant a higher property value lowers the interest rate for mortgage borrowers, making it easy for the mortgage deal to go through and hence increase the revenue. From a cost perspective, a high property value mortgage lowers the credit risk and in turn lowers the guarantee fee charged by GSEs. Therefore, mortgage lenders have strong motivation to maintain a system of appraisal incentives.

Appraisal bias can lead to significant loss to the GSEs and the society as a whole, although it is a favorite to mortgage lenders. Because equity influence default and appraisals are used to estimate borrower's equity in the absence of a transaction price, over-valuation of collaterals results in underestimation of the default risk. The underreported default risk is passed to secondary mortgage market agents and can cause significant consequences. Many times these fraudulent loans end up in foreclosure, resulting in financial losses to mortgage buyers. The total price of mortgage fraud could

be as high as \$120 billion, with the Mortgage Banker Association (MBA) estimating over \$60 billion from its reporting membership (The Denver Post July 20, 2003). For example, among others, fraud appraisals had lead to failures of many thrifts and saving banks, which had to be bailed out by the federal government in the saving and loan crisis in the 1980s. Thousands of foreclosed properties were significantly overvalued and the revues from sales of them were far enough to recoup the losses. A report by the U. S. House of Representatives Committee on Government Operations⁵ concluded that 10 to 15 percent of the \$1.3 billion in losses suffered by private mortgage insurers in 1984-1985 could be attributed to inaccurate and fraudulent appraisals, and that 10 to 40 percent of the \$420 million in loan losses at the Veterans Administration in 1987 was caused by inaccurate or dishonest appraisals or other appraisal-related deficiencies. The FBI recently reported that 10 to 15 percent of all loan applications contains material misrepresentations, such as inflated property valuations (The Denver Post July 20, 2003). “Appraisal bias, spreading like a cancer, is eating away at the industry’s moral foundation” (NAIFA – AppraiserE-Gram, 2001).

Empirical research also documented the evidence of appraisal bias. Using a sample of mortgages purchased by Fannie Mae, Cho and Megbolugbe find that in more than 80 percent of the cases, the appraisal is between 0 and 5 percent above the transaction purchase price, in only 5 percent of the cases is the appraisal lower (Cho and Megbolugbe 1996). Chinloy, Cho and Megbolugbe show that appraisals are systematically higher than purchase data using 3.7 million repeat transactions on mortgages bought by Fannie Mae

⁵ “Impact of Appraisal Problems on Real Estate Lending, Mortgage Insurance, and Investment in the Secondary Market,” a.k.a. “1986 Barnard Report”.

and Freddie Mac by using monthly data from January 1975 to December 1993 (Chinloy, Cho and Megbolugbe 1996).

The alternative world where property appraisers are commissioned by GSEs, instead of mortgage lenders cannot solve the problem either. In this case, the situation is reversed with appraisers share the same incentive with GSEs. GSEs would like to increase the guarantee fee even when the default risk is not high, making appraisers likely to under report the property value. Therefore, reversing the roles does not solve the incentive problem.

In summary, the existence of information asymmetry between mortgage lenders and GSEs creates an environment for a problem to exist. This problem is realized through a phenomenon in refinance transactions referred to as appraisal bias. Because the truthfulness of the property value is costly for GSEs to verify independently, mortgage lenders with their in-house appraisers have incentive to inflate the property value in order to maximize their profit at the expense of GSEs. Complementing the anecdotal evidence and empirical research on the existence of appraisal bias, the remaining parts of this paper attempt to analyze the existence of appraisal bias from a theoretical point of view.

2.3 Literature Review

Information has proven to be very valuable to individuals and entities in the economy because shortage of it can impair the economy. One form of information shortage is reflected by the existence of asymmetric information among economic agents.

Asymmetries of information are pervasive in economic relationships: customers know more about their tastes than firms, firms know more about their costs than the government, and all agents take actions that are at least partly unobservable. Such information asymmetry can have a profound impact on market efficiency and organizational structure (Akerlof 1970, Hart 1989). For example, in Akerlof (1970)'s seminal work on lemons, he shows that a market may function very badly if the informed party cannot communicate effectively the quality of the good it is selling to the uninformed party. His result shows that information asymmetry about product quality can hinder the function of the market from where all quality products are traded to the point where only the worst quality goods are traded, and ultimately the market unravels. The disappearance of markets for high quality commodities caused by information asymmetry reduces market efficiency, whereas such markets, under competitive structure with complete information, could reach Pareto efficiency (Vickers 1995).

The asymmetric information between an informed party and an uninformed party in general takes on two forms: hidden action where the private information is regarding to what an agent does, and hidden information where the private information bears on who the agent is (Salanie 1997). Economists who study the economics of information also take on different paths to attenuate the problem of asymmetric information based on the two types of private information. They include adverse selection models, signaling models and moral hazard models. Adverse selection models, also named as self-selection or screening models, apply to cases where the uninformed party is imperfectly informed of who the informed party is, i.e. the characteristics of him, and initiate the first move

(Stiglitz 1976). Signaling models share the same information availability with the adverse selection model, but differ in that the informed party moves first to signal who they are to the uninformed party (Spence 1971, 1974). Lastly, in moral hazard, also called agency problem cases, the uninformed party is imperfectly informed of the action of the informed party and acts first (Holmström and Milgrom 1987).

In adverse selection cases, the uninformed agent designs a menu of contracts for the informed agents to choose such that the informed agents shall reveal themselves through their choices. Examples of such include life insurance where the insurer offers several insurance packages to tailor for specific risk class, credit extensions where banks use different interest rates to assess borrowers' default risks, firm regulations where regulators set up contracts such that firms accurately reveal their cost in choosing the contracts, and optimal taxations where tax on production help government to maximize social welfare by implementing an allocation of consumption and labor (Maskin and Riley 1984; Stiglitz and Weiss 1981; Laffont 1994; Mirrlees 1986).

Extensions are made on adverse selection models including competition within the informed parties as well as the uninformed parties, risk-averse agents, and asymmetric information on both sides (Rothschild and Stiglitz 1976, Salop and Salop, 1976; Champsaur and Rochet 1989; Salanie 1990). For example, Rothschild and Stiglitz (1976) studied a competitive insurance market and showed that a pooling equilibrium where all people buy the same insurance contract despite accident probability cannot hold. Such a market may end up either with a separating equilibrium where people buy different

insurance contracts in accordance to their accident probability, or there exists no equilibrium at all. As a result, equilibrium strategies can be drastically different from standard adverse selection models in some of those extensions.

Different from screening models, signaling refers to games that the informed party moves first by sending a signal that may reveal information relating to his type. And the uninformed party then tries to decrypt these signals by using interpretative scheme.

Spence classified the signaling into contingent contract and exogenous costly signals (Spence 1976). Contingent contract is defined as contracts where “there involve a menu of options for the seller that are created by virtue of the buyer’s subsequent ability to observe the product quality directly, and, to transact with the seller at that point – hence the terms contingent contract” (Spence 1976). One good example of contingent contract is warranties offered by product manufactures. Heal (1977) stated that existence of a warranty provided incentives for producers to improve quality, at least to the extent to reduce the chances of falling below the product warranty. Spence (1977) perceived warranty as a signal where high quality producers could afford to offer more complete warranty while low quality producers choose not to. However, the actual effect and design of warranty can be more complicated than just a signal (Cooper and Ross 1988).

Exogenous costly signals are defined as “activities engaged in by the seller, which have costs that vary with product quality, independent of the buyer’s response to the activity” (Spence 1976). Education is a good example of an exogenous costly signal. Spence (1974) explored the effect of education as a signal for productivity of employees in the

labor market. By assuming a disutility suffered only by low productivity labors, Spence derived that the successful completion of a higher degree signals the employee's productivity type to prospective employers. Note that the costly education expenditures are initiated before the labor is supplied, which is independent of possible responses of employers; hence the name exogenous costly signal applies. Other signaling and screening devices studied include price (Wolinsky 1983) and reputation (Shapiro 1982, 1983, Rogerson 1983, and Allen 1984) etc⁶.

The reason why exogenous costly signals enable the uninformed party to separate the agents is due to the existence of a signal whose cost varies with the types of the informed party. Spence (1976) suggested that verbal declaration could also function as a potential source of information. However, words are cheap and free of costs, which hardly provide convincing information. Signals where there exists no cost for the agents who send them are referred to as costless signals.

Crawford and Sobel (1982) are the first to analyze cheap-talk games where all signals are costless. Cheap-talk game is a special case of signaling game where the signal sender's messages are costless, nonbinding and non-verifiable. For that reason, such talks cannot be informative in Spence's signaling games because the signal senders have the same preference over the receiver's actions. Hence there always exists a non-informative babbling equilibrium in cheap-talk games where all signals shall be ignored.

⁶ Wolinsky (1983) argued that high price is associated with high quality product because high cost level is required to consistently deliver the quality; hence a markup needs to be earned to support the quality. Shapiro (1983) argued that reputation can also serve as an effective instrument to differentiate product quality given ample returns.

Nevertheless, Crawford and Sobel (1982) show that there also exist equilibria that reveal some information. Their research suggests that more communication can occur through cheap talk when the player's preferences are more closely aligned, but perfect communication cannot occur unless the players' preferences are perfectly aligned.

Examples of cheap talk include policy announcement by the Federal Reserve, veto threats by the president, information transmission in debate and union voice to the management (Stein 1989; Mathews 1989; Austen and Smith 1990; Farrell and Gibbons 1991). I will model the appraisal bias phenomenon based on the cheap-talk game.

The last family of games that reflect information asymmetry is moral hazard, or principal-agent models. In moral hazard settings, an agent takes an action that affects his utility as well as that of the principal; the principal only observes the outcome, which is an imperfect signal of the action taken. As a result, the action the agent would choose spontaneously is not Pareto optimal. Examples of moral hazard appear almost everywhere in the economy. For example, it can be found in the employer-employee relationship, property insurer and the insured relationship, car owner and his mechanic's relationship, patient and doctor relationship. Grossman and Hart (1983) suggest an optimal contract offered to the agent that maximizes the principal's utility. The optimal contract should trade off risk sharing and incentives, which are the basis of moral hazard.

In summary, information asymmetry prevails in economic relationships, which lead to market inefficiency and sub-optimal results for the economy. Researchers who work on

contract theory and information economics develop families of economic models to study those relationships using perfect Bayesian equilibrium concepts and game theory.

2.4 Model

In this section, I will set up a dynamic game with incomplete information to illustrate the relationship between mortgage lenders and GSEs. By solving the perfect Bayesian equilibria of this game, I shall prove that there exists a hybrid equilibrium where it is optimal for mortgage lenders to randomize between maintaining honesty and defrauding given GSEs' actions. This confirms the empirical evidence that appraisal bias exists in some but not all mortgage cases. I shall also provide a scope of parameter conditions such that a pooling equilibrium where all lenders choose to behave honestly/dishonestly exists, as well as conditions for a separating equilibrium to exist where lenders behave differently according to the types of mortgages they receive.

The following section will start with a revisit of the mortgage market structure, an introduction to the game structure, and payoff functions of each player. Throughout the text, I shall introduce assumptions and notations that are necessary for solving the equilibrium conditions. After I solve for each equilibrium condition, I shall present a comparative statistics analysis to measure the effect of parameters on each equilibrium outcome.

2.4.1 Market Structure Revisit

As mentioned in previous section, there are numerous mortgage lenders in the market competing for borrowers. They seek GSEs' guarantee service to ensure timely payment of principal and interests to mortgage investors. In exchange, mortgage lenders pay a guarantee fee to GSEs for the credit risks insurance. GSEs, who serve as credit risk insurers, rely on lenders to provide estimated values of the collaterals to evaluate borrowers' credit risks. Once the risk is known, GSEs will charge a risk based guarantee fee specified for different levels of default risks. I will interchange "GSEs" with "credit risk insurers" in the subsequent texts since their role is to insure the credit risk on mortgages sold to mortgage investors.

2.4.2 Setup

Following Stiglitz (1976), I assume that the types of borrowers are uniformly distributed along a line. I also assume that the primary mortgage market is a Hotelling market where there exists a search cost s that is greater than zero for mortgage borrowers.

Assume there are two types of borrowers in the market; their only difference is that one is endowed with a high value house H , and the other is endowed with a low value house L . The proportion of borrowers who are endowed with high value houses is η , and the proportion of those who are endowed with low value houses is $1 - \eta$. Both types of borrowers want to apply for a mortgage at the same amount V , with the house being the collateral. They search for the lowest interest rate offered in the market at a positive search cost s . The interest rate is exogenously determined such that loans with low

value collaterals and high value collateral are charged differently. Because holding everything else equal a high value property has a low probability of default and a low value property has a high probability of default, interest rate charged on mortgages with high value collaterals is lower than that charged on mortgages with low value collaterals to compensate the credit risks.

The lifetime of a mortgage is assumed to be two periods. In the first period, the lender originates the mortgage and seeks guarantees from credit risk insurers. In the second period, the outcome of the mortgage is realized; a loan either defaults, or pays off. I assume the discount factor between the two periods is one⁷.

Lenders are assumed to know the true property value after the appraisal process. They then report an appraisal value to the mortgage insurer and transfer the mortgage risk at the guarantee price determined based on the reported appraisal value. Everything held equal, high value property has low probability of default and low value property has high probability of default. Therefore, lenders with high value collaterals mortgage will be charged a low guarantee fee whereas lenders with low value collateral mortgage will be charged a high guarantee fee. Assume at the moment that the guarantee fees are exogenously determined.

If no verifiability is available, then this game falls under the classic cheap talk game scenario where all lenders' preference is to claim high value regardless of the true value

⁷ One can assume the discount factor is less than one and still obtain similar result with one more parameter condition on the discount factor.

of the property. Hence a babbling equilibrium is the only outcome. However, market can be restored by letting credit risk insurers costlessly verify fraud and pass the cost of default back to lenders. This cost depends on the realization of the state of nature and the credit risk insurers' actions. This semi-verifiability helps construct non-babbling equilibria in this game.

Credit risk insurers, knowing that lenders have incentive to cheat, will randomly choose to conduct a costly quality review on defaulted loans at a positive cost C . Assume that once reviewed fraudulent loans will be discovered with probability one⁸. Consequently, lenders who originated those loans will be forced to incur the default loss and the fraudulent loans shall be returned to them. In this sense, the signal sent by a mortgage lender on whether the collateral is high value or low value is costless a priori, but can be costly posteriori depending on the outcome of the loan and the credit insurer's action.

Another potential time to review is when the loan is delivered, as opposed to when default occurs. However, just like insurance verification is conducted when the claim is made, not at the time the insurance is purchased, GSEs should also verify the truthfulness of the information when it matters. Verification after default has occurred makes sense because the cost of verification is high and default is a rare event.

Under complete and symmetric information scenario, mortgage insurers can observe the true value of the collaterals, correctly infer the probability of default and hence charge

⁸ In reality, the success rate of fraud discovering is less than one. However, one can assume that the review cost encompasses the cost of unsuccessful reviews.

guarantee fees correspondingly. Mortgage lenders, under information symmetry, do not have incentive to deviate from reporting the property value truthfully. The role of mortgage lenders is simply to transfer information in a complete and perfect fashion. Information asymmetry, where mortgage insurers cannot observe the true default probability of the properties, creates incentives for lenders to over appraise the property to reduce the guarantee fees.

2.4.3 Game Structure

I assume the game played between a mortgage insurer and a mortgage lender is a dynamic game with incomplete information. The mortgage lender is the agent who has private information on a loan's default risk. He sends out a signal regarding the quality of his "product" to the mortgage insurer. The mortgage insurer receives the signal and chooses his action accordingly without the ability to verify the truthfulness of the signal. This game differs from Spence's famous job market signaling game in that the sender's signal is exogenously costly in Spence's model, whereas it is costless and does not depend on lender's type in this model. Therefore, it can be viewed as a "Cheap Talk" game.

Cheap talk game is a special case of signaling games, where the sender's messages are costless, nonbinding, and not verifiable. Such talk cannot be informative in signaling games that require costly messages to separate the types. Thus, for cheap talk to be informative, one necessary condition is that different types of senders have different preferences over the receiver's actions (Gibbons 1992). The model analyzed here,

however, does not satisfy this condition because all lenders prefer a low guarantee fee to a high guarantee fee regardless what types of loans they are dealt with. Hence a babbling equilibrium where all signals are uninformative arises. However, adding a condition that the signals are costly posterior can guarantee a non-babbling equilibrium. This condition is reflected by the fact that the receiver chooses his action after observing the sender's signal as well as the outcome of the loan. The loan outcome conditional on sender's signal helps update the receiver's belief on the lender's type. Therefore, the semi-verifiability of the costless signal can be proven to affect signals sent by different types and hence makes this model different from other games in the "Cheap Talk" family.

The timing of the game is as follows:

1. Nature determines the type of the borrower, which is reflected by the type of property that he is endowed with. The proportion of borrowers who are endowed with high value houses is η , and the proportion of those who are endowed with low value houses is $1 - \eta$. A high value property, denoted by H has a default probability p_H , and a low value property, denoted by L has a default probability p_L . Note that the subscripts H, L correspond to the value of the property, not the value of the probability. I shall use the subscripts in the same fashion in the subsequent notations. It is assumed that $p_H < p_L$ because a high value property has a lower default probability than a low value property.
2. A borrower searches for the lowest possible interest rate offered with a positive search cost $s > 0$. A high value borrower is offered with an interest rate r_H , and a low value borrower is offered r_L . Note that $r_H < r_L$ because high value borrowers

are offered with a market determined low interest rate; low value borrowers are offered with a market determined high interest rate due to credit risk differences. Some low value borrowers may be offered an interest rate as low as a high value borrower receives if the lender chooses to lie about the property value. Because of the positive search cost, I assume not every low value borrower can find a compromising lender within his search cost constraint.

3. The lender, after appraising the property, obtains the true property value H or L , default probability p_H or p_L , and chooses an appraisal value \hat{H} or \hat{L} to report to the credit risk insurer.

4. Borrowers are willing to accept any offer that yields utilities greater than or equal to their reservation utilities. Since the interest rate offered to a high value borrower is lower than that offered to a low value borrower, borrowers with low collateral value prefer to receive the offer designed for the high value borrowers.

5. The mortgage insurer observes the appraisal value and the interest rate offered, but not the property's true value or its true probability of default. It offers a guarantee price $g_{\hat{H}}$ or $g_{\hat{L}}$ to the lender according to the claimed property value

\hat{H} or \hat{L} . The guarantee fee is assumed to be determined exogenously.

6. Nature moves again and reveals the outcome of the loans. A loan either defaults or pays off in its lifetime. The mortgage insurer takes no action toward loans that are safely paid off because he does not incur any credit loss on a paid-off loan regardless of fraud. When a loan defaults, the mortgage insurer will choose whether or not to order a quality review with a probability q at a positive cost

$C > 0$. Once the review is conducted, fraud discovery is guaranteed and the fraudulent lender will be forced to incur the default loss. The review has no effect on honest lenders.

7. Game is over.

The above notations can be summarized as follows:

H denotes a high value property,

L denotes a low value property,

\hat{H} denotes a property with a reported high value, regardless of its true value,

\hat{L} denotes a property with a reported low value, regardless of its true value,

D denotes event of default,

ND denotes event of pay off,

R denotes event of quality review by credit insurer,

NR denotes no quality review.

The probability of certain event is as follows:

The probability of a high value property on the market is:

$$P(H) = \eta \quad (a)$$

The probability of a low value property on the market is:

$$P(L) = 1 - \eta \quad (b)$$

The default probability of a high value property is:

$$P(D | H) = p_H \quad (c)$$

The default probability of a low value property is:

$$P(D|L) = p_L \quad (d)$$

The probability of a quality review on a defaulted loan by the credit insurer is:

$$P(R|D) = q \quad (e)$$

The probability of a quality review on a paid off loan by the credit insurer is:

$$P(R|ND) = 0 \quad (f)$$

The action space for the lender comprises of two parts: one which he reports the collateral value truthfully, the other where he lies about the collateral value. The action space for the insurer has three components: one that he reviews the defaulted loan, the other he does not, the third one is where he takes no action if default does not occur. This is because a paid off loan, regardless if there exists appraisal fraud, does not impact the profitability for the insurer as long as the loan pays off as scheduled.

To simplify the notion, it is fair to assume that the payoff structure on a given mortgage is one where a fixed pass through rate of return is allocated to the investors in the capital market, a fixed administration and servicing fee is allocated to mortgage servicers, and the left over is up for negotiation between lenders and mortgage insurers. Specifically, the pricing structure for a mortgage can be expressed as:

$$r = i + f + T$$

$$T = o + g$$

where r is the coupon/note rate that the borrower pays, i is the pass through rate to the investors, f is servicing and administration fee and T is residual rent. The residual rent is divided between the lender as an origination fee o , and the credit risk insurer as a

guarantee fee g . Given the note rate, pass through rate and servicing fees, the higher the guarantee fee is, the lower the lender's return is.

The payoff structures for each player is as follows:

The combined payoff for the two players in the event of no default is $V \cdot T$, where V and T , as previously introduced denote the loan amount and the left over rent for the mortgage lender and credit risk insurer to share respectively. The mortgage lender receives $V(T - g_{\bullet})$ and the insurer receives $V \cdot g_{\bullet}$ as revenue. The guarantee fee g_{\bullet} is $g_{\hat{H}}$ for reported high value property or $g_{\hat{L}}$ for reported low value property. In the event of a default, the combined payoff is reduced to $V \cdot T - V$, where the defaulted loan amount V needs to be taken off from the combined revenue as a loss. This loss shall be absorbed by the mortgage lender and the credit risk insurer. When the mortgage insurer does not review the defaulted loan, the payoff for the lender is $V(T - g_{\bullet})$ and the payoff for the insurer is $V \cdot g_{\bullet} - V$. If the insurer reviews the loan, the two players' payoffs are $V(T - g_{\bullet}) - V$ for the lender and $V \cdot g_{\bullet} - C$ for the insurer if fraud is detected; they are $V(T - g_{\bullet})$ for the lender and $V \cdot g_{\bullet} - C - V$ for the insurer if fraud is not detected. Once again, the guarantee fee g_{\bullet} here is either $g_{\hat{H}}$ or $g_{\hat{L}}$ for reported high value property and reported low value property respectively.

The game can also be depicted in a normal-form representation. In the normal game representation, I define cheating as inaccurately report the property value. It includes cases that a lender with L value collateral reporting the property value as \hat{H} , and a

lender with H value collateral reporting the property value as \hat{L} to the credit risk insurer.

If a mortgage lender is dealt with a low value collateral L and the loan pays off, the total payoff between the mortgage lender and the credit insurer becomes $V \cdot T$. The mortgage lender's payoff as well as the payoff for the credit risk insurer is as follows depending on if he lies about the property value. Since the credit risk insurer does not review loans that paid off, there is only one action space on his side.

		Credit Risk Insurer	
		No Quality Review	
Mortgage Lender	Cheat	$V(T - g_{\hat{H}})$,	$V \cdot g_{\hat{H}}$
	No Cheat	$V(T - g_{\hat{L}})$,	$V \cdot g_{\hat{L}}$

If a mortgage lender is dealt with a low value collateral L and the loan defaults, the total payoff between the mortgage lender and the credit insurer becomes $V \cdot T - V$. His payoff as well as the payoff for the credit risk insurer is as follows depending on if he lies about the property value and if the credit insurer reviews the loan. Note that now the credit insurer has two actions, review and not review, because the loan has defaulted.

Credit Risk Insurer

		No Quality Review	Quality Review
Mortgage Lender	Cheat	$V(T - g_H^{\wedge}),$ $V \cdot g_H^{\wedge} - V$	$V(T - g_H^{\wedge}) - V,$ $V \cdot g_H^{\wedge} - C$
	No Cheat	$V(T - g_L^{\wedge}),$ $V \cdot g_L^{\wedge} - V$	$V(T - g_L^{\wedge}),$ $V \cdot g_L^{\wedge} - V - C$

If a mortgage lender is dealt with a high value collateral H and the loan pays off, the total payoff between the mortgage lender and the credit insurer becomes $V \cdot T$. His payoff as well as the payoff for the credit risk insurer is as follows.

Credit Risk Insurer

		No Quality Review	
Mortgage Lender	Cheat	$V(T - g_L^{\wedge}),$	$V \cdot g_L^{\wedge}$
	No Cheat	$V(T - g_H^{\wedge}),$	$V \cdot g_H^{\wedge}$

If the high value property H defaults, the total payoff between the mortgage lender and the credit insurer becomes $V \cdot T - V$. His payoff as well as the payoff for the credit risk insurer is as follows. Note that a high value lender will not be penalized if the review

discovers that the loan is actually a high value loan even though it is reported as low value.

Credit Risk Insurer

		No Quality Review	Quality Review
Mortgage Lender	Cheat	$V(T - g_L^{\wedge}),$ $V \cdot g_L^{\wedge} - V$	$V(T - g_L^{\wedge}),$ $V \cdot g_L^{\wedge} - V - C$
	No Cheat	$V(T - g_H^{\wedge}),$ $V \cdot g_H^{\wedge} - V$	$V(T - g_H^{\wedge}),$ $V \cdot g_H^{\wedge} - V - C$

As the payoff function shows, the payoff for a high value H lender when he behaves honestly is always higher than when he cheats. This is because he will be charged a high guarantee fee with no compensation if he does so. Therefore, a high value lender does not have the incentive to “cheat” or underreport the collateral value regardless the outcome of the loan. It is always in a high value lender’s best interest to honestly report the property value.

A dishonest lender with a low value collateral will be penalized only under two conditions, 1) the loan must default, 2) the credit risk insurer must choose to review this defaulted loan. Since credit risk insurers will only choose to review defaulted loans, loans that are paid off would not be subject to review regardless if the lender cheated or not.

No lenders have incentive to over appraise the property value higher than \hat{H} since it is common knowledge that \hat{H} is the highest property value in the market. If a credit risk insurer sees a property value higher than \hat{H} , he would automatically infer that fraudulent behavior has occurred and hence reject the transaction.

Lenders who receive a low value property have incentive to bias the appraisal value upward and report the property as \hat{H} to receive a low guarantee fee. Let's denote a credit risk insurer's prior belief on cheating by low value lenders as e . Because the signal sent by the lender is costless and property type independent, mortgage insurer does not update his belief on cheating after receiving the reported value. The update process takes place only when the outcome of the loan is realized.

The above sections set up the game structure, payoff functions and the action space of mortgage lenders and credit risk insurers. In the following section, I shall solve for the perfect Bayesian hybrid and pure strategy equilibria and provide comparative static analyses of the impact of parameters on each equilibrium outcome.

2.4.4 Hybrid Equilibrium

There exist hybrid and pure strategy equilibria in this game. In this section I shall focus on the hybrid equilibrium where one type of lender chooses to send a signal with certainty and the other type of lender randomizes between mimicking the first type by

choosing the first type's signal and separating from the first type by choosing a different signal. By model setup, only lenders with low value properties have incentive to mimic the ones with high value properties. We hence analyze the case in which the low value lender randomizes.

Suppose the high value lender chooses to honestly report the property value, but the low value lender randomizes between being honest with probability $1 - e$ and dishonest with probability e .

The probabilities for certain events according to Bayes's rules are as follows.

Using equations (a)-(f), the probability of observing default outcome on a reported high value property is:

$$P(\hat{H}, D) = P(\hat{H}, D | L) \cdot P(L) + P(\hat{H}, D | H) \cdot P(H) = p_L e(1 - \eta) + p_H \eta \quad (1)$$

The probability of observing default outcome on a reported low value property is:

$$P(\hat{L}, D) = P(\hat{L}, D | L) \cdot P(L) + P(\hat{L}, D | H) \cdot P(H) = p_L(1 - e)(1 - \eta) \quad (2)$$

The probability of observing payoff outcome on a reported high value property is:

$$P(\hat{H}, ND) = P(\hat{H}, ND | L) \cdot P(L) + P(\hat{H}, ND | H) \cdot P(H) = (1 - p_L)e(1 - \eta) + (1 - p_H)\eta \quad (3)$$

The probability of observing payoff outcome on a reported low value property is:

$$P(\hat{L}, ND) = P(\hat{L}, ND | L) \cdot P(L) + P(\hat{L}, ND | H) \cdot P(H) = (1 - p_L)(1 - e)(1 - \eta) \quad (4)$$

The probability of observing default outcome on a reported high, but in fact low value property is:

$$P(\hat{H}, D, L) = P(\hat{H}, D | L)P(L) = p_L e(1 - \eta) \quad (5)$$

The probability of observing default outcome on a reported high and indeed high value property is:

$$P(\hat{H}, D, H) = P(\hat{H}, D | H)P(H) = p_H \eta \quad (6)$$

The probability of observing default outcome on a reported low, but in fact high value property is:

$$P(\hat{L}, D, H) = P(\hat{L}, D | H)P(H) = 0 \quad (7)$$

The probability of observing default outcome on a reported low and indeed low value property is:

$$P(\hat{L}, D, L) = P(\hat{L}, D | L)P(L) = p_L(1 - e)(1 - \eta) \quad (8)$$

The probability of observing payoff outcome on a reported high, but in fact low value property is:

$$P(\hat{H}, ND, L) = P(\hat{H}, ND | L)P(L) = (1 - p_L)e(1 - \eta) \quad (9)$$

The probability of observing payoff outcome on a reported high and indeed high value property is:

$$P(\hat{H}, ND, H) = P(\hat{H}, ND | H)P(H) = (1 - p_H)\eta \quad (10)$$

The probability of observing payoff outcome on a reported low, but in fact high value property is:

$$P(\hat{L}, ND, H) = P(\hat{L}, ND | H)P(H) = 0 \quad (11)$$

Lastly, the probability of observing payoff outcome on a reported low and indeed low value property is:

$$P(\hat{L}, ND, L) = P(\hat{L}, ND | L)P(L) = (1 - p_L)(1 - e)(1 - \eta) \quad (12)$$

Derived based on the above probabilities, the followings are the beliefs credit risk insurers have on the true value of the property upon observing the signal sent by the lender, i.e., the appraisal value (\hat{H}, \hat{L}) , as well as the outcome of the loans (D, ND) . A credit risk insurer will choose the probability of review based on his belief, a best response function with respect to lender's strategy.

A credit risk insurer's belief that a property has low value given that it is reported as high value and has defaulted is:

$$P(L | \hat{H}, D) = \frac{P(\hat{H}, D, L)}{P(\hat{H}, D)} = \frac{p_L e(1-\eta)}{p_H \eta + p_L e(1-\eta)} \quad (13)$$

His belief that a property has high value given that it is reported as high value and has defaulted is:

$$P(H | \hat{H}, D) = \frac{P(\hat{H}, D, H)}{P(\hat{H}, D)} = \frac{p_H \eta}{p_H \eta + p_L e(1-\eta)} \quad (14)$$

His belief that a property has low value given that it is reported as low value and has defaulted is:

$$P(L | \hat{L}, D) = \frac{P(\hat{L}, D, L)}{P(\hat{L}, D)} = \frac{p_L(1-e)(1-\eta)}{p_L(1-e)(1-\eta)} = 1 \quad (15)$$

His belief that a property has high value given that it is reported as low value and has defaulted is:

$$P(H | \hat{L}, D) = \frac{P(\hat{L}, D, H)}{P(\hat{L}, D)} = 0 \quad (16)$$

His belief that a property has low value given that it is reported as high value and has paid off is:

$$P(L | \hat{H}, ND) = \frac{P(\hat{H}, ND, L)}{P(\hat{H}, ND)} = \frac{(1 - p_L)e(1 - \eta)}{(1 - p_H)\eta + (1 - p_L)e(1 - \eta)} \quad (17)$$

His belief that a property has high value given that it is reported as high value and has paid off is:

$$P(H | \hat{H}, ND) = \frac{P(\hat{H}, ND, H)}{P(\hat{H}, ND)} = \frac{(1 - p_H)\eta}{(1 - p_H)\eta + (1 - p_L)e(1 - \eta)} \quad (18)$$

His belief that a property has low value given that it is reported as low value and has paid off is:

$$P(L | \hat{L}, ND) = \frac{P(\hat{L}, ND, L)}{P(\hat{L}, ND)} = \frac{(1 - p_L)(1 - e)(1 - \eta)}{(1 - p_L)(1 - e)(1 - \eta)} = 1 \quad (19)$$

Lastly, his belief that a property has high value given that it is reported as low value and has paid off is:

$$P(H | \hat{L}, ND) = \frac{P(\hat{L}, ND, H)}{P(\hat{L}, ND)} = 0 \quad (20)$$

Before we solve for the hybrid equilibrium solution, we first look at how the beliefs change with respect to changes on the parameters.

$$\frac{\partial P(H | \hat{H}, D)}{\partial e} = -\frac{p_H p_L \eta (1 - \eta)}{[p_H \eta + p_L e(1 - \eta)]^2} < 0 \quad (21)$$

$$\frac{\partial P(H | \hat{H}, D)}{\partial p_H} = \frac{\eta [p_H \eta + p_L e(1 - \eta)] - p_H \eta^2}{[p_H \eta + p_L e(1 - \eta)]^2} = \frac{\eta p_L e(1 - \eta)}{[p_H \eta + p_L e(1 - \eta)]^2} > 0 \quad (22)$$

$$\frac{\partial P(H | \hat{H}, D)}{\partial p_L} = -\frac{e(1 - \eta)}{[p_H \eta + p_L e(1 - \eta)]^2} < 0 \quad (23)$$

$$\frac{\partial P(H | \hat{H}, D)}{\partial \eta} = \frac{p_H[p_H\eta + p_L e(1-\eta)] - p_H\eta(p_H - p_L e)}{[p_H\eta + p_L e(1-\eta)]^2} = \frac{p_H p_L e}{[p_H\eta + p_L e(1-\eta)]^2} > 0 \quad (24)$$

The above comparative static shows the directions of change on beliefs with respect to changes on four parameters. First of all, a credit risk insurer's belief that a property is high value given that its stated value is high and the loan has defaulted decreases if lenders are more likely to cheat. Cetera paribus the higher the probability of cheating by lenders, the lower the probability that a loan is truly a high value loan conditional on event of default. Secondly, the belief that a property is high value given that its stated value is high and the loan has defaulted increases if the high value loan's default probability rises or the low value loan's default probability falls or if there are more high value loans in the market. The higher default probability of high value loans, or the lower default probability of low value loans, the more likely that a defaulted yet claimed to be high value loan is truly a high value loan. If the proportion of high value loans in the market increases, the likelihood of a defaulted yet claimed to be high value loan is indeed a high value loan also increases.

Now I shall solve for the hybrid equilibrium solution. In order for a hybrid equilibrium to exist, it must be the case that the two players are indifferent between strategies. I.e., the payoffs from cheating have to equal the payoffs from being honest for a low value lender. Similarly, the payoffs from review have to equal the payoffs from not review for a credit risk insurer.

The expected payoffs for a low value lender when he cheats is:

$$E(\pi) | cheating = p_L q [V(T - g_{\hat{H}}) - V] + p_L (1 - q) V(T - g_{\hat{H}}) + (1 - p_L) V(T - g_{\hat{H}}) \quad (25)$$

Three outcomes could happen when a low value lender chooses to cheat. The loan could pay off safely with probability $1 - p_L$; or it could default but not chosen for reviews with probability $p_L (1 - q)$; or it could default and be chosen for reviews with probability $p_L q$. The expected payoff from cheating thus equals the probability of cheating, default and reviewed multiplies the payoffs from being caught (depicted in the 1st part), plus the probability of cheating, default, not reviewed multiplies the payoffs from not being caught (depicted in the 2nd part), plus the probability of cheating, not default multiplies the payoff from not being caught (depicted in the 3rd part).

The expected payoff for a low value lender when he does not cheat is simply his payoff from being honest

$$E(\pi) | honest = V(T - g_{\hat{L}}) \quad (26)$$

Equate the above two expected payoffs allows us to solve for the equilibrium probability of review that makes a low value lender indifferent between cheating and being honest.

$$\begin{aligned} p_L q [V(T - g_{\hat{H}}) - V] + p_L (1 - q) V(T - g_{\hat{H}}) + (1 - p_L) V(T - g_{\hat{H}}) &= V(T - g_{\hat{L}}) \\ \Rightarrow -q p_L &= (T - g_{\hat{L}}) - (R - g_{\hat{H}}) \\ \Rightarrow q^* &= \frac{g_{\hat{L}} - g_{\hat{H}}}{p_L} \end{aligned} \quad (27)$$

Now let's turn to the insurer's payoff. Given the insurer's belief that high value lenders will not misreport loan values, all reported low value loans must be true low value loans. Therefore, the insurer only needs to review defaulted loans with high reported values \hat{H} .

Two outcomes arise when review is conducted on a defaulted \hat{H} loan. The loan is either a true high value loan H or a true low value loan L . The credit risk insurer either transfers the default loss to the lender if the loan is found to be low value, or incurs the default loss if it is found to be high value. Hence, the expected payoff for an insurer when he reviews a defaulted loan with high reported value is the probability of discovering the true type of the property (H or L), given observing a defaulted \hat{H} multiplies the payoffs from discovering the true type. His expected payoff is as follows when he reviews a defaulted \hat{H} loan:

$$\begin{aligned}
 E(\pi) | review &= \frac{p_L e(1-\eta)}{p_H \eta + p_L e(1-\eta)} (V \cdot g_{\hat{H}} - C) + \frac{p_H \eta}{p_H \eta + p_L e(1-\eta)} (V \cdot g_{\hat{H}} - C - V) \\
 &= \frac{p_L e(1-\eta)(V \cdot g_{\hat{H}} - C) + p_H \eta (V \cdot g_{\hat{H}} - C - V)}{p_H \eta + p_L e(1-\eta)} \quad (28)
 \end{aligned}$$

The 1st probability is simply the probability that a loan has low value given that a high reported value and event of default are observed on that loan, i.e. $P(L | \hat{H}, D)$. The second term is the probability that a loan has high value given that a high reported value and default occurred on that loan, i.e. $P(H | \hat{H}, D)$.

The credit risk insurer's expected payoff if he does not review a defaulted high reported value loan is simply the revenue from insuring a \hat{H} minus the cost from default.

$$E(\pi) | \wedge \text{review} = V \cdot g_{\hat{H}} - V \quad (29)$$

Equate the two expected payoffs solves for the equilibrium probability of cheating that makes credit risk insurers indifferent between reviewing and not reviewing a defaulted loan with a high reported value.

$$\begin{aligned} \frac{p_L e(1-\eta)(V \cdot g_{\hat{H}} - C) + p_H \eta(V \cdot g_{\hat{H}} - C - V)}{p_H \eta + p_L e(1-\eta)} &= V \cdot g_{\hat{H}} - V \\ \Rightarrow e\{p_L(1-\eta)(V - C)\} &= p_H \eta \cdot C \\ \Rightarrow e^* &= \frac{p_H \eta \cdot C}{p_L(1-\eta)(V - C)} \end{aligned} \quad (30)$$

$$q^* = \frac{g_L^{\wedge} - g_H^{\wedge}}{p_L} \text{ and } e^* = \frac{p_H \eta \cdot C}{p_L(1-\eta)(V - C)} \text{ establish the equilibrium probability of review}$$

and cheating that make the mortgage lender and the credit risk insurer indifferent between cheating and not cheating and review and not review.

If $q > q^*$ then a mortgage lender will always remain honest and report the true property value with probability one.

If $q = q^*$ then a mortgage lender will randomizes between being honest and dishonest.

If $q < q^*$ then a mortgage lender will always cheat and misreport low value property as high value with probability one.

If $e > e^*$ then a credit risk insurer will always review defaulted loans with probability one.

If $e = e^*$ then a credit risk insurer will randomize between reviewing a defaulted loan and not reviewing it.

If $e < e^*$ then a credit risk insurer will always choose not to review any defaulted loans with high reported values with probability one.

Above result shows the hybrid equilibrium conditions. The following comparative static shows the relationship between the equilibrium probability of review, equilibrium probability of cheating and the parameters.

$$\frac{\partial e^*}{\partial p_H} = \frac{\eta C}{p_L(1-\eta)(V-C)} > 0 \quad (31)$$

$$\frac{\partial e^*}{\partial p_L} = \frac{\eta C}{(1-\eta)(V-C)} \cdot \frac{-1}{p_L^2} < 0 \quad (32)$$

$$\frac{\partial e^*}{\partial C} = \frac{p_H \eta p_L (1-\eta)(V-C) + p_H \eta C p_L (1-\eta)}{[p_L(1-\eta)(V-C)]^2} = \frac{p_H \eta p_L (1-\eta)V}{[p_L(1-\eta)(V-C)]^2} > 0 \quad (33)$$

$$\frac{\partial e^*}{\partial \eta} = \frac{p_H C p_L (1-\eta)(V-C) + p_H \eta C p_L (V-C)}{[p_L(1-\eta)(V-C)]^2} = \frac{p_H C p_L (V-C)}{[p_L(1-\eta)(V-C)]^2} > 0 \quad (34)$$

The comparative static result shows how the equilibrium condition on the cheating probability changes with regard to changes on parameters. First of all, as the high value loan's default probability increases, the equilibrium probability of cheating that makes a credit insurer indifferent between review and not review also increases. This means that as more and more high value loans default, reviews will discover less and less frauds

given the previous equilibrium cheating probability. Hence a higher cheating probability by lender is needed in order for an insurer to randomize between reviewing and not to review. Secondly, as probability of default increases for low value loans, cheatings are more likely to be detected given the same probability of review. Hence, a lower probability of cheating by lender is required to sustain indifference between review and not for an insurer. Thirdly, the equilibrium cheating probability also increases with cost of review and proportion of high value loans in the market. As cost of review for insurers rises or proportion of high value loans increases, reviews become more costly or less likely to generate fraud detections. As a result, the probability of cheating by lenders must rise in order for insurers to randomize between review and not review.

Similarly, the following comparative static shows the relationship between the equilibrium probability of review with each parameter.

$$\frac{\partial q^*}{\partial p_L} = (g_L^{\wedge} - g_H^{\wedge}) \cdot \frac{-1}{p_L^2} < 0 \quad (35)$$

$$\frac{\partial q^*}{\partial g_L^{\wedge}} = \frac{1}{p_L} > 0 \quad (36)$$

$$\frac{\partial q^*}{\partial g_H^{\wedge}} = \frac{-1}{p_L} < 0 \quad (37)$$

The result shows that the equilibrium probability of review that makes a lender indifferent between cheating and not cheating decreases if default probability of low value loans rises. It increases with guarantee fees for low value loans and decreases with guarantee fees for high value loans. It is not a direct function of default probability of

high value loans. The intuition is the following: as the default probability of low value loans rises, it becomes easier for credit risk insurers to detect fraud given the same level of cheating. Therefore, a lower probability of review is required to induce randomization between cheating and not cheating by lenders. An increased guarantee fee for low value loans or a decreased guarantee fee for high value loans generate higher incentive to cheat by lenders and hence calls for a higher probability of review to sustain the hybrid equilibrium.

2.4.5 Pure Strategy Equilibria

The above section completes the analysis on the hybrid equilibrium condition of the game. In this section, I shall take a look at the other four possible pure-strategy perfect Bayesian equilibria in this game. They are: pooling on reporting high value \hat{H} ; pooling on reporting low value \hat{L} ; separation with high value lender reporting high value \hat{H} and low value lender reporting low value \hat{L} ; and separation with high value lender reporting low value \hat{L} and low value lender reporting high value \hat{H} .

Because guarantee fee is exogenously determined, a reported \hat{H} is always charged less than an \hat{L} is. Furthermore, review outcome does not influence true H value loans.

Therefore, a true H value lender will never have the incentive to report the property value as \hat{L} . Hence pooling equilibrium on \hat{L} or separating equilibrium with H reporting as \hat{L} and L reporting as \hat{H} does not exist.

2.4.5.1 Pooling Equilibrium on \hat{H}

Now let's investigate the other pooling equilibrium where both types of lenders report the property value as \hat{H} regardless of its true value. Suppose there exists such equilibrium where the lender's strategy is to report the property value as high regardless of its true value, i.e., $e = 1$. The insurer's information set corresponding to the lender's strategy should be on the equilibrium path and his belief shall be determined by Bayes' rule and the lender's strategy.

The insurer's belief that a property is truly high value given that it is reported to be high is:

$$P(H | \hat{H}) = \frac{P(\hat{H} | H)P(H)}{P(\hat{H} | H)P(H) + P(\hat{H} | L)P(L)} = \frac{1 \cdot \eta}{1 \cdot \eta + 1 \cdot (1 - \eta)} = \eta \quad (38)$$

Similarly, his belief that a property has a low value given that its value is reported to be high is just $1 - \eta$.

After the loan's outcome is realized, an insurer's belief that a property is indeed high value given that a reported high value and default are observed on the loan is:

$$P(H | \hat{H}, D) = \frac{P(\hat{H}, D, H)}{P(\hat{H}, D)} = \frac{p_H \eta}{p_H \eta + p_L (1 - \eta)} \quad (39)$$

His belief that a property has low value given that it is reported to be high value and event of default has occurred is:

$$P(L | \hat{H}, D) = \frac{P(\hat{H}, D, L)}{P(\hat{H}, D)} = \frac{p_L(1-\eta)}{p_H\eta + p_L(1-\eta)} \quad (40)$$

His belief that a property has high value given that it is reported to be high value and the loan has paid off is:

$$P(H | \hat{H}, ND) = \frac{P(\hat{H}, ND, H)}{P(\hat{H}, ND)} = \frac{(1-p_H)\eta}{(1-p_H)\eta + (1-p_L)(1-\eta)} \quad (41)$$

And lastly, his belief that a property has low value given that it is reported to be high value and the loan has paid off is:

$$P(L | \hat{H}, ND) = \frac{P(\hat{H}, ND, L)}{P(\hat{H}, ND)} = \frac{(1-p_L)(1-\eta)}{(1-p_H)\eta + (1-p_L)(1-\eta)} \quad (42)$$

Given the insurer's belief, his expected payoff if he chooses to review a defaulted loan is the expected payoff from reviewing a fraudulent loan and reviewing an honest loan:

$$\begin{aligned} E(\pi) | review &= \frac{p_L(1-\eta)}{p_H\eta + p_L(1-\eta)} (V \cdot g_{\hat{H}} - C) + \frac{p_H\eta}{p_H\eta + p_L(1-\eta)} (V \cdot g_{\hat{H}} - C - V) \\ &= \frac{p_L(1-\eta)(V \cdot g_{\hat{H}} - C) + p_H\eta(V \cdot g_{\hat{H}} - C - V)}{p_H\eta + p_L(1-\eta)} \end{aligned} \quad (43)$$

If he chooses not to review the defaulted loan, his expected payoff is simply:

$$E(\pi) | \wedge review = V \cdot g_{\hat{H}} - V \quad (44)$$

Compare the expected payoffs of review and not review we obtain that

$$\begin{aligned}
& E(\pi) | review - E(\pi) | \hat{review} \\
&= \frac{p_L(1-\eta)(V \cdot g_{\hat{H}} - C) + p_H\eta(V \cdot g_{\hat{H}} - C - V)}{p_H\eta + p_L(1-\eta)} - (V \cdot g_{\hat{H}} - V) \\
&= \frac{p_L(1-\eta)(V - C) - p_H\eta \cdot C}{p_H\eta + p_L(1-\eta)} \tag{45} \\
&= \frac{p_L(1-\eta)V}{p_H\eta + p_L(1-\eta)} - C
\end{aligned}$$

Therefore, an insurer's best response following lender's pooling strategy on \hat{H} depends on the sign of $\frac{p_L(1-\eta)V}{p_H\eta + p_L(1-\eta)} - C$, i.e., the relative size of review cost and average loss

from default in the market. If $\frac{p_L(1-\eta)V}{p_H\eta + p_L(1-\eta)} - C > 0$ or default loss is greater than the

cost of review, then an insurer's best response to a pooling strategy on \hat{H} is to review with probability one, i.e. $q = 1$. Low value and high value lenders' expected payoffs under pooling on \hat{H} are $V(T - g_{\hat{H}}) - p_L \cdot V$ and $V(T - g_{\hat{H}})$ respectively. An insurer

earns $E(\pi) | review = \frac{p_L(1-\eta)(V \cdot g_H - C) + p_H\eta(V \cdot g_H - C - V)}{p_H\eta + p_L(1-\eta)}$, greater than his

payoff when he does not review.

To determine whether both types of lenders are willing to pool on \hat{H} , we need to specify how the insurer would react to \hat{L} . If the credit insurer's best response to both \hat{L} and default D is not to review, then a high value lender's payoff regardless of the outcome of the loan remains at $V(T - g_{\hat{H}})$ if he reports \hat{H} , which always exceeds his expected

payoff $V(T - g_{\hat{L}})$ regardless of the loan's outcome if he chooses to report \hat{L} . Under the same best response of not to review if \hat{L} and D are observed, a low value lender's payoff from truthfully reporting the property value is $V(T - g_{\hat{L}})$ regardless of loan's outcome. If the credit insurer's best response to \hat{H} and D is to always review, the mortgage lender's expected payoff from misreporting the property value to be \hat{H} is then $V(T - g_{\hat{H}}) - p_L \cdot V$. Comparing the two expected payoffs, if a low value lender's payoff from reporting \hat{L} exceeds his expected payoff from reporting \hat{H} , i.e., $V(T - g_{\hat{L}}) - p_L \cdot V > V(T - g_{\hat{H}}) - p_L \cdot V$ then no pooling equilibrium on \hat{H} where both types of lenders pool on \hat{H} and the credit insurer reviews all defaulted loans with certainty shall exist when insurer's best response to \hat{L} and D is not to review. Only if

$$g_{\hat{L}} - g_{\hat{H}} - p_L > 0 \text{ and } \frac{p_L(1-\eta)V}{p_H\eta + p_L(1-\eta)} - C > 0 \text{ does such pooling equilibrium on } \hat{H}$$

exist when insurer's best response to \hat{L} and D is not to review.

It remains to consider the insurer's belief at the information set corresponding to \hat{L} and D as well as the optimality of his choice given this belief. The insurer's belief that a loan has low value when observing \hat{L} and D is:

$$P(L | \hat{L}, D) = \frac{P(\hat{L}, D, L)}{P(\hat{L}, D)} = \frac{p_L(1-e)(1-\eta)}{p_L(1-e)(1-\eta)} = 1 \quad (46)$$

His belief that a loan has high value when observing \hat{L} and D is:

$$P(H | \hat{L}, D) = \frac{P(\hat{L}, D, H)}{P(\hat{L}, D)} = 0 \quad (47)$$

His expected payoffs under review and no review conditional on observing \hat{L} and D are

$$E(\pi | \text{review}, \hat{L}, D) = V \cdot g_{\hat{L}} - C - V \quad (48)$$

$$E(\pi | \wedge \text{review}, \hat{L}, D) = V \cdot g_{\hat{L}} - V \quad (49)$$

Since review always yields fewer payoffs than no review when \hat{L} and D are observed, the insurer's optimal strategy to \hat{L} and D will always be not to review. Therefore, a pooling equilibrium on \hat{H} exists under the conditions that $g_{\hat{L}} - g_{\hat{H}} - p_L > 0$ and

$$\frac{p_L(1-\eta)V}{p_H\eta + p_L(1-\eta)} - C > 0. \text{ The insurer's best response under the above parameters is to}$$

review if he sees \hat{H} and D , not to review if he sees \hat{L} and D . Low value and high value

lenders each earns $V(T - g_H) - p_L \cdot V$ and $V(T - g_H)$ respectively. The credit risk

$$\text{insurer earns } E(\pi) | \text{review} = \frac{p_L(1-\eta)(V \cdot g_H - C) + p_H\eta(V \cdot g_H - C - V)}{p_H\eta + p_L(1-\eta)}.$$

If $\frac{p_L(1-\eta)V}{p_H\eta + p_L(1-\eta)} - C < 0$ or review cost is greater than the loss from default, an

insurer's best response to a pooling strategy on \hat{H} is not to review with certainty, i.e.,

$q = 0$. Low value and high value lenders' expected payoffs under pooling on \hat{H} are both

$V(T - g_{\hat{H}})$. An insurer earns $E(\pi) | \wedge \text{review} = V \cdot g_{\hat{H}} - V$. Since an insurer's best

response toward \hat{L} and D is also not to review, both types of lenders earn $V(T - g_{\hat{H}})$

regardless of default outcome when they report \hat{H} , which always exceed their payoffs $V(T - g_{\hat{L}})$ when reporting \hat{L} . Therefore, pooling equilibrium on \hat{H} also exists when

$$\frac{p_L(1-\eta)V}{p_H\eta + p_L(1-\eta)} - C < 0. \text{ The credit insurer's best response is not to review any}$$

defaulted loans. Both types of lenders earn $V(T - g_{\hat{H}})$, the insurer earns

$$E(\pi) |^{\wedge} \text{review} = V \cdot g_{\hat{H}} - V.$$

In summary, the existence of pooling equilibrium where both types of lenders report the property values as high regardless of its true value depends on parameters. Under the

conditions that $g_{\hat{L}} - g_{\hat{H}} - p_L > 0$ and $\frac{p_L(1-\eta)V}{p_H\eta + p_L(1-\eta)} - C > 0$, the credit risk insurer's

best response to lenders pooling on \hat{H} is to review any defaulted loans with high reported values and not to review any defaulted loans with low reported values. There

also exists another pooling equilibrium where both types of lenders pool on \hat{H} . Under

the conditions that $g_{\hat{L}} - g_{\hat{H}} - p_L > 0$ and $\frac{p_L(1-\eta)V}{p_H\eta + p_L(1-\eta)} - C < 0$, the credit risk

insurer's best response to lenders pooling on \hat{H} is always not to review any defaulted

loans. Those parameters and equilibria constructed correspond to the optimal hybrid

equilibrium e^* and q^* established previously. For example, when $g_{\hat{L}} - g_{\hat{H}} > p_L$, it is

concluded that the optimal review probability for a hybrid equilibrium to exist is greater

than one, i.e. $q^* > 1$. Therefore, $q < q^*$ and lender will always cheat. When

$\frac{p_L(1-\eta)V}{p_H\eta + p_L(1-\eta)} - C < 0$, the optimal cheating probability for a hybrid equilibrium to

exist is greater than one, i.e., $e^* > 1$. Therefore, $e < e^*$, and the credit risk insurer will always choose not to review.

2.4.5.2 Separating Equilibrium Where H Reports \hat{H} and L Reports \hat{L}

Now let's take a look at the feasible separating equilibrium where both types of lenders honestly report the property value, i.e. $e = 0$. Under this scenario, the credit insurer's

beliefs on the true value of a property when he sees \hat{H} and \hat{L} must be

$$P(H | \hat{H}) = \frac{P(\hat{H} | H)P(H)}{P(\hat{H} | H)P(H) + P(\hat{H} | L)P(L)} = \frac{1 \cdot \eta}{1 \cdot \eta + 0 \cdot (1 - \eta)} = 1 \quad (50)$$

$$P(L | \hat{L}) = \frac{P(\hat{L} | L)P(L)}{P(\hat{L} | H)P(H) + P(\hat{L} | L)P(L)} = \frac{1 \cdot (1 - \eta)}{0 \cdot \eta + 1 \cdot (1 - \eta)} = 1 \quad (51)$$

The credit risk insurer's beliefs on the property's true value H or L after observing the

lender's signal \hat{H} or \hat{L} , and the loan's outcome D or ND are:

The probability that a property has high value given that it is reported to be high value

and the loan has defaulted is $P(H | \hat{H}, D) = \frac{P(\hat{H}, D, H)}{P(\hat{H}, D)} = 1$; the probability that a

property has low value given that it is reported to be high value and the loan has

defaulted is $P(L | \hat{H}, D) = \frac{P(\hat{H}, D, L)}{P(\hat{H}, D)} = 0$; the probability that a property is high value

given that it is reported as high value and the loan has paid off is

$$P(H | \hat{H}, ND) = \frac{P(\hat{H}, ND, H)}{P(\hat{H}, ND)} = 1; \text{ the probability that a property has low value given}$$

that it is reported as high value and the loan has paid off is

$$P(L | \hat{H}, ND) = \frac{P(\hat{H}, ND, L)}{P(\hat{H}, ND)} = 0.$$

His expected payoff from reviewing the defaulted loan given his beliefs is

$$E(\pi) | \text{review} = 0 \cdot (V \cdot g_{\hat{H}} - C) + 1 \cdot (V \cdot g_{\hat{H}} - C - V) = V \cdot g_{\hat{H}} - C - V$$

His expected payoff from not to review is $E(\pi) | \wedge \text{review} = V \cdot g_{\hat{H}} - V$. An insurer will

always chooses not to review when both types of lenders honestly report the property value as long as review cost $C > 0$. Low value lender and high value lender each earns

$V(T - g_{\hat{L}})$ and $V(T - g_{\hat{H}})$ respectively. It remains to check if the lender's strategy is

optimal given the insurer's strategy. If the low value lender deviates by reporting \hat{H} , the lender earns $V(T - g_{\hat{H}})$ under the no review strategy by the credit risk insure. This

payoff exceeds the low value lender's payoff of $V(T - g_{\hat{L}})$ from reporting honestly. Thus

no separating Bayesian equilibrium exists when the review cost $C > 0$. When the review cost $C = 0$, the insurer will always choose to review any defaulted loans. The low value lender and the high value lender's expected payoffs are still $V(T - g_{\hat{L}})$ and $V(T - g_{\hat{H}})$

respectively. If the low value lender deviates, his expected payoff will be

$V(T - g_{\hat{H}}) - p_L \cdot V$. Deviation makes sense only if $g_{\hat{L}} - g_{\hat{H}} - p_L > 0$. Therefore, a

separating equilibrium exists when 1) review is costless, i.e. $C = 0$ and 2) the low value

property's default probability is greater than the difference between the two guarantee fees, $g_L - g_H - p_L < 0$, i.e., the cost of cheating outweighs the benefit from cheating.

In summary, the existence of a separating equilibrium where both types of lenders honestly report the property value also depends on the parameters. It exists only when the review is costless and $g_L - g_H - p_L < 0$.

Under the situation where guarantee fees are exogenously determined, the existence of Bayesian equilibria in this game depends on parameters. There are cases where review cost is so high, compared with the default loss, that the credit risk insurer chooses not to review defaulted loans and incurs the credit loss generated by low value lender's fraudulent behaviors. There are also cases where the probability of default for low value properties' is low enough, compared with the difference in guarantee fees, that low value lenders are encouraged to cheat. Hybrid equilibrium where lenders randomize between cheating and not cheating and insurers randomize between review and not review also exists under certain parameters conditions.

2.5 Conclusion

This paper analyzes a phenomenon in mortgage market induced by the information asymmetric structure between mortgage lenders and credit risk insurers. This phenomenon is referred to as appraisal bias, where mortgage lenders under information asymmetry, have incentive to behave in ways that maximize their utilities at the expense of credit risk insurers, in particular GSE's.

The game played between mortgage lenders and mortgage insurers can be viewed as a cheap talk, for the signals sent by lenders are costless, non binding, and non-verifiable. Consequently, a babbling equilibrium where all signals are uninformative exists. However, by changing the non-verifiably information to be semi-verifiable, this paper establishes how semi-verifiability in a cheap talk game structure helps construct a non-babbling perfect Bayesian equilibrium in an asymmetric information environment.

This paper assumes that the guarantee fee, which determines the profit of credit risk insurers are exogenous. In reality, this fee ought to be determined endogenously. In subsequent studies, I shall look at cases where mortgage insurers endogenously determine the guarantee fees based on expected default rate.

Chapter 3: An Empirical Analysis of Mortgage Termination With Appraisal Bias – Maximum Likelihood Estimation of A Proportional Hazard Competing Risks Model With Grouped Duration Data

3.1 Introduction

Residential mortgage market has received a great deal of attention in the past decades due to its importance to the economy. First of all, residential mortgage market stimulates the economy by generating jobs, wages and tax revenue through construction of new homes. According to National Association of Home Builders, the construction of 1,000 single-family homes can generate on average 2,448 full-time jobs in construction and construction related industries; it can also generate \$79.4 million in wages and \$42.5 million in combined federal, state and local revenues and fees (National Association of Home Builders – Economic and Housing Data). Secondly, residential mortgage market strengthens social stability by increasing and expanding homeownerships across the nation. United States' social stability is partially reflected by having the highest homeownership of any major economy in the world. Currently, U.S. homeownership has topped 69.0 percent as of 2004Q3 (U.S. Census Bureau – Housing Vacancy Survey Third Quarter 2004). Lastly, residential mortgage market is crucial to the prosperity of U.S. economy. Residential fixed investment and housing services attributed to 14 percent of U.S. total GDP in 2003 (Bureau of Economic Analysis: Latest NIPA Table for GDP). In dollar terms, residential mortgage debt outstanding (MDO) in the U. S. totaled \$7.8 trillion as of year-end 2003, higher than United States government total debt outstanding today. This amount was more than double the amount of residential mortgage debt outstanding 10 years ago, or \$3.4 trillion as of year-end 1993. By 2013, mortgage debt is estimated to total \$17.2 trillion with an 8.25 percent annual increase (Federal Reserves

Bulletin, Freddie Mac Special Commentary, Fannie Mae Annual Report). The growing size of residential mortgage market and its importance in maintaining a stable society generates great interest in understanding the economics of mortgages.

Understanding residential mortgage market requires the knowledge of borrower's default and prepayment actions. Default is defined as "failure to make required debt payments on a timely basis or to comply with other conditions of an obligation or agreement."

Prepayment is defined as "payment of all or part of a debt prior to its due date"

(Merriam-Webster). In mortgage terminology, default is realized when borrowers fail to make a certain number of mortgage payments consecutively. It results in the "transfer of the legal ownership of the property from the borrower to the lender either through the execution of foreclosure proceedings or the acceptance of a deed in lieu of foreclosure"

(Giliberto and Houston 1989). Prepayment is realized when borrowers payoff the current mortgage obligation either through sale of the property or refinancing the current mortgage with a new mortgage. Prepayment stops scheduled cash flow of principal and interest payments from the borrowers to the mortgage lenders. Mortgage market considers default and prepayment as two options available to borrowers to terminate mortgage contracts before the amortization date.

Evaluating the value of these two options is crucial to the pricing decisions of mortgage contracts. Pricing of mortgage contracts requires thorough understanding of borrowers' default and prepayment behaviors to take the termination risks into consideration. Pricing mortgage contracts is also complicated because default and prepayment options are

distinct yet interdependent. For example, once a borrower decides to default or prepay on a mortgage, he or she can no longer exercise default or prepayment options in the future. Furthermore, differences in risk preferences and other idiosyncrasy across borrowers may influence borrowers behaviors differently. All of the above imply that adequate pricing of mortgages calls for a thorough understanding of the mortgage borrower behavior and requires appropriate models of default and prepayment risks.

It is well established that the contingent claims model developed by Black and Scholes (1973), Merton (1973), Cox, Ingersoll, and Ross (1985) provides a coherent and useful framework for analyzing mortgage borrowers' behaviors. In their framework, prepayment is viewed as a call option and default as a put option. Well-informed borrowers in a perfect competitive market are assumed to exercise either of these two options whenever they can to increase their wealth. Borrowers in the absence of transaction cost can increase their wealth by defaulting on a mortgage contract, purchase the same property with a lower monthly payment for the same remaining term when the market value of the mortgage exceeds the value of the house. They can increase their wealth by prepaying and refinancing the mortgage for the same remaining term at a mortgage rate less than the coupon rate when the market value of the mortgage exceeds par value (Kau and Keenan 1995). Alternatively, one can think of borrowers take on prepayment or default actions to minimize the cost.

Although researchers quickly recognize the applicability of contingent claims option models on mortgage termination, virtually all the early studies that applied option models

to mortgage termination focus on either default or prepayment option, but not both. It is not until recent time that joint estimation of the prepayment and default options in mortgage studies becomes well known (Deng, Quigley and Van Order 2000). Because a homeowner who exercises a default option today gives up not only the option to default tomorrow, but also the option to prepay tomorrow, the two options are interdependent and need to be jointly estimated (See Kau, et al., 1995 for a theoretical outline of the relationship between the two options).

As explained above, the joint estimation of the two options calls for competing risks models. Since default and prepayment can be viewed as hazards, duration models provide a convenient analytical tool to analyze the two terminations. Duration models are well established in biometrics and labor economics with seminal works by Cox (1972), Prentice and Gloeckler (1978), Lancaster (1979), Kennan (1985), Katz (1986), Kiefer (1988), etc. It was first applied to borrower behavior in the mortgage market in 1986 (Green and Shoven, 1986) and has received extended attention following recent works by Deng, Quigley and Van Order (2000).

Deng, Quigley and Van Order (2000) contribute to the literature by making competing risks duration models useful in analyzing an overlooked aspect of mortgage market, namely the unobserved heterogeneity of borrowers. The unobserved heterogeneity of study subjects is well recognized in duration model approaches to biometric research (Kalbfleisch and Prentice, 1980). For example, research in survival time after medical treatment discovers that those patients who are least physically fit are more likely to exit

the sample of subjects than those who are fit (Kalbfleisch and Prentice, 1980). Heckman and Singer (1985) also recognize the same issue when applying duration models to study unemployment spells in labor markets.

The analogy in mortgage termination studies lies in the well-observed different prepayment rate among borrowers. Borrowers who are most financially astute are more likely to exercise prepayment options to terminate mortgage contracts, and those who are left in the sample are more likely to be financially uninformed. This implies that any sample of surviving mortgage borrowers is successively more likely to include disproportionate fractions of those less financially astute. In addition, financially astute borrowers, had they remained in the sample, are also less likely to default than the average population that remains in the sample. Therefore, without accounting for unobserved heterogeneity, one is likely to draw inaccurate inferences on the significance of explanatory variables.

However, there are still critical aspects of residential mortgage markets that previous researchers fail to capture. Specifically, previous studies on residential mortgage suffer from lack of data to account for origination appraisal bias. Origination appraisal bias arises due to the information asymmetry between mortgage investors and mortgage originators. Mortgage originators, who do not share the same objective as mortgage investors, have incentive to over appraise the value of the collateral in order to profit at the expense of the investors. Theoretical proof on the existence of appraisal bias is presented in the previous chapter. There has not been any study that investigates the

effect of origination bias on default and prepayment in residential mortgage market due to data limitation⁹. Origination bias needs to be controlled in the estimation of mortgage terminations. Without correcting for appraisal bias, one is likely to reach less accurate conclusions on default or prepayment hazards.

This paper attempts to achieve the following goals:

- Jointly estimate default and prepayment outcomes as competing risk proportional hazard models with grouped data
- Use the exact likelihood function instead of the approximated likelihood function upon which previous researchers relied
- Examine to what extent the option pricing theory can explain default and prepayment behavior with single-family residential mortgages and test if variables related to characteristics of the borrower, mortgage and underlying property are predictive to mortgage terminations
- Model the unobserved borrower heterogeneity jointly with competing-risks in Heckman-Singer semiparametric approach and compare the results with one that does not control for unobserved heterogeneity.
- Explicitly correct for origination appraisal bias at loan level by using a unique dataset that contains repeated transactions property values and compare the results with those that do not control for appraisal bias

⁹ Ciochetti *et al.* (2003) analyzed a different aggregated originator bias in commercial mortgages. Their research is based on 2043 commercial loans originated by a single lender who may have loose underwriting standards or stringent underwriting standards. They refer to this sampling bias as originator bias. To reduce it, they weight each observation in their sample based on how representative that observation type is relative to its population equivalent.

The model improves on early studies by using an exact likelihood function instead of the approximated quasi likelihood function that is commonly used. The estimation results suggest that measures of appraisal bias are significant in determining default probabilities; a high bias leads to a big hazard of default. They are, however, not significant in explaining prepayment. The results confirm previous researchers' findings that the call and put options are interdependent. They provide empirical support for the option pricing approach to explain default and prepayment behaviors of the borrowers. The results also confirm previous findings that non-option related variables, which measure characteristics of the borrower, mortgage and underlying collateral and regional economic conditions are also predictive in default and prepayment. Joint estimation of the default and prepayment risks allows the flexibility of the two risks to be interdependent. Unobserved borrower heterogeneity proves to be an important factor in accounting borrower behaviors. Ignoring the heterogeneity can lead to less optimal model fitness. Unlike previous studies, the two-by-two mass point heterogeneity distribution applied in this paper allows for independence between default heterogeneity and prepayment heterogeneity¹⁰.

The remainder of the paper is organized into the following sections. Section 2 reviews previous literature on mortgage terminations. It in itself can be broken down into early literature review, economic theory, option theory review, and reviews on econometric methodology on proportional hazard models. Section 3 discusses the dataset utilized in

¹⁰ Previous works, including Deng et. Al (2000), Huang and Ondrich (2002), restrict the heterogeneity between the two risks to be monotonic dependent. Although easy to implement, this assumption is counter intuitive to what mortgage studies have found: people who are more likely to prepay are less likely to default than those who is less likely to prepay.

this paper and provides summary statistics on important variables. Section 4 presents the exact specification of the econometric model and estimation procedure. Section 5 presents the results of the estimations and a final section summarizes the main conclusions and provides suggestions for future research.

3.2 Literature Review on Mortgage Termination

3.2.1 Early Literature Review

Early literature on mortgage termination contributes to the line of research by suggesting and testing variables that are important in explaining borrower behaviors on default and prepayment. The variables tested range from loan characteristics, borrower characteristics, property characteristics and economic conditions. Without sufficient economic theory and econometric models, early studies treat default and prepayment behaviors separately and investigate them individually.

Beginning with works by Jung (1962), Page (1964), and Von Furstenberg (1969, 1970a, 1970b), early studies on mortgage terminations attempt to explain default outcome by mortgage characteristics at origination using aggregate data. Their choices of explanatory variables include origination loan-to-value (LTV) ratio, origination interest rate, mortgage term and origination home equity. Jung (1962), despite lack of formal economic theory, is the first to identify positive correlation between LTV ratios and default risk using aggregate data from 31 savings and loans institutions. Page (1964) and Von Furstenberg (1969) are among the first to use the regression approach and provide empirical evidence on the importance of origination LTV ratio, mortgage term, and home equity at the time of origination to predict default. For example, Von Furstenberg (1969,

1970a, 1970b) finds that home equity at origination is the most important predictor of default risk, with default rates increasing by seven times if LTV ratios are raised from 90 to 97 percent. He also found that mortgage default risk increases with the age of the loan up to the loan's third or fourth year of age, after which it declines.

In addition to mortgage characteristics, early studies on mortgage termination in the form of default also investigate the importance of borrower and property characteristics. The impact of borrower and property attributes receive fewer consensus than mortgage characteristics however. Characteristics that show consistent effect on mortgage default include self-employment identification, indicators for occupation stability and variables that reflect the fluctuation of household income (Herzog and Earley, 1970; Von Furstenberg 1969). Characteristics on which researchers have not arrived at consensus include household income, borrower age, marital status, number of dependents and debt-to-income ratio at origination (Von Furstenberg 1969; Herzog and Earley, 1970; Morton 1975; Sandor and Sosin 1975; William, Beranek and Kenkel 1974).

Some of the disagreement on the effect of borrower and property characteristics on default is as follows. Von Furstenberg (1969)'s result shows that household income negatively and significantly impacts default risk. He attributes this effect to the correlation between income and LTV ratio and concludes that household income by itself is not deterministic of default. Herzog and Earley (1970) test borrower age, marital status, number of dependents and debt-to-income ratio at origination and conclude that none of the above has any significant effect on default. Although confirmed by Morton (1975)

and Sandor and Sosin (1975) separately, a different conclusion was reached if debt-to-income ratio was measured categorically (William, Beranek and Kenkel, 1974). Unlike Herzog and Earley (1970), Willam et.al find that all else equal, a borrower with debt-to-income ratio higher than 30 percent is more likely to default than other borrowers.

Property characteristics including local real estate markets, property conditions, and local unemployment rates are also among the variables that have been tested. Von Furstenberg and Green discover that mortgages in suburban areas are less risky to default than those in central cities (Von Furstenberg and Green 1976). Similarly, high default risk is also found to be associated with bad property conditions and neighborhoods with high crime rate (Williams Beranek and Kenkel 1974).

Compared to studies on default, there is less early research that focuses on prepayment behavior. Early studies on prepayment include works by Dunn and McConnell (1981), Buser and Hendershott (1984), and Titmand and Torous (1989) etc. Research on prepayment flourishes when option theory is introduced to the analysis of mortgage terminations.

In summary, early literature on mortgage termination separately analyzes variables influencing default or prepayment in lieu of formal theory. Variables analyzed include origination loan characteristic, property characteristics, borrower characteristics, and economic conditions. Researchers made no attempts to provide a theoretical basis for borrower behavior at the time of mortgage termination. However, they still made

significant contribution to the line of research by suggesting and testing variables that are candidates to determine mortgage termination.

3.2.2 Economic Theory and Option Theory Review

The first connection between economic theory and mortgage borrower behaviors are established in studies conducted in 1980s. Studies on mortgage terminations in 1980s utilize economic theory of utility maximization and test previous research results based on the assumption that borrowers rationally decide whether to carry out the mortgage contract or not in the course of maximizing their utility over time (Jackson and Kasserman 1980; Campbell and Dietrich 1983; Vandell and Thibodeau 1985; Zorn and Lea 1989; Cunningham and Capone 1990).

Based on McFadden's consumer choice model, most studies assume that mortgage borrowers have multiple choices in their choice set at each payment period during the life of the mortgage. A borrower can choose to continue payment as scheduled, stop payment and default, or prepay the mortgage. A borrower at each decision-making period chooses one of three choices that maximize her utility over time. In summary, a borrower is assumed to maximize a utility function defined over a vector of mutually exclusive qualitative choices, S , and a vector of exogenous state variables, X . Each of the utility maximizing choices can be represented as a conditional probability, all of which sum to one, conditioned on X . McFadden's random utility model was applied to derive each conditional probability and multinomial logit estimation was used to estimate covariates

that impact termination outcome (Vandell and Thibodeau 1985; Zorn and Lea 1989; Cunningham and Capone 1990)¹¹.

In addition to connecting mortgage borrower behavior with economic theory, studies applying the optimization model of consumer choice also contribute to the literature by introducing contemporaneous measures on mortgage characteristics to explain termination. Campbell and Dietrich (1983) find that both origination and contemporaneous LTV ratios have significant positive effects on default decision. Their results provide evidence of the importance of time varying covariates on mortgage termination. Borrower characteristics such as income variability are also found to have a significant positive effect on multi-period model for mortgage termination (Campbell and Dietrich 1983).

The second yet breakthrough connection between mortgage borrower behavior and theory is also made in this time period. In 1973, Black and Scholes (1973) developed a

¹¹ For example, Vandell and Thibodeau (1985) apply the following model to analyze default. The choice set in their model includes only two choices: continue to pay or default. The payoff function if borrowers choose to default can be expressed as follows:

$$W_D = (Y - R - Q_r)(1 + r_i) + W(1 + r_i)$$

The payoff function if borrowers choose to continue payment can be expressed as:

$$W_C = (Y - R - Q)(1 + r_0) + (V_T - L_T) + W(1 + r_i)$$

Where

W_D is payoff function if borrower defaults, W_C is payoff function if borrower continues payment, Y is real annual after-tax household income, R is required real non-discretionary expenditures (other than housing), Q_r is required real rent on new unit (gross rent plus utilities, etc.) r_i is expected real return on non housing investment, W is current real non-housing wealth, Q is required real after tax payment on mortgage (plus taxes, insurance, and other ownership costs) r_0 is expected opportunity cost of borrowing or return on lending, V_T is expected real market value of current home, and L_T is expected real outstanding loan balance on current mortgage.

Borrowers are expected to default instead of continue to pay their mortgage if W_D is greater than W_C .

contingent claims model and obtained an actual closed-form solution to the problem of valuing a call option on an underlying asset for short-run scenarios in which the interest rate is regarded as constant. Their results imply that investors can arrive at the value of a put or call option on a certain stock provided a few parameters such as current price and volatility of the stock are known. Their results have significant implications to mortgage studies because they provide motivations for default and prepayment behaviors, quantify the value of each behavior and predict under what circumstance shall default and prepayment occur.

The put and call options are defined financially as follows. A put option is an agreement that gives an investor the right, but not the obligation, to sell a stock, bond, commodity or other financial instrument at a specified price (also called strike price or exercise price) within a specific time period. The put option is usually purchased to protect against a fall in price. A call option, on the other hand, is an agreement that gives the holder the right but not the obligation to purchase an asset for a pre-determined price (also called strike price or exercise price) at or before the expiration date of the option. A call option is bought with the expectation of a rise in prices (<http://www.cftc.gov>).

In order to appreciate Black and Scholes results, one needs to realize that an option is a derivative asset whereas the stock itself is the fundamental asset. All market forces that govern prices, i.e., demand and supply reflected by individual preferences and firm technologies, have directly set the value of the stock, and the value of the option follows that of the stock. Since the payoff obtained from an option can also be obtained by

continuous costless adjustment of a portfolio consisting of the same underlying stock and some risk free bonds, the value of the option must equal the value of the sum of stock and the risk free bonds in order to prevent arbitrage. Because the value of the sum of stock and the risk free bonds are known, so is the value of the option. Thus Black and Scholes conclude that the value of an option can be calculated without any reference to individual risk attitudes or to the mean value of the stock price movement that govern the option's risk (Black and Scholes 1973). Volatility of the stock does matter in determining the value of an option to the extent that it determines the range of likely stock movements (Black and Scholes 1973).

Turning to studies on mortgage termination, the analogy between a mortgage on a house and an option on a stock is quite close. Both can be considered investment tools and both have underlying asset that back the value of the investment. In addition, similar to uncertainties in an option on a stock, there are also uncertainties in mortgages on real estate properties. The two fundamental sources of uncertainties in mortgages are interest rate risk and house price risk. If one views the house as a traded asset, then default portion of a mortgage can be viewed as a put option whereas prepayment can be viewed as a call option. By exercising default option, a borrower turns over possession of the house in exchange for abandoning payments; by exercising prepayment option, a borrower gains the right to the house by paying off the loan. Between the two uncertainties, house price risk is more associated with default option whereas interest risk is more associated with prepayment option.

Based on results from option-based model, the conditions driving default and prepayment behaviors are (1) conditions in the capital market – namely current and expected future interest rates that represent the opportunity cost of funds, and (2) the current and expected future price of the property. The two conditions are specified as stochastic processes representing expectations about the future pattern and volatility of these measures (Cunningham and Hendershott 1984; Hendershott and Van Order 1987; Kau, Keenan, Muller and Epperson 1987; Buser and Hendershott 1984; Brennan and Schwartz 1985; Kau, Keenan, Muller, and Epperson 1992, 1995; Kau and Keenan 1996; Titman and Torous 1989; etc.).

Take Kau, Keenan, Muller, and Epperson (1992, 1995) for example, a standard mean-reverting process is applied to interest rates and a standard lognormal process is applied to the property value trends. A borrower's prepayment decision is determined by whether the value of the mortgage under contract interest rate (also called coupon rate) is less than that of a new mortgage under the market interest rate. A default decision is determined by whether the value of the mortgage under the market interest rate is greater than the current value of the property (details in Kau et al. 1992). At each decision point, a borrower decides which options to choose by comparing the market value of her mortgage with the par value and the market value of her property. Property value is assumed to be observable to borrowers. Solution to a partial differential equation solved backward is the current value of the mortgage¹².

¹² Following Kau et al. (1992), the value of a mortgage $M(c, r, H, B, k)$ depends on the coupon rate c , a vector of relevant interest rates r , the property value H , the outstanding balance B , the age of the loan k , and other parameters. For simplicity, they assume there is only one interest rate that determines the yield curve. Also assume a continuous house price change with an instantaneous mean and a standard

Note that a mortgage borrower does not need to solve for the differential equation in order to know when to exercise either option (Kau et al. 1995). All she needs to know is whether or not she can refinance the remaining mortgage for the same term at a mortgage rate lower than her current interest rate. If the answer is yes, prepayment option shall be exercised immediately. If the borrower knows that she can obtain a new mortgage on the same property for the same remaining term yet with a lower than current payment, default option shall be exercised immediately. Hence the option-based model predicts

deviation σ_H . The interest rate denoted by r and the value of the house denoted by H are assumed to follow a stochastic process:

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

$$\frac{dH}{H} = (r - d)dt + \sigma_H dz_H,$$

$$dz_r dz_H = \rho(r, H, t)dt,$$

where θ is the mean value of the interest rate r , γ is the rate of convergence for the interest rate, d is the imputed rent payout rate (dividend), σ_r is the volatility of the interest rate, $\sigma_r \sqrt{r}$ is the instantaneous standard deviation of the term structure, σ_H is the instantaneous standard deviation of the house price, dz_r and dz_H are standard Brownian motion with $E(dz) = 0$ and $E(dz^2) = dt$, and ρ is the correlation between the disturbance to the term structure and the disturbance to the house price (Kau et al. 1992).

Kau et al. (1995) shows that under the perfect capital market assumption and the Local Expectations Hypothesis, the value of the mortgage M satisfies:

$$\frac{1}{2} r \sigma_r^2 \frac{\partial^2 M}{\partial r^2} + \rho \sqrt{r} H \sigma_r \sigma_H \frac{\partial^2 M}{\partial r \partial H} + \frac{1}{2} H^2 \sigma_H^2 \frac{\partial^2 M}{\partial H^2} + \gamma(\theta - r) \frac{\partial M}{\partial r} + (r - d)H \frac{\partial M}{\partial H} + \frac{\partial M}{\partial t} - rM = 0,$$

a second order partial differential equation such that the value of the mortgage equals the risk-adjusted expected present value of its net cash flows. This follows from Black and Scholes (1973). The value of $M(c, r, H, B, k)$ and the optimal default and prepayment strategy are determined simultaneously. $M(\cdot)$ is determined by choosing r^* and H^* that minimize $M(\cdot)$ given the above equation. The optimal level of r and H are functions of parameters underlying the stochastic process of interest rate and house value as well as c, r, H, B, k (Kau et al. 1995). The decision rules are such that borrower choose to exercise default option when the house value falls to H^* , and prepay when the current interest rate falls to r^* .

that prepayment or default option will be immediately exercised once either option is “in the money”.

Many recent studies have applied the above option-based model to mortgage market by modeling default as a put option and prepayment as a call option. They calculate the value of the put and call options and tested whether default and prepayment occur as option theory predicts. Hendershott and Van Order (1987) and Kau and Keenan (1995) provide survey to literature that relates to option-based mortgage pricing.

Early applications fail to recognize the interdependence between default and prepayment and tend to focus on either of them however. There are many studies that use option theory to model default. For example, Foster and Van Order (1984) apply the option theory to mortgage default using FHA data from 1960 through 1978. They estimate an equity position as measured by contemporaneous LTV ratios over time and create a number of variables that represent the percentage of loans with negative equity for each year in the sample. Their estimates on the current and lagged equity variables explain over 90 percent of the variance. Cunningham and Hendershott (1984) and Epperson, Kau, Keenan, and Muller (1985) apply option models to price default risk by modeling default as a put option. Foster and Van Order (1984) and Quigley and Van Order (1995) estimate default model empirically in an option-based framework using fixed rate residential mortgages. Buser, Hendershott and Sanders (1985), Cox, Ingersoll, and Ross (1980), Findlay and Capozza (1977), Kau et al. (1985, 1990, 1993) and Schwartz and Torous

(1991) analyze residential adjustable rate mortgages in option-based models. . Quercia and Stegman (1992) and Vandell (1993) give reviews on empirical default models¹³.

Just like studies that focus on default, there are also many studies that rely on option theory to analyze prepayment behavior. For instance, Dunn and McConnell (1981), Buser and Hendershott (1984) analyze prepayment option of adjustable-rate (ARM) and fixed-rate mortgages (FRM). Titmand and Torous (1989) analyze special case of commercial mortgages where prepayment is prohibited through yield maintenance features and prepayment penalties. Green and Shoven (1986), Hendershott and Van Order (1987), Schwartz and Torous (1989), Stanton (1994), Caplin Freeman and Tracy (1997), to name a few, analyze mortgage termination by prepayment using aggregate data. Findley and Capozza (1997) analyzed the prepayment options of adjustable-rate mortgages (ARM) and fixed rate mortgages (FRM). Archer, Ling and McGill (1996) use American Housing Survey data from 1985 to 1987 to examine the influence of post-origination income and collateral constraints on prepayment behavior, after controlling for the value of the call option.

In summary, the connection between option theory and mortgage termination behaviors provides a theoretical tool to analyze default and prepayment incentives and outcomes.

Researchers in mortgage studies calculate the extent to which the put and call options are in the money, based on which they predict when default and prepayment occur.

¹³ Research emphasizing financial variables that directly bear on a mortgage asset also include those by Cunningham and Capone (1990) and Jackson and Kaserman (1980). Empirical studies employing the option-theoretic approach include those of Archer and Ling (1993); Foster and Van Order (1984, 1985); and Quigley and Van Order (1991, 1992). Please see Quercia and Stegman (1992) for a comprehensive review of the empirical literature.

A straightforward application of Black and Scholes (1973)'s results on mortgage terminations provides us with a "frictionless" model in the sense that there is no transaction cost associated with either default or prepayment. The "frictionless" model assumes that borrowers can pay off the outstanding mortgage balance without penalty. It also assumes that borrowers can purchase a new property with equal value as the defaulted one and enter into a new loan agreement after immediately exercise the default option with no transaction or reputation cost. Following its assumption, the model predicts that borrowers will immediately exercise either option whenever default or prepayment option is "in the money", and hence the name "ruthless" termination behavior is used to refer to it.

However, there exists empirical evidence that shows mortgage borrowers do not exercise their default or prepayment options in the same manner as investors exercise financial options (Vandell 1995; Green and LaCour-Little 1999). In particular, mortgage borrowers do not ruthlessly exercise either option as the frictionless model implies (Foster and Van Order 1984, 1985).

For instance, Foster and Van Order (1984, 1985) find that only 4.2 percent of residential mortgages with contemporaneous LTV ratios in excess of 110 percent default, compared to a predicted 100 percent default outcome based on the ruthless model. The authors attribute this friction to transaction costs that can vary across borrowers. Transaction cost can be measured in the forms of loss of equity, search costs, and opportunity costs facing

different borrowers. For instance, a borrower who defaults today might lose her creditworthiness and face a limited access to credit in the future (Foster and Van Order 1985).

Other researchers also provide evidence of default non-ruthlessness. Vandell and Thibodeau (1985) use a set of disaggregate loan histories from a saving and loan association to evaluate the influences of both the equity and cash flow on default as well as the influence of non-financial trigger events such as divorce and unemployment. They conclude that an expected negative net equity of 10 percent is to cause less than a 5 percent likelihood of default on average. Their results also show that other than equity, measures of put option value, source of income, length of employment, level of wealth endowment and the condition of the neighborhood also affect the likelihood of default. Furthermore, some of these effects appear to have greater influences on default than the equity effect does. Their results suggest that solvency and trigger event such as loss of income due to social disruptions plays an important role in default. Deng, Quigley and Van Order (1996) also provide findings supporting the importance of trigger events such as unemployment and divorce in mortgage terminations. Vandell (1990, 1992) find similar evidence of “hesitated” default behavior among commercial mortgage borrowers using data from the American Council of Life Insurance. These types of models are referred to as “non-ruthless” models.

Evidence of non-ruthless exercise of the prepayment option is also present. For example, Archer, Ling and McGill (1996) analyze prepayment behavior and conclude that higher

annual debt-to-income and LTV ratios are negatively related to prepayments after controlling for call option values. Caplin, Freeman and Tracy (1997) find that regional recessions reduce prepayment rates by as much as 50 percent in states with declining property markets. Peristiani, et al. (1997)'s results suggest that poor consumer credit history as well as high contemporaneous LTV significantly reduce the probability of prepayment. These empirical findings are intuitive because if the property value decline below the loan balance, additional funds will be required to prepay the current mortgage. Similarly, a borrower whose income or financial position deteriorates may be unable to prepay due to constraints on credit quality or debt-to-income ratios.

Evidence on the existence of the non-ruthless option exercise behaviors as well as the importance and the impact of non-option related variables on mortgage terminations make testing both option and non-option variables a necessity. Most studies include both option and non-option related variables and conclude that they are both important in affecting mortgage default and prepayment. Most studies conclude that the probability of prepayment is an increasing function of the market price of the mortgage; it is a negative function of the contemporaneous LTV, the probability of negative equity, the local unemployment rate, minority status, borrower income and low credit score indicator (Clapp and Deng 2002). On the other hand, the probability of default is an increasing function of the market value of the mortgage, the contemporaneous LTV, the probability of negative equity, the local unemployment rate, and the debt-to-income ratio. It is a decreasing function of borrower's credit history.

One issue related with mortgage termination that is rarely analyzed is the existence of the unobserved heterogeneity among borrowers. Although no investigation has been done until recently, unobserved heterogeneity can selectively censor the observations and disproportionately impact the survival sample. If borrowers differ in their astuteness toward interest rate, then those who are financially savvy are more likely to exercise prepayment option and those who are financially unsophisticated are more likely to remain in the at risk population for default. Heterogeneity can also be caused by different opportunities such as unmeasured changes in property values and perceptions or abilities among borrowers. One example given in Vandell (1995) is that if borrowers who reside in the low end of the distribution of property values also tend to have high cost to default, the default option will be underestimated due to under exercise of the default option. Unobserved heterogeneity remains to be an important factor influence mortgage terminations and one needs to correctly account of it. Deng Quigley and Van Order (2000) is the first to control for unobserved heterogeneity in mortgage terminations to explain why some borrowers are less likely to exercise either option than other borrowers in a semiparametric framework.

Finally, virtually all early research based on option models tend to focus on one, not both termination outcomes. The most common approaches used in the literature rule out the possibility of default when valuing the right to prepay, and rule out the possibility to prepay when considering default¹⁴. Only until a series of papers by Kau et al. (1992,

¹⁴For instance, Schwartz and Torous (1989) analyzed Ginnie Mae mortgage prepayment experience by assuming the mortgages were free of default risk. Green and Shoven (1986) and Quigley and van Order (1990) made analogous assumptions in the analysis of prepayment behavior. Cunningham and Hendershott (1984) analyze mortgage default using a single hazard model.

1995); Kau and Keenan (1995); and Titman and Torous (1989) did the interdependent relationship between default and prepayment becomes theoretically established. A borrower who exercises default option today forfeits the option to default in the future as well as the option to prepay today or in the future and vice versa. It is important to realize that one can never ascertain whether default is optimal by considering only the value of the default option; one must look at the entire termination option set to decide whether termination is called for and what form it should take (Kau and Keenan 1995).

Recent studies recognize the importance of the competing risk nature and jointly estimate the two termination outcomes. Cunningham and Capone (1990) use multinomial logit estimation on a pool of Federal Housing Association (FHA) insured loans to analyze the concurrent loan termination (default and prepayment) experience at each payment period of FRM and ARM residential mortgage borrowers. Zorn and Lea (1989) undertake a similar analysis on a sample of Canadian ARM borrowers. Foster and Van order (1985) estimate simultaneous models of default and prepayment using pool data of FHA loans. Schwartz and Torous (1993) estimate joint hazard using Poisson regression on aggregate data. Deng, Quigley and Van Order (1996), Deng (1997) and Deng Quigley and Van Order (2000) jointly analyze residential mortgage default and prepayment behavior using disaggregated loan level data. Deng et al. (2000) further extend their studies by accounting for borrower heterogeneity to explain why some borrowers are less likely to exercise either option than other borrowers in a semiparametric framework. Huang and Ondrich (2002) apply a competing risk proportional hazard model on FHA insured multifamily mortgages and employ a bivariate Heckman-Singer nonparametric random

effect distribution to control for unobserved heterogeneity. They conclude that failure to control for it leads to severe downward biases in the coefficient estimates.

In summary, it is well established that option-based contingent claims model provides connection between formal economic and financial theory and mortgage borrower behavior. Although challenged, option related variables, specifically variables measuring the market value of the mortgages remain to be explanatory in determining mortgage terminations. Along with variables reflecting transaction costs and trigger events, option related variables continue to draw attention from researchers and are included in estimations of mortgage terminations. Competing risks models are called for when estimating mortgage terminations because it is well known by now that default and prepayment are interdependent and need to be jointly estimated. Borrower heterogeneity, if exists, should be accounted for if one needs to correctly draw inference on the effect of explanatory variables on either option.

3.2.3 Reviews of Econometric Methodology

Many estimation methodologies are proposed to study mortgage terminations. Until recently, the most commonly used models have been bivariate logit under a single risk assumption and multinomial logit model (MNL) under multiple risk assumptions (See Campbell and Dietrich (1983); Cunningham and Capone (1990); Arch, Ling and McGill (1996); Clapp, Goldberg, Harding and LaCour-Little (2001); Calhoun and Deng (2002)).

Bivariate logit model automatically rules out the risk of default when analyzing prepayment and rules out the risk of prepayment when analyzing default. It treats default as a censored observation when analyzing prepayment and treats prepayment as a censored observation when analyzing default. By construction, it does not allow for competing risks among different termination outcomes.

Unlike bivariate logit model, multinomial logit model allows for direct competition among different risks. By requiring all risks sum to unity, MNL ensures that an increase in one termination probability must be offset by a decrease in probability for one or more of the alternatives and hence guarantees competition among risks. However, MNL does not allow correlations among different risks because by construction it assumes independence across different risks. This assumption violates the empirical findings that risks are interdependent in mortgage terminations.

Multinomial logit also implies independence of irrelevant alternatives (IIA). IIA says that the odds ratio for any pairs of choices is independent of any third alternative and eliminating one of the choices should not change the ratios of probabilities for the remaining ones. Thus IIA implication limits the role of multinomial logit model in mortgage termination analysis when two choices are close.

An example illustrating the point above can be found in Clapp and Deng (2002)'s model on mortgage terminations where the authors attempt to model three termination outcomes including default, prepayment by refinance, and prepayment by sale of property.

Prepayment by refinance and prepayment by sale of property are considered closer risks than that of default. Not surprisingly, their results show that MNL models generate small and insignificant coefficients for the risk of prepayment by sale of property than all other risks.

One additional pitfall of using discrete choice models other than their inflexibility to accommodate critical assumptions on mortgage terminations is their inadequacy to analyze duration data. Discrete choice models such as logit are easy to employ, yet suboptimal to analyze continuous durations of mortgage life. Unlike discrete consumer choice problems, mortgage termination is a duration problem with underlying deterministic hazard functions. Discrete choice models lack the capability to analyze the shape of these hazards. Similar duration nature arises in other fields such as survival analysis in biostatistics, and unemployment spell studies in labor economics. The duration nature of mortgage calls for sophisticated econometric estimation techniques.

The estimation technique that is suitable to handle duration analysis is developed based on Cox's seminal work on proportional hazard model (Cox 1972). Cox (1972)'s proportional hazard model provides a flexible tool to analyze censored survival data. It allows for time varying and time invariant covariates and combines the advantages of both parametric and nonparametric estimation approaches for model building.

The Cox proportional hazard model is built on the basis of defining hazard function and survivor function. Assuming the probability density function of the duration of the event,

in this case, from loan origination to termination (default or prepayment), at t is $f(t)$; the cumulative probability distribution is $F(t)$. The hazard function is defined as the probability density of termination between time t and $t + \Delta$, conditional on surviving up to time t :

$$h(t) = \lim_{\Delta \rightarrow 0} \frac{\Pr(t < T < t + \Delta | T \geq t)}{\Delta} = \frac{f(t)}{1 - F(t)} = \frac{f(t)}{S(t)},$$

Where $S(t) = 1 - F(t)$ is called the survivor function.

The Cox proportional hazard model assumes that $h(t)$ is a product of a baseline hazard common to all individuals at each time period $h_0(t)$ and a component that depends on the covariates $z_i(t)$:

$$h_i(t) = h_0(t) \exp\{z_i(t)' \beta\},$$

$h_0(t)$ is the unknown baseline hazard at time t ; $z_i(t)$ is the vector of covariates that influences the termination decision for observation i at time t ; $z_i(t)$ can contain covariates that are time varying or time invariant; and β is the parameter vector to be estimated. The convenient exponential specification ensures that the hazard rate under different values of covariates is always positive. Theoretically, the hazard function is continuous and can take any nonnegative value. This flexibility, however, makes the empirical identification and estimation of the model a nontrivial exercise.

Several approaches have been proposed to estimate the hazard function. The simplest one is through parametric assumption on the distribution of the baseline hazard, most commonly Exponential, Logistic, Lognormal or Weibull, followed by estimation on the rest unknown functional parameters. However, studies in labor economics especially in unemployment duration have shown empirical evidence on inconsistencies between the assumed distribution and the actual hazard (Kiefer, 1988; Han and Housman 1990; Katz and Meyer 1990; McCall 1993, 1994, 1995)¹⁵. This is because the shape of the baseline hazard is often irregular and hence a simple parametric form is unlikely to well approximate it. The functional form assumption inevitably exerts constraints on the shape of the underlying hazard, forbidding the estimated hazard to reflect the real hazard. Meyer (1995) points out that the baseline hazard is often central to tests of economic hypotheses and imposing a shape for the baseline hazard may lead to incorrect conclusions about economic hypotheses.

In order to relax the assumption on the baseline hazard, Cox proposes Partial Likelihood specification (CPL) (Cox 1975; Cox and Oakes 1984). CPL estimation technique only requires the existence of a common stationary baseline hazard function $h_0(t)$ for all subjects with no specified functional forms. The likelihood function under this scenario is decomposed into two separate parts; one contains the unknowns in the baseline hazard function, the other contains unknowns of the partial likelihood of the proportional

¹⁵ For example, Katz and Meyer (1990) find irregular spikes in the hazard for unemployment spell at 26-27 weeks and 52-53 weeks due to termination of unemployment insurance. Similar spikes have been documented by independent studies.

changes (Cox 1975). Therefore, β can be identified without parametric restrictions on the baseline function because the baseline hazard can be factored out¹⁶.

This estimation methodology is called partial likelihood because CPL uses likelihoods contributed from observed terminations instead of all the individuals in the sample. It is also a semiparametric approach in the sense that one applies nonparametric estimation in the baseline hazard functions and parametric in the specifications of the proportional change (Cox 1975; Cox and Oakes 1984; Gill 1984).

Since researchers are also interested in the shapes of baseline hazard, remedies are proposed to CPL to estimate the baseline function. They include a two-step procedure

¹⁶Following Cox (1972), Miller et al. (1980), Andersen (1982) and Kalbfleisch and Prentice (1980), let $T_i, i = 1, \dots, N$ be independent continuously distributed positive random variables representing the times of termination of N individuals, each of whom can only be observed on a fixed time interval $[0, c_i]$ for certain censoring times $c_i, i = 1, \dots, N$. Suppose each individual i has hazard rate specified above: $h_i(t) = h_0(t) \exp\{z_i(t)' \beta\}$, where $h_0(t)$ is a fixed unknown baseline hazard rate for an individual whose $z \equiv 0$. Let $\mathfrak{R}(t) = \{i : T_i \geq t, c_i \geq t\}$ denote the risk set at time t , i.e., the set of individuals under observation at time t . Given $\mathfrak{R}(t)$ and that at time t only one individual in $\mathfrak{R}(t)$ is observed to terminate, the probability that it is precisely individual i can be expressed as $\frac{\exp\{z_i(t)' \beta\}}{\sum_{j \in \mathfrak{R}(t)} \exp\{z_j(t)' \beta\}}$; note

that $h_0(t)$ is cancelled out in the numerator and the denominator.

Cox then proposed that the statistical inference on β could be carried out by considering

$$L(\beta) = \prod_{i: T_i < c_i, i=1}^N \frac{\exp\{z_i(T_i)' \beta\}}{\sum_{j \in \mathfrak{R}(T_i)} \exp\{z_j(T_j)' \beta\}}$$

as a likelihood function for β , to which standard large sample maximum likelihood theory applies. Each term in the above product is the probability that at time T_i of an observed termination, it is precisely individual i who is observed to terminate. After the likelihood is decomposed to two parts, common maximum likelihood estimation approaches shall be applied to estimate β . Also see Lancaster (1990), chapter 8.2.5 and Amemiya (1985), chapter 11.2.7.

where regression coefficients β are first identified through CPL. These estimates are then employed in the full likelihood estimation to obtain the necessary parameters for a flexible specified baseline hazard function, either fully unspecified (Meyer 1990) or some high-order polynomial function of time (See Fu, LaCour-Little and Vandell (2000) for an application).

Although extremely flexible and powerful, CPL still faces pitfalls when applied to economic data, specifically in cases where surviving data is grouped with many ties in failure time (i.e., numerous individuals are with the same reported duration).

Grouped data with ties in failure time are common in economic duration data. For example, length of unemployment spells is measured in weeks; mortgage terminations are reported in months or quarters. The grouping data naturally makes ties a frequent phenomenon. However, CPL assumes that duration data are continuous and is not capable to handle grouped data with ties. As illustrated in Prentice and Gloeckler (1978), CPL technique suffers from computational inconvenience in the event of grouped data with ties. CPL is a product of terms, one from each distinct failure time. If m failure times are tied at time t and n individuals are at risk prior to t , the partial likelihood contribution involves a summation over all possible subsets of size m from the n at risk¹⁷. With many failures occurring at the same time, the likelihood becomes intractable and is theoretically and computationally infeasible to obtain.

¹⁷The CPL partial likelihood contains a term for every subset of the risk set at time t , which contains exactly d_t elements, where d_t is the number of failures occurring at time t when ties occur. Denote \mathfrak{R}_t as the risk set (those alive and uncensored just prior to t) at t , then the number of such terms is

A series of papers by Prentice and Gloeckler (1978), Meyer (1986, 1990), Han and Hausman (1990) have proposed grouped data version of the proportional hazard model to fill the gap. The grouped data proportional hazard model, also referred to as “flexible” proportional hazard method, is able to obtain computationally feasible estimates for β and the corresponding survival function in the presence of ties.

The “flexible” proportional hazard model assumes that failure times are grouped into intervals and regression vector is allowed to be time-dependent, but fixed within a specific time interval. As in CPL, the hazard function for individual i at time t is still in the form of $h_i(t) = h_0(t) \exp\{z_i(t)' \beta\}$, where $h_0(t)$ is the unknown baseline hazard at time t ; $z_i(t)$ is the vector of covariates.

Following Prentice and Gloeckler (1978)’s specification, failure times are grouped into intervals $A_s = [a_{s-1}, a_s)$, $s = 1, \dots, r$ with $a_0 = 0, a_r = \infty$. Failure times in A_s are recorded as t_s and regression vector $z_i = z_i(t_s)$ for individual i in interval A_s is fixed within the time interval A_s . The probability of observing a failure time t_s on an individual i with regression vector z_i is

$$[1 - \alpha_s^{\exp\{z_i(t_s)\beta\}}] \prod_{j=1}^{s-1} \alpha_j^{\exp\{z_i(t_j)\beta\}},$$

$\frac{R_t!}{(R_t - d_t)!(d_t)!}$. Even without ties, a full parametric likelihood function with continuously changing baseline hazard rates is computationally difficult to converge (Meyer 1995).

where $\alpha_j = \exp\{-\int_{\alpha_{j-1}}^{\alpha_j} h_0(u)du\}$ is the conditional survival probability in A_j for an individual with $z_i(t_j) = 0$; $h_0(t)$ is the baseline hazard. The probability of surviving to the beginning of A_s is $P(t_s, z_i) = \prod_{j=1}^{s-1} \alpha_j^{\exp\{z_i(t_j)\beta\}}$.

Let (t_k, δ, z) represent the survival time, censoring indicator ($\delta = 0$ if censored, $\delta = 1$ if failure) and regression vector for a study subject. Prentice and Gloeckler (1978) show that the contribution to the likelihood from this subject is $\{1 - \alpha_k^{\exp\{z\beta\}}\}^\delta \prod_{j=1}^{k-1} \alpha_j^{\exp\{z\beta\}}$. The likelihood function of the sample is the product of the above term over all individuals in the sample.

With specified grouping intervals the discrete model above includes a finite number of parameters $(\alpha_1, \dots, \alpha_{r-1})$ and β . Those estimators are computationally feasible to obtain.

Meyer (1986) compares estimates obtained from Prentice and Gloeckler (1987) method and those obtained from parametric estimations. His results show that the semiparametric estimators yield more plausible and consistent coefficients than those from Weibull models if the baseline hazard shape assumption is incorrect. After several distribution simulations, Meyer (1986) concludes that the Prentice and Gloeckler (1987) method allows for consistent estimation of β when the form of the baseline hazard is unknown. The conventional parametric assumption on the baseline hazard, on the other hand, results in inconsistent estimates of β when the assumed baseline is incorrect.

After the development of the flexible proportional hazard model, researchers are armed with a necessary tool to confront grouped economic data. However other economic issues remain to be solved. One of such issues is how to treat unobserved heterogeneity presented in the population.

Economists who work on hazard models have been emphasizing the importance to control for unobserved individual characteristics or heterogeneity in the estimation of the hazards (Lancaster 1979; Lancaster and Nickell 1980; Heckman and Singer 1982, 1984, 1986; Han and Hausman 1990; McCall 1996; Meyer 1990, 1995; Kiefer 1988). As one expects, characteristics of individuals that cause a lower hazard would be more present among the surviving individuals. Unobserved individual heterogeneity has been shown to affect failure time in labor economics, specifically in the studies of unemployment spells (Lancaster 1979; Lancaster and Nickell 1980). For example, researchers discover that individuals who remain in the survival sample in unemployment spell studies tend to be disproportionately high benefit, non-white and aged.

If incorporated in CPL model, the unobserved heterogeneity will generate multiple integrals of the same order as the number of individuals in the risk set, which makes estimation difficult, if not impossible (Meyer 1986). Fortunately, the grouped data proportional hazard specification can easily accommodate the individual heterogeneity.

There are two ways to specify the heterogeneity: parametric and nonparametric heterogeneity. Examples of the former include distribution assumptions on the heterogeneity in the population. For example Lancaster (1979), Han and Hausman (1990), and Meyer (1990) have assumed Gamma distribution on the heterogeneity; others have assumed lognormal distribution or discrete distributions (Meyer 1986). Examples of the latter begin with works by Heckman and Singer (1982). Heckman and Singer (1982, 1984, 1986) propose a nonparametric technique for the estimation of the heterogeneity distribution, which generates a discrete distribution on the heterogeneity with a finite number of mass points. They argue that estimates of structural parameters obtained from ad hoc choices of the heterogeneity parametric distributions are very sensitive to the choice of mixing distributions. Instead, they develop a consistent nonparametric maximum likelihood estimator for the distribution of unobservable and a computational strategy for implementing it. Heckman-Singer heterogeneity distribution is a discrete distribution with J points, each with associated probability p_j that sums to one. Their research suggests that parametric estimation on the heterogeneity is very unstable and nonparametric estimation should be employed (Heckman and Singer 1982). Meyer (1986), McCall (1996, 1997), Deng, Quigley and Van Order (2000), Ciochetti, Deng, Gao and Yao (2002), Huang and Ondrich (2002) adopt this method and assume nonparametric heterogeneity in estimating hazards.

A natural extension of single-risk duration analysis with grouped data (Prentice and Gloeckler 1978; Kiefer, 1988) is how to modify it to apply to the competing risks framework with more than one risk. Competing risks models are named such because the

shortest realized risk-specific duration makes the durations of other risks right censored. As illustrated in previous paragraphs, competing risks models are appropriate when multiple risks are interdependent and hence should be estimated jointly. For example, studies on unemployment spells terminated by full time employment or part time employment, and mortgage terminations by default or prepayment are cases where competing risks models are called for. In the case of mortgage terminations, the two termination outcomes are correlated. If good quality mortgagors exercise prepay option ruthlessly in early age of the mortgage whereas bad quality mortgagors cannot exercise prepayment option due to institutional constraints, then the at risk population will be adversely selected as time goes on. Therefore, one shall correct for the sample selection problem by modeling termination hazard model in a competing risks framework.

Previous studies deal with competing risk in a very restricted manner. Many assume restricted parametric form on the risks such as bivariate lognormal distribution (Diamond and Hausman 1984). Others assume independence between the risks on the ground that unrestricted interdependent risks model is unidentifiable (Kalbfleisch and Prentice 1980; Katz 1986).

Han and Hausman (1988) provide proofs for identification of the bivariate competing risks proportional hazard model under certain regularity conditions and clear the way to estimate competing risk model with unrestricted correlation. Based on their results, Han and Hausman (1990), Sueyoshi (1992), and McCall (1996) suggest a maximum likelihood estimation approach to expand the single risk flexible proportional hazard into

an unrestricted competing risk framework. Their approach estimates the competing risks simultaneously in a proportional hazard specification, allows for unrestricted correlation among different risks, and accounts for presence of individual unobserved heterogeneity. Time varying covariates are also allowed in their approach. The difference among them is that Han and Hausman (1990) adopt piece-wise linear integrated baseline hazards in the estimation and assume parametric form on the heterogeneity whereas Sueyoshi (1992) and McCall (1996) assume the integrated baseline hazard takes the form of a fifth order polynomial specification and adopt Heckman and Singer (1982)'s approach to non-parametrically estimate the heterogeneity jointly with the baseline hazard and the coefficients on the covariates.

Following McCall (1996), let T_a, T_b be risk a and risk b , t_a, t_b be the duration of a subject until it is terminated by risk a and risk b respectively. The joint survivor function conditional on z, θ_a, θ_b , $S(t_a, t_b | z, \theta_a, \theta_b) = \Pr(T_a > t_a, T_b > t_b | z, \theta_a, \theta_b)$, is defined as

$$S(t_a, t_b | z, \theta_a, \theta_b) = \exp\left\{-\theta_a \sum_{t=1}^{t_a} \exp[\alpha_a(t) + z(t)\beta_a] - \theta_b \sum_{t=1}^{t_b} \exp[\alpha_b(t) + z(t)\beta_b]\right\}.$$

Here z is time varying or time invariant covariate, θ_a, θ_b are parameters for unobserved heterogeneity, and $\alpha_a(t), \alpha_b(t)$ are the log of integrated baseline hazard rate for risk type a and b between $t-1$ and t , i.e., $\alpha_s(t) = \log\left(\int_{t-1}^t h_{0s}(u)du\right)$. $h_{0s}(u)$ is the underlying

continuous time baseline hazard function, $s = a, b$. McCall assumes $\alpha(t)$ is in the form of a fifth order polynomial in t .

The competing risk nature makes only the first realized termination observable, that is $t = \min(t_a, t_b)$. Define $F_a(k | \theta_a, \theta_b)$ as the probability of termination by risk a in period k , $F_b(k | \theta_a, \theta_b)$ as the probability of termination by risk b in period k , and $F_c(k | \theta_a, \theta_b)$ as the probability that the surviving spell last more than k period.

Following McCall (1996), the above probabilities can be expressed as

$$F_a(k | \theta_a, \theta_b) = S(k, k | \theta_a, \theta_b) - S(k+1, k | \theta_a, \theta_b) - 0.5\{S(k, k | \theta_a, \theta_b) + S(k+1, k+1 | \theta_a, \theta_b) - S(k, k+1 | \theta_a, \theta_b) - S(k+1, k | \theta_a, \theta_b)\},$$

$$F_b(k | \theta_a, \theta_b) = S(k, k | \theta_a, \theta_b) - S(k+1, k | \theta_a, \theta_b) - 0.5\{S(k, k | \theta_a, \theta_b) + S(k+1, k+1 | \theta_a, \theta_b) - S(k, k+1 | \theta_a, \theta_b) - S(k+1, k | \theta_a, \theta_b)\},$$

$$F_c(k | \theta_a, \theta_b) = S(k, k | \theta_a, \theta_b).$$

McCall uses the term

$0.5\{S(k, k | \theta_a, \theta_b) + S(k+1, k+1 | \theta_a, \theta_b) - S(k, k+1 | \theta_a, \theta_b) - S(k+1, k | \theta_a, \theta_b)\}$, as an adjustment because durations are measured in discrete time rather than continuously.

The unconditional probability is then given by

$F_m(k) = \sum_{j=1}^J p_j F_m(k | \theta_{aj}, \theta_{bj})$, where $m = a, b, c$, and $j = 1, \dots, J$ is number of distinct and

unobserved types of people and p_j defines the relative frequency that these types of people occur in the population.

The log likelihood function of the fully specified competing risk proportional hazard model with unobserved heterogeneity is given by

$$L = \sum_{i=1}^N \{I_{ia} \log(F_{ia}(k_i)) + I_{ib} \log(F_{ib}(k_i)) + I_{ic} \log(F_{ic}(k_i))\},$$

with N being the total number of subjects, I_{ia}, I_{ib}, I_{ic} being the indicators that equals one if the i th subject is terminated by risk a , risk b , or censoring respectively and zero otherwise.

McCall's method is widely adopted by recent studies on mortgage termination (Deng, Quigley and Van Order 2000; Deng and Quigley 2000; Ambrose and LaCourLitle 2001; Ciochetti, Deng, Gao and Yao 2002; Huang and Ondrich 2002) despite that the likelihood function with the adjustment term he proposed is an ad hoc approximation of the sample likelihood function.

A recent paper by An (2004) shows that competing risks proportional hazard model with grouped duration data is unidentifiable if the baseline hazard is assumed to be non-parametric; functional form assumptions on the baseline hazard are needed for any

consistent estimation and meaningful inference. The non-identification for the competing risks model is purely due to data grouping and is irrelevant to the existence of heterogeneity. Unlike in single risk proportional hazard model with grouped data where the likelihood function depends on the baseline hazard only through the discrete values of the integrated hazard function evaluated at limited numbers of points, in competing risks model the likelihood function depends on the values of all risks' baseline hazards at each time between each grouped intervals. Provided sample size is always smaller than infinite, functional form assumptions on the baseline hazard are required in grouped data proportional hazard competing risks model.

An (2004) also shows that under certain parametric assumption such as piecewise constant baseline hazards, the sample likelihood function has an explicit analytical form and there is no need for approximation once one makes the assumption that the baseline hazards are piecewise constant. The piecewise constant baseline assumption ensures that the integrated baseline are in the form of piecewise linear and the likelihood function depends on the baseline hazards only through fixed number of discrete values of the integrated baseline hazards. Provided the number of groups in the sample is smaller than sample size, the parameters are identifiable.

Recent studies by Deng, Quigley and Van Order (2000), Deng and Quigley (2000), Ambrose and LaCourLitle (2001), Ciochetti, Deng, Gao and Yao (2002) adopted McCall's approximation formula of the likelihood function, which is inconsistent with their piecewise constant assumption of the baseline hazard. In addition, specific to

mortgage termination models is the fact that default is an extremely rare event, with a hazard rate of less than one percent of that of the prepayment hazard rate. Therefore, McCall's one half one half split between the two risks, although might be suitable in unemployment spell studies, is inappropriate to apply to mortgage studies. More work is needed where an explicit analytical form of the likelihood function is available in mortgage studies. In addition, Deng et al. (2000a, 2000b) and Huang and Ondrich (2002) attempt to control for unobserved heterogeneity by Heckman-Singer's mass point nonparametric approach. Their specification limits the heterogeneity in a restrictive form and hence forces the results to be in contrast to empirical evidence. I attempt to apply a different Heckman-Singer heterogeneity approach to relax the restrictions.

In summary, the econometric estimation technique on mortgage terminations advances from discrete choice model to proportional hazard model to semiparametric flexible proportional hazard model to competing risk proportional hazard model with grouped data with heterogeneity. As the technique sophisticates, so are the issues addressed. Up to now, researchers on mortgage terminations are able to address unrestricted form of correlation between competing risks along with unrestricted form of individual heterogeneity. They can also investigate the shapes of the hazards to further understand termination behaviors.

The above section gives a comprehensive review of the literature review, including review on early studies, economic theory, option theory and review on econometric methodology. In the following section, I attempt to apply the option theory to test the

non-ruthless model specification while controlling for the unobserved heterogeneity. I attempt to improve the literature by introduction an unmeasured, yet important factor that significantly influence mortgage default. In addition, I apply Heckman-Singer's unobserved heterogeneity to allow for a more flexible correlation between heterogeneities on default and prepayment than the specification in Deng, Quigley and Van Order (2000). Lastly, I will adopt a latest result by An (2004) on the identification of the sample likelihood function in the estimation of competing-risk models under the proportional hazards assumption with grouped duration data and introduce the exact likelihood function to replace the commonly used approximation formula.

3.3 Data

3.3.1 Dataset

In this section, I shall provide a brief explanation of the dataset that I employed in the paper. The empirical analysis is based on individual mortgage history data maintained by a large mortgage institution. The database contains 508,219 observations on single-family mortgage loans originated from 1992 to 2000. All loans are one unit and owner occupied properties with 30 or 15 years of fixed-rate mortgage terms. All loans are observed from their origination dates till December 2003 or till their termination dates. The maximum length a loan can be observed in the sample is 48 quarters.

For each mortgage loan, the available information includes the origination and termination time, termination reason (prepayment, default, or censoring), the original loan amount, the initial loan-to-value ratio, the mortgage contract interest rate (coupon rate), the monthly housing expense including principal, interest and insurance payment,

state, region and the zip code in which the property is located. Borrower information such as credit score, income, and debt-to-income ratio are also presented in the dataset. Since all loans mortgage term are either 15 or 30 years, and the observation window is 12 years, no loan exits the sample as a matured loan. Therefore, the analysis is confined to loans that were either terminated or censored at the end of 2003. The duration is measured in quarters.

This database is unique in the sense that the observations in it are paired in the following way: each property has appeared chronologically twice in the dataset. The first time the property enters the database is through a purchase transaction where a borrower obtains a mortgage from purchasing the property. The second time the same property enters the database is in the form of either a refinance or another purchase transaction, which occurs later in time than the first purchase transaction. Hence there are two pairs of transactions in the database: purchase to refinance and purchase to purchase. Purchase to purchase pair is obtained when the title of the property exchanges hands and the owner of the property changes after the transaction. Purchase to refinance pair is realized when the property owner refinances her current mortgage and obtain a new mortgage on the same property; no ownership changes are involved¹⁸.

The first purchase transaction provides the property's market value that is agreed upon between a buyer and a seller. It serves as the base property value from which the property appreciates. Since the housing market competitively decides the purchase price of the

¹⁸ The appraisal bias measure is obtained through the purchase to refinance transactions.

property, we assume no systematic bias exists on the value of the purchase transaction property. By the same reason, the purchase-to-purchase transaction is assumed to exhibit mean zero value bias. The property value of the refinanced transaction, however, is not determined in the same manner. Because there is no demand for the property, the value of the property is not determined by the market. It is rather what a lender designated appraiser appraises at. Due to the information asymmetric structure described in the previous chapter, lenders and appraisers have an incentive to inflate the property value. Therefore, there is potential appraisal bias in the property value of the refinanced transaction.

3.3.2 Appraisal Bias

The structure of the secondary mortgage market creates asymmetric information between mortgage lenders and mortgage buyers. A profit maximizing lender would manipulate the application, e.g., overvalue the collateral, in order to achieve favorable outcome without being subjected to penalties. Fraud on loans delivered to Fannie Mae or Freddie Mac is directly undetectable because these two government-sponsored enterprises are limited to participate in the secondary market only by Charter rules set up by Congress. Therefore, an objective third party is required to verify the truth of the collateral values. This job is done entirely by the real estate appraisal industry.

The real estate appraisal industry certifies the value of collaterals for mortgage transactions and provides an estimate of the property value through an appraisal process. Their estimate helps mortgage lenders and borrowers determine if their investment is

worthy. In theory, real estate appraisers should maintain independence of mortgage borrowers and lenders in order to present objective and unbiased opinion on the value of the property. The system is set up to protect mortgage lenders and ensure that they can recoup their investments if forced to foreclose on a property, it also protects borrowers from paying more than what their homes are worth. However, recent changes in mortgage market structures and intensive competition among appraisers prevent real estate appraisers to fulfill their duty objectively.

In the past when the residential mortgage market was dominated by low loan to value (LTV) purchase money loans to high quality borrowers directly from lenders, the appraiser was considered a fact checker for a newspaper or magazine. His job boiled down to making sure the contracted-for purchase price was reasonable. Very occasionally, appraisers were asked to determine if the borrower would have enough equity in the transaction so that the lender would be protected in case of default or if the lender should require private mortgage insurance (MI). Most lenders simply counted on rising housing prices, making such decisions irrelevant.

The independent third-party fact-checking function of appraisers has been compromised due to a drastic change in the mortgage market environment. Today, every aspect of that worry-free environment that existed throughout most of the 1970s and 1980s is gone. For example, in 1990s, refinancing and home equity loans had come to dominate the residential mortgage market, making purchase money mortgage less than one quarter of mortgage transactions. Currently, high LTV loans account for 25 percent of mortgage

productions, and mortgage brokers who do not hold mortgages in their portfolios originate more than half of all loans in the primary market. In addition, consumer delinquencies are on the rise and create a growing pool of sub-prime borrowers.

Appraisers are tempted to follow these market trends to secure their future businesses. In the worst-case scenarios, they achieve the goal by inflating stated property values. Instead of knowing the purchase price in a sale transaction, the appraiser finds out the size of the underlying mortgage that the homeowner wants to refinance or how big a credit line he wants to establish. Often appraisers are given this information by the homeowner or the mortgage originator/lender. Or, with matrix pricing by LTV becoming more and more popular in the sub-prime market, the appraiser finds out what interest rate the homeowner wants to qualify for and report the property value accordingly. In other cases, appraisers inflate property value in order not to be responsible for imposing the cost of private mortgage insurance on a homeowner, a percentage of the unpaid principal balance, which could be as high as half of the mortgage payment. Competition has accentuated this trend. Even in the peak refinancing year of 1994, there were only an average of 100 transactions per appraiser generating an average income of \$25,000 to \$30,000. Their average incomes declined by more than one third over the next two years. Given that most defaults do not happen in the first year of loan origination and a steady house price appreciation, appraisers have incentive to inflate property prices to attract business without getting penalized.

Pressure from mortgage lenders is another reason for appraisers to inflate property appraisal values. “The people who certify what homes are worth when they are sold or refinanced say they are being pressured to inflate their numbers to ensure that the lending deals go through”, cited from the Denver Post (The Denver Post, July 20, 2003).

The property valuation bias created could lead to substantial losses for the economy. For example, it was inflated appraisals and rampant property flipping that were among the many factors that led to the savings and loan crisis of the 1980s, which led to the ultimate failure of a large number of US’s thrifts while significantly weakening many lending institutions. Many of the thrifts subsequently had to be closed or bailed out by the federal government, which seized and then resold thousands of overvalued properties.

Since then, the real estate appraisal industry has received severe criticism. On September 23, 1986, U. S. House of Representatives Committee on Government Operations approved and adopted a report, titled “Impact of Appraisal Problems on Real Estate Lending, Mortgage Insurance, and Investment in the Secondary Market”. The so-called “1986 Barnard Report,” named after then chairman, Congressman Douglas Barnard, Jr., concluded that more than 800 federally insured savings and loan associations have “significant appraisal deficiencies” and on top of that, more than 300 were declared insolvent or placed in “problem status” by federal regulators. The Committee also reported that 10-15% of the \$1.3 billion in losses suffered by private mortgage insurers in 1984-1985 could be attributed to inaccurate and fraudulent appraisals, and that 10-40% of

the \$420 million in loan losses at the Veterans Administration in 1987 was caused by inaccurate or dishonest appraisals or other appraisal-related deficiencies.

The findings and conclusions of this report stated that a wide range of corrective measures would have to be developed and instituted by federal regulatory authorities, the appraisal industry, and the real estate finance and investment interests. Following the report, a series of mandated reforms and regulatory requirements in the appraisal profession were introduced. They include the emergence of the Appraisal Foundation in 1987; Title XI of the Financial Institutions Reform, Recover, and Enforcement Act of 1989 (know as FIRREA); Uniform Standards of Professional Appraisal Practice; state licensing; quality control measures; and appraisal review guidelines.

Despite a series of appraisal regulations, updates, mandates, procedural guidelines, and advisory opinions, faulty and fraudulent appraisals still remain a national problem. The FBI recently reported that 10 to 15 percent of all loan applications contains material misrepresentations, such as inflated property valuations (The Denver Post July 20, 2003). Many times these fraudulent loans end up in foreclosure, resulting in financial losses to mortgage buyers. Their estimates showed the total price of mortgage fraud could be as high as \$120 billion, with the Mortgage Banker Association (MBA) estimating over \$60 billion from its reporting membership. The Washington Post's recent article on mortgage fraud quotes an FBI official that "mortgage fraud has the potential to become a 'national epidemic' that could expose lenders to hundreds of millions of dollars in losses." (The Washington Post, September 18, 2004; Page E01). "Appraisal bias, spreading like a

cancer, is eating away at the industry's moral foundation" (NAIFA – AppraiserE-Gram, 2001).

OFHEO (the Office of Federal Housing Enterprise Oversight), the federal oversight agency for Fannie and Freddie, has also identified appraisal bias as an issue in determining how much capital they should set aside for any market downturns (risk based capital requirements)¹⁹. The Office of Federal Housing Enterprise Oversight is directed by the Federal Housing Enterprises Financial Safety and Soundness Act of 1992 to develop a risk-based capital regulation for Freddie Mac and Fannie Mae (collectively, the Enterprises). The regulation specifies the risk-based capital stress test that will determine the amount of capital an Enterprise is required to hold to maintain positive capital throughout a ten-year period of economic stress. The results of the risk-based capital stress test will be used to determine each Enterprise's risk-based capital requirements and, along with the minimum capital requirement, to determine each Enterprise's capital classification for purposes of possible supervisory action.

Appraisal bias, defined by OFHEO, can result from “the perceived tendency of appraisers, as agents of primary mortgage lenders, to impart an upward bias to a home value to insure that a home sale is made”²⁰. OFHEO recognizes that appraisal bias could affect the rates generated by the stress test if the method of computing the house price index were changed in some way to account for appraisal bias or if appraisal bias were found to be significantly different in more recent data than in the historical data used to

¹⁹ Federal Register: April 13, 1999 (Volume 64, Number 70), Page 18083-18132.

²⁰ Federal Register: April 13, 1999 (Volume 64, Number 70), Page 18125.

estimate the models²¹. As of September 30, 2004, Fannie Mae's risk-based capital requirement was \$18.342 billion. Fannie Mae's total capital of \$38.762 billion on that date exceeded the risk-based capital requirement by \$20.420 billion (OFHEO News Release Thursday, December 21, 2004). As of September 30, 2004, Freddie Mac's risk-based capital requirement was \$5.749 billion. Freddie Mac's total capital of \$34.397 billion on that date exceeded the requirement amount by \$28.648 billion (OFHEO News Release Thursday, December 30, 2004). OFHEO continues to conduct analysis on appraisal bias and its effect on risk-based capital requirement for the GSEs; no conclusive findings on how the risk-based capital shall be impacted are available at this moment. Nevertheless, OFHEO's awareness of appraisal bias and its willingness to investigate the impact of appraisal bias on risk-based capital requirement for the GSEs suggests that appraisal bias is an economically significant phenomenon.

The following figures help quantify the importance of appraisal bias. As of January 2005, 16 percent of outstanding subprime market refinance loans are reported to have exactly 80 percent origination LTV ratio, translating into \$35.57 billion in unpaid principal balance²². This LTV ratio allows the borrowers to pay the minimum required down payment while avoiding primary mortgage insurance payments. According to the estimation data used in this paper, the mean ratio of reported property value to predicted property value for refinance loans with 80 percent LTV is 1.05; the standard deviation is 0.12. Therefore, 97.72 percent of those loans should have predicted LTV less or equal to

²¹ Federal Register: April 13, 1999 (Volume 64, Number 70), Page 18125.

²² The total outstanding balance for refinance loans in the subprime market is \$219.0 billion as of January 2005.

103 percent²³. 50 percent of those loans should have predicted LTV greater than 84 percent, which would require primary mortgage insurance coverage between 12 to 35 percent of the unpaid principal balance; the exact coverage also depends on other loan characteristics²⁴. For simplicity, let's assume these loans would require on average 15 percent mortgage insurance coverage. Multiplying their unpaid principal balance amount, the coverage will translate into \$5.3 billion of loss cushion for default. This amount of money is what mortgage buyers are able to recoup from primary mortgage insurers should the loans default during the lifetime of a loan. The loss within a shorter period of time depends on the default rate within shorter period. If we assume that the cumulative default rate within 3 years is 1 percent, then the loss generated by defaults in 3 years would be \$0.35 billion, out of which \$53 million shall be recovered. However, these loans are not required to purchase mortgage insurance since they are reported to have exactly 80 percent LTV ratio, making the would have been there loss cushion not recoupable. Since default risk is strongly determined by LTV as literature establishes, the underreported LTV due to appraisal bias will generate unmeasured default risk and subsequently un-recoupable loss to mortgage buyers.

Because most lenders only use their in house appraisers, I assume the appraisers and the lenders are one entity and share the same interest. Appraisal bias in this paper refers to the actions taken by lenders and appraisers that intentionally overvalue the collateral value to facilitate the successful completion of a mortgage transaction. Origination

²³ Two standard deviation away from the population mean on a normal distribution.

²⁴ The standard mortgage insurance coverage is 12, 25, 30 and 35 percent for loans with LTV 85 percent and under, 85.01 to 90 percent, 90.01 to 95 percent, 95.01 to 100 percent respectively for 30 year, standard coverage, primary residence, purchase or rate term refinance, A rating credit loans. Source: www.radian.biz/rates.

appraisal bias arises due to the information asymmetry between mortgage investors and mortgage originators²⁵. Mortgage originators, whose compensation is contingent upon business volume, have an incentive to influence appraisers to provide favorable valuations to the mortgage investors. Mortgage investors and secondary mortgage market agents do not have access to the individual collateral values independently and lack the capability to verify the lender reported property value. This structure creates incentive for lenders to behave in ways that contract to the best interest of the mortgage investors.

For example, holding everything else constant, a higher property value lowers the interest rate for mortgage borrowers, and mortgage borrowers are more inclined to accept an offer with a lower interest rate. Therefore, mortgage lenders are willing to report a higher than actual property value to mortgage buyers in the secondary market in returns for increasing business provided the inflated value is not detectable. The higher than actual property value is realized through lender ordered appraisals where the appraisers intentionally over appraise the underlying collateral value. This systematic fraudulent action is referred to as origination appraisal bias.

The existence of appraisal bias is solely due to the structure between primary mortgage market and secondary mortgage market. Due to the investment nature of mortgages and the popularity of mortgage derivatives, i.e., mortgaged backed securities (MBS), mortgage lenders typically do not hold mortgages in their portfolios. Instead, they sell them to secondary mortgage market in return for MBS, which can be traded frequently in the capital market. MBSs, depending on who the issuers are, carry guaranties on the

²⁵ Please refer to chapter 2 on the theoretical derivation of appraisal bias.

credit risk of the underlying mortgages. Investors on the capital market are free of default risk if they purchase MBSs issued by certain institutions, e.g., Fannie Mae and Freddie Mac and Ginnie Mae. Those institutions guarantee the timely payment of mortgage principals and interest. In this scenario, it is the MBS issuers, not the mortgage lenders who bare the credit risk of the mortgages. This gives mortgage lenders incentive to behave in ways that maximize their profit at the expense of MBS issuers.

Because appraisals are used to estimate borrower's equity in the absence of a transaction price and equity influence default, over-valuation of collaterals results in underestimation of the default risk. By passing the inflated default risk to secondary mortgage market agents, mortgage lenders have strong motivations to maintain a system of appraisal incentives.

Empirical and experimental/behavioral research has documented the existence of appraisal bias. Empirical wise, Cho and Megbolugbe (1996) and Chinloy, Cho and Megbolugbe (1997) find out that over 95 percent of the appraised values on a sample of 600, 000 residential mortgages purchased by Fannie Mae exceed that of the property pending sale price. Noordewier, Harrion, and Ramagopal (2001) study 1,428 residential mortgages and conclude that mortgages whose collateral values are over-valued, compared to their comparable value, exhibit higher default risk than those whose collaterals do not have overvaluation.

Experimental/behavior research conducted through survey and experimental design present findings that reinstate the existence of appraisal bias. Survey and experiment conducted by Kinnard, Lenk and Worzala (1997) reveal that appraisers acknowledge they are under pervasive client pressure and have a tendency to succumb to it when exerted by important clients. Another study finds that unsophisticated pressures from lenders are applied to appraisers in the forms of threat of withholding fee payments or reduction in future assignments (Levy and Schuck 1999). Another experiment shows an asymmetric response to transaction price feedback among appraisers (Hansz and Diaz 2001). Appraisers do not adjust downward their estimates when told their previous estimate on an unrelated property is too high. However, when told that their value estimates on a previous appraisal was too low regardless if it is so, appraisers respond by adjusting upward their next valuation on an unrelated assignment. Hansz and Diaz interpret this asymmetry as a routine response to pervasive appraiser-lender concerns.

Theoretical derivation on appraisal bias can be found in chapter 2. Chapter 2 concludes that under certain conditions, it is an equilibrium strategy for lenders to randomize between truthfully reports the property value and intentionally inflates the property value. It is also an equilibrium strategy for mortgage buyers to randomize between takes the reported value as such and costly investigate the soundness of the reported property value.

There has not been any study on mortgage termination that explicitly corrects for origination appraisal bias when calculating the put and call option at loan level²⁶.

Ignoring origination appraisal bias leads to mis-measurement of equity related variables and biased estimators. In order to correct for appraisal bias, one needs to know the true market value of the property instead of relying on lender reported value. Unfortunately, limited data on property value prevents previous studies to address this concern.

Fortunately, this study utilizes a unique set of dataset where all sample observations have an earlier purchase transaction. As mentioned earlier, one can match each refinance loan record with an earlier purchase transaction where the purchase price of the property at that time is available. In addition, we have access to a quarterly weighted repeated sales house price index separately for 100,000 zip code levels provided by the same institution that also provides the estimation dataset. This index provides an estimate of the course of house prices in each zip code area and each quarter since loan originations. Combining the purchase price with the zip code level quarterly weighted repeated sales house price index one can derive the calculated property value at the second transaction's origination time. The difference between the calculated and the lender reported property value at second loan's origination time reflects the appraisal bias. The appraisal bias for purchase transactions is eliminated because the purchase price is assumed to be determined competitively by demand and supply in the market.

²⁶ Ciochetti et al. (2002) studies lender level aggregated originator bias in commercial mortgage. They define originator bias as the non-representative underwriting behavior (too loose or too stringent) by each lender. They correct it by weighing each sample observation based on how representative that observation type is relative to its population equivalent.

The methodology that creates the house price index (HPI) is based on the weighted repeated sales (WRS) 3-stage estimation procedure proposed by Bailey, Muth and Nourse (1963) and implemented by Case and Shiller (1987; 1989)²⁷. To validate the appropriateness of their methodology, I compute predicted property value using Case and Shiller's WRS index on purchase loans and refinance loans. Then I compare the predicted property value with the actual purchase transaction price and the lender-reported refinance property price. My comparison suggests that the WRS model over predicts the purchase price by 0.5 percent and under predicts the refinance property value by 2 percent on average²⁸. This indicates that Case and Shiller's methodology does well in predicting actual purchase transaction prices. The under prediction of the refinance property values suggests that there exists appraisal bias in the refinance transactions.

The above measure is still at most a proxy for appraisal bias because researchers do not observe the actual market value of the property. However, it at least eases the concern of appraisal bias at the mean level. I limit the observation to loans that have prior purchase transactions records only because in theory the purchase price should bare no bias in any direction. On the other hand, a prior refinance transaction might have been subjected to appraisal bias and hence the estimated current property value based on an inflated prior property value may lead to underestimated appraisal bias.

²⁷ Please refer to Appendix B for the documentation on weighted repeat sale house price index.

²⁸ The average actual purchase price is \$179,580.51 with a standard deviation of 72,138.96, and the average refinance price is \$173,317.81 with a standard deviation of 71,940.72.

3.3.3 Explanatory Variables

As explained earlier, the borrower treats prepayment option (call option) as an option to pay off the unpaid balance prior to maturity, where she “buys” the mortgage from the mortgage lender for the amount of the remaining balance. In the interpretation of default option (put option), the borrower “sells” the property back to the mortgage lender at the property market value at the time of default.

Following literature on option pricing model, the variable influencing the exercise of the call option is the gap between the market value of the mortgage and the face value of the mortgage. This variable can be computed by the ratio of the present discounted value of the unpaid mortgage balance at the current market rate relative to the value discounted at the contract interest rate (coupon rate)²⁹.

The variable influencing the exercise of the put option is the amount of negative equity associated with the property at the time of default. It can be measured by the equity ratio between contemporaneous property value and market value of the mortgage.

Unfortunately, we do not observe the exact market property value for each individual house. Once again, we rely on the zip code level house price index to measure the house price appreciation in each quarter. Applying the estimates of the zip code level house price index to the origination house value gives us the quarterly market house value. The

²⁹ The current quarterly market interest rate used here is the average interest rate charged by lenders reported by Freddie Mac’s quarterly market survey (<http://www.freddiemac.com>). This rate varies by quarter across five major US regions.

ratio between the market house value and the mortgage market value gives us proxy for put option value³⁰.

This measure is by far an approximation for the put option value that a borrower endows in each period. An apparent improvement is one that considers the tax benefit that the interest payment on a mortgage brings to the borrower. The interest payment on a mortgage is tax deductible and hence reduces the put option value to the borrower. However, lack on information on the household size prevents me to locate the county, state and federal specific marginal tax bracket that should be applied to each borrower. One the other hand, one also needs to subtract property tax from the interest payment tax benefit. The property tax is based on county level property tax rate that I do not obtain. Arbitrary assumptions on the amount of the tax benefit and tax payment does not make the put option measurement more accurate than what it is. Therefore, I assume the tax benefit and tax payment cancel out each other and hence the exclusion of both does not impose a threat to the estimation. The put option value measures the probability that the equity value falls below market value and should be considered a probability measure.

The econometric model presented in the next section uses variables that can be categorized as mortgage and property characteristics, borrower characteristics, and economic conditions as covariates to explain default and prepayment behavior. Based on those variables, I construct the put and call option value, the transaction cost, and the trigger events that determine default and prepayment behavior.

³⁰ Please refer to Appendix A for construction of the call and put option variable.

Mortgage and property characteristics:

Mortgage Size – Mortgage size is included to control for differences in the fixed-cost component of transaction costs associated with default and prepayment. In practice, the fixed-cost component of the transaction costs is proportionately smaller for large mortgages than for small mortgages. Therefore, we expect to see borrowers with large mortgages are more likely to invoke the default or prepayment options. On the other hand, the size of the mortgage also indicates the financial affordability of the borrowers. The fact that a borrower chooses to borrow a large size mortgage demonstrates her financial confidence and stability, which also reflects a low default probability. And we should expect borrowers who are less financially stable, as represented by a small loan amount, are more likely to default. The value of the natural logarithm of mortgage size at origination is used in the estimation³¹.

Put and Call Option Value – The put and call option values are constructed using loan age, monthly principal and interest payment, mortgage note rate, current local market interest rate, mortgage term, property value at origination and market value of the property. One expects the borrowers with high put option values are more likely to default and borrowers with high call option values are more likely to prepay. Since the two are correlated, we are also interested in how the put and call variables influence the other option.

³¹ $\log(m)$ is used in the duration model as one of the covariates, where m is the mortgage amount at origination.

Borrower Credit Score at Origination – Borrower credit score serves as measures for borrower’s capability to manage debt. A historically capable borrower is better positioned to manage her credit risk and to avoid default than an incapable one.

Borrower Debt-to-Income Ratio – Debt-to-income ratio is negatively correlated with prepayment and positively correlated with default. It is hard for borrowers with high debt-to-income ratio to find lenders who are willing to refinance their current mortgage due to their financial constraints. For the same reason, borrowers with high debt-to-income ratio are likely to exercise default option for there is little financial cushion.

Economic conditions:

State Level Quarterly Unemployment Rate – Unemployment rate serves as proxy for trigger event. One expects that properties in high unemployment regions are more likely to default than those that are in low unemployment regions. Previous studies also suggest that high unemployment rates lower the prepayment hazard, indicating that liquidity constraints make prepayment difficult for unemployed borrowers (Deng, Quigley and Van Order 1996).

3.3.4 Summary Statistics of Explanatory Variables

Table I presents the means and standard deviations of the time invariant variables by termination reasons. The summary statistics shows that origination LTV, borrower credit score, borrower monthly income and debt to income ratio all move in the direction that literature suggested. Defaulted loans have high origination LTV, low credit scores, low

monthly income, and high debt to income ratios at origination. In comparison, loans that have prepaid during the 48 quarters observation window tend to have high loan amount, high credit score, high monthly income and low debt-to-income ratio. In addition, defaulted loans also have higher appraisal bias than active loans or prepaid loans as of December 2003. Whether or not these variables possess significant explanatory power to loan terminations remain to be tested.

Table I. Descriptive Statistics of Time Invariant Variables by Loan Termination Type

Variables	Total	As of 2003/12		
		Active	Prepaid	Defaulted
Original Loan Amount (\$)	124,034.63 (47,777.49)	108,123.95 (44,287.54)	127,629.48 (47,805.67)	118,181.00 (46,718.68)
Original Loan to Value Ratio	0.74 (0.14)	0.71 (0.15)	0.75 (0.14)	0.84 (0.90)
Original Note Rate	0.0715 (0.0057)	0.0687 (0.0050)	0.0722 (0.0057)	0.0750 (0.0065)
Borrower FICO	718.95 (50.67)	720.47 (53.57)	718.91 (49.80)	660.78 (54.40)
Monthly Income (\$)	5,933.56 (3,571.54)	5,568.67 (3,401.16)	6,020.13 (3,606.31)	4,992.14 (2,796.04)
Monthly Debt (\$)	1,680.44 (842.03)	1,511.47 (800.72)	1,718.39 (846.37)	1,663.18 (846.91)
Debt to Income Ratio	0.3073 (0.1190)	0.2967 (0.1228)	0.3094 (0.1180)	0.3515 (0.1166)
Appraisal Bias (Predicted LTV - Reported LTV)	0.0231 (0.0992)	0.0152 (0.1031)	0.0247 (0.0976)	0.0911 (0.1588)
Appraisal Bias (Reported Value - Predicted Value)/Predicted Value	0.0341 (0.1589)	0.0240 (0.1633)	0.0362 (0.1573)	0.1175 (0.2074)

Notes: Standard deviations are in parenthesis. All loans originated from 1992 to 2000. Values are reported as of the time of loan origination. An active loan is one that had not yet terminated by 12/31/2003. Based on 508,219 observations.

The following two tables present summary statistics of the time variant variables, one by loan age, the other by calendar time.

Table II presents means and standard deviation of time variant variables by duration/loan age. The data shows that the put option monotonically decreases as the loans age, with the biggest decrease occurring from origination to the end of first year. The decreasing trend then flattens to approach zero from 3rd to 5th year, and returns to its original speed from 5th to 7th year and slows down again to stay relatively flat after a loan reaches 7 years of age. This indicates that default is more likely to occur after a loan reaches its 5 years of age. The call option fluctuates as loan ages, indicating that its value is less correlated with the age of the loan. This result is not surprising because call option is defined by a loan's origination interest rate and what is being offered in the market quarterly. The decreasing trend observed in house price reflects the fact that the surviving pool consists more low value property loans than high value property loans.

Table II. Descriptive Statistics for Time Variant Variables By Loan Duration

Variables	At Origination	1Yr	3Yr	5Yr	7Yr	9Yr
House Price	172,066.69 (72,740.74)	171,102.72 (72,298.20)	169,558.04 (71,872.93)	161,033.89 (69,661.33)	155,580.68 (70,729.64)	158,383.18 (72,897.68)
Call Option	0.0089 (0.0423)	-0.0265 (0.0802)	0.0063 (0.0602)	0.0603 (0.0880)	0.0081 (0.0878)	0.0259 (0.0796)
Put Option	0.1075 (0.1346)	0.0615 (0.1063)	0.0297 (0.0682)	0.0218 (0.0556)	0.0072 (0.0309)	0.0024 (0.0144)
Squared Term of Call Option	0.0019 (0.0030)	0.0071 (0.0072)	0.0037 (0.0109)	0.0114 (0.0154)	0.0078 (0.0163)	0.0070 (0.0087)
Squared Term of Put Option	0.0297 (0.0618)	0.0151 (0.0467)	0.0055 (0.0277)	0.0036 (0.0179)	0.0010 (0.0102)	0.0002 (0.0036)
State Unemployment Rate	4.6106 (1.1795)	4.3506 (1.0266)	5.0027 (1.0603)	5.4610 (1.2222)	4.7471 (1.2329)	5.7703 (0.9222)

Notes: Standard deviations are in parenthesis. All loans originated from 1992 to 2000. Values are reported by loan age. Based on 508,219 loan observations

Table III presents summary statistics of time variant variables by calendar time. It shows that the call option value is increasing over the year except for a dip in 1999Q4,

indicating that prepayment activities shall be clustered before and after 1999Q4. The value of call option is the lowest in 1999Q4, indicating that the prepayment option is least likely to be exercised. State level quarterly unemployment rate decreases from 1992Q4 to 1999Q4 and reaches its lowest level in 1999Q4, after which it then starts to climb back up again. Values of put options fluctuate by calendar time, its absolute value is less informative since it is reflecting a mixture of loans of different ages. The house price reflects the general market appreciation trend occurred by calendar time. Once again, the dip in 2003Q4 is resulting from the low property loans left in the surviving pool.

Table III. Descriptive Statistics for Time Variant Variables By Calendar Time

Variables	1993Q4	1995Q4	1997Q4	1999Q4	2001Q4	2003Q4
House Price	158,727.60 (73,639.30)	157,125.43 (70,645.67)	160,424.41 (68,139.59)	170,899.53 (71,713.89)	172,997.26 (73,289.79)	158,304.79 (70,550.63)
Call Option	0.0269 (0.0532)	0.0084 (0.0556)	0.0304 (0.0576)	-0.0595 (0.0383)	0.0352 (0.0608)	0.0453 (0.0917)
Put Option	0.0546 (0.1103)	0.0626 (0.1308)	0.0880 (0.1370)	0.0343 (0.0672)	0.0384 (0.0843)	0.0193 (0.0559)
Squared Term of Call Option	0.0036 (0.0054)	0.0032 (0.0052)	0.0042 (0.0059)	0.0050 (0.0040)	0.0049 (0.0080)	0.0105 (0.0123)
Squared Term of Put Option	0.0151 (0.0506)	0.0210 (0.0696)	0.0265 (0.0673)	0.0057 (0.0223)	0.0086 (0.0363)	0.0035 (0.0210)
State Unemployment Rate	6.4108 (1.4127)	5.3717 (1.2180)	4.5971 (0.9873)	4.0484 (0.7753)	5.5298 (0.8311)	5.9265 (0.9790)

Notes: Standard deviations are in parenthesis. All loans originated from 1992 to 2000. Values are reported by loan age. Based on 508,219 loan observations

3.4 Model

In this section, I present the derivation and specification of the competing risk proportional hazard model with unobserved heterogeneity.

Let T_d, T_p be the latent variables representing the duration of a mortgage until it is terminated in the form of default or prepayment respectively. Denote $T = \min(T_d, T_p)$ as

the observable duration, where only the risk that occurs earlier will be observed.

Therefore, it is known that an observation's default duration is longer than its prepayment duration if a prepayment is observed. Likewise, an observation's prepayment duration is longer than its default duration if a default is observed. If an observation remains active by the end of the 48th quarter, we conclude that both its default and prepayment duration exceed the length of the maximum observation window.

Let z denote a vector of time invariant covariates and x_t denote a vector of time varying covariates that is fixed between $[t-1, t)$, but can vary from period to period. Both variables influence the termination decision for the borrower at each period. Furthermore let $\theta = (\theta_d, \theta_p)$ be the two unobserved heterogeneity factors associated with default and prepayment risks respectively.

Competing risk model under proportional hazard specifications has the following assumptions (Han and Hausman 1990; Sueyoshi 1992; McCall 1996; An 2004; Wooldridge):

Assumption 1 (Conditional Independence) Conditional on the observed and unobserved heterogeneity (x_t, z, θ) , the two risk-specific durations T_d, T_p are independent.

Assumption 2 (Proportional Hazards) Conditional on $(x_1, \dots, x_t, z, \theta)$, the hazard function for T_d, T_p is $h_{ji}(t | x_1, \dots, x_t, z, \theta) = h_{0j}(t) \exp(x_{it} \beta_j + z_i \gamma_j + \theta_{ij})$, $j = d, p$ respectively.

The hazard at time t only depends on the covariates at time t .

McCall (1994) specifies the heterogeneity as multiplicative to the proportional hazards, specifically in the form of $h(t | z(s), s \in [0, t], \theta) = h_0(t) \exp(z(t)\beta)\theta$. McCall's specification and the one in assumption 2 are equivalent where his heterogeneity equals the exponential of what is specified in assumption 2. The advantage of the latter is that by putting the heterogeneity parameter inside the exponential function, we are guaranteed a positive value.

For each observation i , researchers can observe two covariate vectors x_{it}, z_i , and an interval A_i such that $T_i \in A_i$, where A_i is either $[k-1, k)$ or $[k-1, \infty)$, $k = 1, \dots, \infty$.

Researchers also observe two censoring indicator, δ_{pi}, δ_{di} . The value of δ_{pi} equals one for prepayment and zero for default; the value of δ_{di} equals one for default and zero for prepayment for observation i .

Combining assumption 1, 2 and utilizing the relationship between probability density

function, survivor function and hazard function, i.e., $h(\cdot) = \frac{f(\cdot)}{S(\cdot)}$, the joint density

function of (T_d, T_p) conditional on (x_t, z, θ) can be expressed as:

$$\begin{aligned} f(u, v | x_{it}, z_i, \theta_i) &= f_d(u | x_{it}, z_i, \theta_i) \cdot f_p(v | x_{it}, z_i, \theta_i) \\ &= h_d(u | x_{it}, z_i, \theta_i) \cdot h_p(v | x_{it}, z_i, \theta_i) \cdot \exp\left\{-\int_{s=0}^u h_d(s | x_{is}, z_i, \theta_i) ds - \int_{s=0}^v h_p(s | x_{is}, z_i, \theta_i) ds\right\} \end{aligned}$$

The joint survivor function is

$$\begin{aligned}
S(u, v | x_{it}, z_i) &= E_{\Theta} \{ \Pr(T_d \geq u, T_p \geq v | x_{it}, z_i) \} \\
&= E_{\Theta} \left\{ \exp \left\{ - \int_{s=0}^u h_{0d}(s) \cdot \exp(x_{is} \beta_d + z_i \gamma_d + \theta_{di}) ds - \int_{s=0}^v h_{0p}(s) \cdot \exp(x_{is} \beta_p + z_i \gamma_p + \theta_{pi}) ds \right\} \right\}
\end{aligned}$$

where E_{Θ} denotes the expectation operator with respect to the distribution of the heterogeneity, which is assumed to be independent from x_i, z_i .

Denote the risk specific integrated baseline hazard and time varying covariate

$$\int_{s=0}^u h_{0j}(s) \cdot \exp(x_{is} \beta_j) ds \text{ as } \Lambda_{ji}(u), j = d, p \text{ and the time invariant covariate}$$

$\exp(z_i \gamma_j + \theta_{ji})$ as $\phi_{ji}, j = d, p$. After simplification, the joint survivor function can be written as

$$S(u, v | x_{it}, z_i) = E_{\Theta} \{ \exp \{ -\Lambda_{di}(u) \cdot \phi_{di} - \Lambda_{pi}(v) \cdot \phi_{pi} \} \}$$

An observation can be classified in the following three types:

Type I. A default subject is observed with $T_i \in [k-1, k)$ for $0 \leq k \leq K$ and $\delta_{di} = 1$,

Type II. A prepayment subject is observed with $T_i \in [k-1, k)$ for $0 \leq k \leq K$ and $\delta_{pi} = 1$,

Type III. A subject is observed with $T_i \in [k-1, \infty)$ for $0 \leq k \leq K$ and $\delta_{di} = 0, \delta_{pi} = 0$.

A Type III observation contributes to the sample likelihood function in the following form:

(1)

$$\Pr(T_i \geq k-1, \delta_{di} = 0, \delta_{pi} = 0 | x_{it}, z_i) = E_{\Theta}[\Pr(T_i \geq k-1 | x_{it}, z_i)] = E_{\Theta}[S(k-1, k-1 | x_{it}, z_i)]$$

Because this observation has survived the end of the observation window, its contribution to the likelihood is its joint survivor function at $k-1$.

The contribution to the sample likelihood of a Type I observation is:

$$\begin{aligned} & \Pr(k-1 \leq T_i < k, \delta_{di} = 1 | x_{it}, z_i) \\ (2) \quad & = E_{\Theta}[\Pr(k-1 \leq T_i < k, \delta_{di} = 1 | x_{it}, z_i)] \\ & = E_{\Theta}\left\{ \int_{t=k-1}^k h_{di}(t) \cdot \exp\{-\Lambda_{di}(t) \cdot \phi_{di} - \Lambda_{pi}(t) \cdot \phi_{pi}\} dt \right\} \end{aligned}$$

The above expression is the probability of an observation surviving up to period $k-1$, multiplied by its default hazard rate between period $k-1$ and k . It represents the area size where default occurs between $k-1$ and k , and prepayment risk occurs after default.

To gain intuition, the above integral can be thought of as

$$\int_{t=k-1}^k f_d(t) \cdot S_p(t) dt, \text{ where } f_d(t) \text{ measures the density that default occurs at time } t \text{ and}$$

$S_p(t)$ represents the likelihood that prepayment occurs after period t . Because

$$\begin{aligned} \Pr(k-1 \leq T_d \leq k, T_p > T_d) &= \int_{t=k-1}^k \int_t^{\infty} f_d(t) \cdot f_p(s) ds dt = \int_{t=k-1}^k [f_d(t) \int_t^{\infty} f_p(s) ds] dt \\ &= \int_{t=k-1}^k f_d(t) \cdot S_p(t) dt \end{aligned}$$

The contribution to the sample likelihood of a Type II observation is:

(3)

$$\begin{aligned}
& \Pr(k-1 \leq T_i < k, \delta_{pi} = 1 \mid x_{it}, z_i) \\
&= E_{\Theta}[\Pr(k-1 \leq T_i < k, \delta_{pi} = 1 \mid x_{it}, z_i)] \\
&= E_{\Theta} \left\{ \int_{t=k-1}^k h_{pi}(t) \cdot \exp\{-\Lambda_{di}(t) \cdot \phi_{di} - \Lambda_{pi}(t) \cdot \phi_{pi}\} dt \right\}
\end{aligned}$$

This is simply the probability of an observation surviving up to period $k-1$, multiplied by its prepayment hazard rate between period $k-1$ and k . It represents the area size where prepayment occurs between $k-1$ and k , and default occurs after prepayment.

Notice that unlike in the single risk grouped data model, where the likelihood depends on the baseline hazard only through the discrete values of integrated hazard function evaluated at limited number of grouping points, the above integrals depend on the values of $h_{0d}(t)$ and $h_{0p}(t)$ for all t between the interval $k-1$ and k . This is because in single risk model, the likelihood of an un-censored observation is reduced to

$$\int_{t=k-1}^k \{h_i(t) \cdot \exp(-\Lambda_i(t) \cdot \phi_i)\} dt .$$

$$\text{Since } \Pr(k-1 \leq T < k) = \int_{t=k-1}^k f(t) dt = F(k) - F(k-1) = S(k-1) - S(k) ,$$

the above integral can also be expressed as $S(k-1) - S(k)$.

In the absence of time varying covariates, the proportional contribution to the hazard function can be taken out of the integral and the integrand is then left with only the baseline hazard. Therefore, the functional form of the baseline is irrelevant; its integrated value at discrete number of grouping points is what needs to be estimated. Provided the

number of grouping points is less than the sample size, consistent estimates can be achieved.

In the competing risk framework with grouped data one needs parameterization of the baseline hazard to solve the exact solution of the integrals even without the presence of

time varying covariates. The integral can be thought of as $\int_{s=k-1}^k f(x) \exp(-F(x) - G(x)) ds$.

One cannot solve for it without the knowledge of $f(x)$ and $g(x)$. An (2004) provides proof of the unidentifiable problem for competing risks model under proportional hazard specification with grouped duration data.

However, exact solution can be achieved under certain assumptions on the baseline hazard function. Following Han and Hausman (1990) and An (2004), a piece-wise constant assumption is made on the baseline hazard.

Assumption 3 (Piece-wise Constant Baseline Hazard) the baseline hazard function is piece-wise constant:

$$\begin{aligned} h_{0j}(t) &= \alpha_{j,k}, \\ t &\in [k-1, k), \\ j &= d, p \end{aligned}$$

Under assumption 3, the two integrated baseline hazard functions $\int_0^t h_{0j}(s)ds, j = d, p$ are piece-wise linear with interval-specific slopes α_{jk} . With this assumption, the integrals in

(2) and (3) have analytical expression and become

(2)'

$$\begin{aligned}
& E_{\Theta} \left\{ \int_{t=k-1}^k h_{di}(t) \cdot \exp\{-\Lambda_{di}(t) \cdot \phi_{di} - \Lambda_{pi}(t) \cdot \phi_{pi}\} dt \right\} \\
&= E_{\Theta} \left\{ \int_{t=k-1}^k \right. \\
&= h_{di}(k) \cdot \exp\{-\Lambda_{di}(k-1) \cdot \phi_{di} - \alpha_{dk}(t-(k-1)) \cdot \exp(x_{it} \beta_d) \cdot \phi_{di} \\
&\quad \left. - \Lambda_{pi}(k-1) \cdot \phi_{pi} - \alpha_{pk}(t-(k-1)) \cdot \exp(x_{it} \beta_p) \cdot \phi_{pi}\} dt \right\} \\
&= E_{\Theta} \left\{ \frac{\alpha_{dk} \exp(x_{ik} \beta_d) \cdot \phi_{di}}{\alpha_{dk} \exp(x_{ik} \beta_d) \cdot \phi_{di} + \alpha_{pk} \exp(x_{ik} \beta_p) \cdot \phi_{pi}} \cdot \right. \\
&\quad \left. \exp\left\{-\sum_{s=0}^{k-1} \{\alpha_{ds} \exp(x_{is} \beta_d) \cdot \phi_{ds} + \alpha_{ps} \exp(x_{is} \beta_p) \cdot \phi_{ps}\} \cdot \right. \right. \\
&\quad \left. \left. \{1 - \exp\{-\alpha_{dk} \exp(x_{ik} \beta_d) \cdot \phi_{di} - \alpha_{pk} \exp(x_{ik} \beta_p) \cdot \phi_{pi}\}\} \right\} \right\}
\end{aligned}$$

(3)'

$$\begin{aligned}
& E_{\Theta} \left\{ \int_{t=k-1}^k h_{pi}(t) \cdot \exp\{-\Lambda_{di}(t) \cdot \phi_{di} - \Lambda_{pi}(t) \cdot \phi_{pi}\} dt \right\} \\
&= E_{\Theta} \left\{ \int_{t=k-1}^k h_{pi}(k) \cdot \exp\{-\Lambda_{di}(k-1) \cdot \phi_{di} - \alpha_{dk}(t-(k-1)) \cdot \exp(x_{it} \beta_d) \cdot \phi_{di} \right. \\
&\quad \left. - \Lambda_{pi}(k-1) \cdot \phi_{pi} - \alpha_{pk}(t-(k-1)) \cdot \exp(x_{it} \beta_p) \cdot \phi_{pi}\} dt \right\} \\
&= E_{\Theta} \left\{ \frac{\alpha_{pk} \exp(x_{ik} \beta_p) \cdot \phi_{pi}}{\alpha_{dk} \exp(x_{ik} \beta_d) \cdot \phi_{di} + \alpha_{pk} \exp(x_{ik} \beta_p) \cdot \phi_{pi}} \cdot \right. \\
&\quad \left. \exp\left\{-\sum_{s=0}^{k-1} \{\alpha_{ds} \exp(x_{is} \beta_d) \cdot \phi_{di} + \alpha_{ps} \exp(x_{is} \beta_p) \cdot \phi_{pi}\} \cdot \right. \right. \\
&\quad \left. \left. \{1 - \exp\{-\alpha_{dk} \exp(x_{ik} \beta_d) \cdot \phi_{di} - \alpha_{pk} \exp(x_{ik} \beta_p) \cdot \phi_{pi}\}\} \right\} \right\}
\end{aligned}$$

Denote $\omega = \frac{\alpha_{pk} \exp(x_{ik} \beta_p) \cdot \phi_{pi}}{\alpha_{dk} \exp(x_{ik} \beta_d) \cdot \phi_{di} + \alpha_{pk} \exp(x_{ik} \beta_p) \cdot \phi_{pi}}$ and

$$1 - \omega = \frac{\alpha_{dk} \exp(x_{ik} \beta_d) \cdot \phi_{di}}{\alpha_{dk} \exp(x_{ik} \beta_d) \cdot \phi_{di} + \alpha_{pk} \exp(x_{ik} \beta_p) \cdot \phi_{pi}}$$

And denote

$$\begin{aligned}
\Lambda(k-1) &= \exp\left\{-\sum_{s=0}^{k-1} \{\alpha_{ds} \exp(x_{is} \beta_d) \cdot \phi_{di} + \alpha_{ps} \exp(x_{is} \beta_p) \cdot \phi_{pi}\} \cdot \right. \\
&\quad \left. \{1 - \exp\{-\alpha_{dk} \exp(x_{ik} \beta_d) \cdot \phi_{di} - \alpha_{pk} \exp(x_{ik} \beta_p) \cdot \phi_{pi}\}\} \right\}
\end{aligned}$$

Since the probably of the duration ends in interval $[k-1, k)$ given x_t, z, θ is the sum of the probability of termination by default and probability of termination by prepayment, i.e.,

$$\begin{aligned}
& \Pr(k-1 \leq T_i < k \mid x_{it}, z_i, \theta_i) \\
&= \Pr(k-1 \leq T_i < k, \delta_{di} = 1 \mid x_{it}, z_i, \theta_i) + \Pr(k-1 \leq T_i < k, \delta_{pi} = 1 \mid x_{it}, z_i, \theta_i) \\
&= \omega \cdot \Lambda(k-1) + (1-\omega) \cdot \Lambda(k-1) \\
&= \Lambda(k-1)
\end{aligned}$$

It is clear that one needs to assign weights ω and $1 - \omega$ to default and prepayment risks respectively when calculating the contribution to the likelihood function. McCall (1996) arbitrarily assigns $\omega = 0.5$ rather than estimating ω and studies by Deng et al (2000), Ciochetti et al (2001), Amborse and LaCour-Little (2001), and Huang and Ondrich (2002) all adopt his formula in contrast to their piecewise constant baseline hazard assumptions. As illustrated, the exact solution of the likelihood is achievable once the baseline hazards are assumed to be piecewise constant, hence one should not approximate ω ; one should rather estimate it using the data.

The unobserved heterogeneity is accounted for by the Heckman-Singer nonparametric approach. Heckman-Singer heterogeneity distribution is a discrete distribution with M points (Heckman and Singer 1984). The joint distribution of (θ_d, θ_p) is modeled by assuming that there are $m = 1, \dots, M$ unobserved types of individuals that occur in the population with relative frequency p_m (an individual in group m is characterized by the doublet of location parameters $(\theta_{dm}, \theta_{pm})$). Here, I assume a two-by-two discrete support for the heterogeneity in the model.

Specifically, I assume there are two types of individuals for each risk; their joint distribution and associated probabilities are presented in the following table.

θ_p & θ_d	θ_d^1	θ_d^2	Marginal Probability of θ_d
-------------------------	--------------	--------------	------------------------------------

θ_p^1	p_{11}	p_{12}	$p_{11} + p_{12}$
θ_p^2	p_{21}	p_{22}	$p_{21} + p_{22}$
Marginal Probability of θ_d	$p_{11} + p_{21}$	$p_{12} + p_{22}$	1

Two sets of conditions must be satisfied on the eight parameters for the heterogeneity.

1. The probabilities must sum to unity:

$$\sum_i \sum_j p_{ij} = 1$$

2. The heterogeneity is mean 1 and distributed independently of x_i, z . Due to the additive heterogeneity specification in this model, the mean 1 normalization is approximated by the following condition:

$$E[\theta_j] = 0, j = d, p$$

Condition 1 and the fact that all probabilities must range between 0 and 1 are met through the following transformation:

$$p_{11} = \frac{\exp(\rho_{11})}{1 + \exp(\rho_{11}) + \exp(\rho_{12}) + \exp(\rho_{21})},$$

$$p_{12} = \frac{\exp(\rho_{12})}{1 + \exp(\rho_{11}) + \exp(\rho_{12}) + \exp(\rho_{21})},$$

$$p_{21} = \frac{\exp(\rho_{21})}{1 + \exp(\rho_{11}) + \exp(\rho_{12}) + \exp(\rho_{21})},$$

$$p_{22} = \frac{1}{1 + \exp(\rho_{11}) + \exp(\rho_{12}) + \exp(\rho_{21})}$$

Condition 2 requires that

$$\theta_d^2 = -\frac{p_{11} + p_{21}}{p_{12} + p_{22}} \cdot \theta_d^1,$$

$$\theta_p^2 = -\frac{p_{11} + p_{12}}{p_{21} + p_{22}} \cdot \theta_p^1$$

Therefore, only five out of the eight heterogeneity parameters are free to be estimated.

They are

$$(\rho_{11}, \rho_{12}, \rho_{21}, \theta_d^1, \theta_p^1)$$

Deng et al. (2000a, 2000b) and Huang and Ondrich (2002) also adopt the Heckman-Singer mass point heterogeneity but in a very restrictive form. Both of their specifications assume the same probability for default and prepayment heterogeneity. Specifically, they assume three types of individuals: (θ_d^1, θ_p^1) with probability p_1 , (θ_d^2, θ_p^2) with probability p_2 and (θ_d^3, θ_p^3) with p_3 , where $p_1 + p_2 + p_3 = 1$. Hence their methods restrict θ_d, θ_p to be dependent and positively correlated. Translated to mortgage terms, their specifications restrict that high prepayment heterogeneity must be associated with high default heterogeneity, which contradicts to empirical evidence. Empirical studies show that a borrower who is more risky to prepay is less prone to default and they have smaller default probability had they not prepaid compare to the average at risk population. My specification relaxes the dependence between default and prepayment heterogeneity and allows them to move freely.

After specifying the heterogeneity, one can define the unconditional probability contribution to the likelihood function corresponding to the three types of observations as $l_c(k), l_d(k), l_p(k)$. They are respectively:

(4)

$$l_c(k) = \Pr(T_i \geq k-1, \delta_{di} = 0, \delta_{pi} = 0) = E_{\Theta}[S(k-1, k | \theta)] = \sum_{i=1}^2 \sum_{j=1}^2 p_{ij} S(k-1, k-1 | \theta_{ij})$$

(5)

$$l_d(k) = \Pr(k-1 \leq T_i < k, \delta_{di} = 1) = E_{\Theta}[\Pr(k-1 \leq T_i < k, \delta_{di} = 1 | \theta)] = \sum_{i=1}^2 \sum_{j=1}^2 p_{ij} \omega_{ij} \cdot \Lambda(k-1 | \theta_{ij})$$

(6)

$$l_p(k) = \Pr(k-1 \leq T_i < k, \delta_{pi} = 1) = E_{\Theta}[\Pr(k-1 \leq T_i < k, \delta_{pi} = 1 | \theta)] = \sum_{i=1}^2 \sum_{j=1}^2 p_{ij} (1 - \omega_{ij}) \cdot \Lambda(k-1 | \theta_{ij})$$

The log likelihood function of the competing risks model over all observations is

$$\log l = \sum_{i=1}^N \delta_{di} \log l_d(k_i) + \delta_{pi} \log l_p(k_i) + (1 - \delta_{di} - \delta_{pi}) \log l_c(k_i)$$

3.5 Estimation Results

3.5.1 Non-Ruthless Specifications

In this section, I present the estimation results of the non-ruthless model specification with controls for appraisal bias. For comparison, I also provide estimation results for the

ruthless model with controls for appraisal bias, and ruthless and non-ruthless models without controls for appraisal bias in appendix C.

All covariates in the estimations are between zero and unity or normalized to be within that range. Variables that have gone through normalization include credit score (divided by 1000), LTV (divided by 100), and logarithm of the origination loan value (divided by 10). For estimation purpose, the baseline hazard functions are assumed to be exponentials of the piecewise constant, specifically,

$$\begin{aligned}h_{0,j}(t) &= \exp(\alpha_{j,k}), \\t &\in [k-1, k), \\j &= d, p\end{aligned}$$

This specification allows the baseline hazard parameters to move inside the proportional changes; the only change stemming from it is the interpretation of α_{jk} . Specifically, α_{jk} now represents the natural logarithm of the baseline hazards. In addition, moving α_{jk} inside the proportional portion of the hazard function allows me to estimate an intercept within the proportional portion while getting rid of one degree of freedom from the baseline portion.

As indicated previously, the difference between ruthless and non-ruthless model assumptions is that the former assumes frictionless exercise of mortgage termination options where borrowers immediately exercise the default or prepayment option as soon as it is “in the money”. The latter one assumes that in addition to option values, there exist other deterministic factors that either trigger or hinder the exercise of mortgage terminations.

In order to test the two assumptions, I set up a ruthless and a non-ruthless model specification, where the former includes only the measures of financial incentives represented by the put and call options and the latter includes variables that measure borrower characteristics, property characteristics and trigger events in addition to the financial option variables in the ruthless specification. Both specifications include measures for appraisal bias.

The likelihood ratio test between the ruthless model and the non-ruthless model with appraisal bias returns a chi-square value of 6354.0 with degree of freedom 10, which favors the non-ruthless model specification. Table IV presents the maximum likelihood estimation of the competing risks proportional hazard model with grouped duration data for the non-ruthless model specification with controls for appraisal bias. For comparison purposes, estimation results of the ruthless model with controls for appraisal bias, ruthless and non-ruthless model without controls for appraisal bias are presented in Appendix C³².

Table IV. Maximum Likelihood Estimation of Competing Risks Proportional Hazard Model – Model 1: Non-Ruthless Model With Controls for Appraisal Bias

The parameters remain stable in different runs where I change the starting value of the maximum likelihood estimations. The parameters in all models are similar in their magnitudes and carry the same sign as those obtained from the conditional default and prepayment models employed by multinomial logit estimations.

	Prepayment	Default
baseline hazard parameters		
2 years of age	-0.7982 (0.2360)	-4.1952 (0.1899)
3 years of age	-0.6360 (0.2347)	-3.6671 (0.2186)
4 years of age	-0.4914 (0.2352)	-3.5411 (0.2502)
5 years of age	-0.1022 (0.2355)	-3.3920 (0.2798)
6 years of age	-0.1413 (0.2523)	-3.5903 (0.3494)
>6 years of age	-0.2464 (0.2514)	-2.9831 (0.3467)
Intercept	-8.7107 (1.0062)	4.1920 (2.5321)
Credit Score	0.3750 (0.5596)	-11.5939 (1.6428)
Debt to Income Ratio	0.0996 (0.2321)	1.9804 (0.5609)
Loan to Value Ratio	-0.1459 (0.2699)	4.9514 (0.8703)
Loan Amount	4.8957 (0.7943)	-6.2484 (2.0413)
Appraisal Bias	-0.4257 (0.5069)	8.8821 (1.3146)
Squared Term of Appraisal Bias	0.9237 (1.3914)	-8.6355 (2.4953)
Put Option	-1.2851 (0.6489)	9.2795 (1.5912)
Call Option	8.2279 (0.7078)	-1.9420 (1.2151)
Squared Term of Put Option	2.4291 (1.6010)	-8.5282 (2.6957)
Squared Term of Call Option	-5.0647 (3.0011)	2.3761 (5.0914)
State Unemployment Rate	0.0589 (0.0305)	0.2513 (0.0675)
Loglikelihood	-478,297.05	

Note: Standard errors reported in parentheses.

Risks are estimated jointly with maximum likelihood approach.

The parameters for the first year's baseline hazards are fixed so that two intercepts, one for prepayment and one for default, can be freely estimated inside the proportional hazards.

The baseline hazards for prepayment and default exhibit the following trend. The prepayment baseline hazard increases as loan ages and reaches its highest rate around a loan's fifth year of age then starts decreasing afterwards. The default baseline hazard reaches its first peak around the fifth year of a loan's age and then makes a sharp dip before it reaches its highest value after a loan has survived for at least seven years.

The put and call option variables move in the direction as the theory predicts. Instead of evaluating the signs of the parameters for the put and call options, 48 partial derivatives of the hazard functions with respect to put and call options at period 1 to period 48 evaluated at the sample mean value of each option at period 1 to period 48 are computed due to the existence of the squared option terms. The results show that across all periods, a higher put option value leads to a higher default hazard and similarly, a higher call option value causes the prepayment risk to increase. Setting the partial derivatives with respect to each option variable to zero solves for the turning point value, exceeding which the effect of each option on termination outcomes becomes negative. None of the call option value exceeds the turning point threshold, and only less than half of one percent of put option value exceeds the turning point threshold.

The estimation results also indicate that high probability of negative equity reduces the prepayment hazard across all periods, whereas the impact of call option on default hazard is insignificant. This indicates that a borrowers with high negative equity are less likely to prepay, either because mortgage lenders are less likely to extend credit to those deemed high risk of default or because their own financial unawareness. In a favorable interest rate market where market rate reduces significantly, such as in 1999-2000 and 2001-2002 periods, and call options across all types of mortgages increases substantially, one fails to find a significant impact of call options on default decision. This might explain that the default population has less financial awareness than the prepayment population. Both the squared terms of the put and call option impact their respective default and prepayment hazards significantly and negatively, indicating that the default and prepayment hazards increase with put and call option value in a decreasing way.

The put option value, which measures the probability of negative equity in each period is calculated based on LTVs and house values reported by lender at origination³³. These values might be subjected to appraisal bias. The measures of appraisal bias attempt to control for it and shed light on the effect of appraisal bias on mortgage terminations.

Appraisal bias arises when the appraiser/lender intentionally inflates the property's value to influence the result of the mortgage application evaluated by automated underwriting systems from large secondary mortgage market participants who decide if and how to extend credit lines to borrowers. The influenced results can be represented in several ways: an approval of the mortgage application switched from a prior rejection, a lowered

³³ Details of how to compute probability of negative equity are provided in Appendix A.

interest rate for the borrower from a prior high interest rate, or a higher cash out amount than what would have been given under the unbiased collateral value. Borrowers who are perceived to be “safe” for a certain level of pricing decisions are in fact “risky” and their true default risk is not captured due to manipulation of the origination collateral value.

Measures of appraisal bias can serve as indicators on how desperate and incapable a borrower is financially. If she needs a high valuation bias in order to qualify an already stretched thin mortgage burden, her default risk is undoubtedly higher than those who would qualify for the same mortgage without inflating the collateral value. Therefore, one should expect measures of appraisal bias are predictive to default probabilities and the higher the appraisal bias, the higher the default risk.

The appraisal bias used in the model measures the difference between lender-reported LTV at loan origination and the mark-to-market house price based predicted LTV at origination. The squared term of appraisal bias is also included in the model to capture potential non-linear effect on mortgage terminations³⁴.

Model 1 shows that coefficients on appraisal bias and its squared term are both significant in determining default probability. However, neither of them is significant in determining prepayment probability. This indicates that property valuation bias at loan origination does not influence prepayment behavior. Both prepayment by selling the house and prepayment by refinance require another appraisal of the current property

³⁴ The likelihood ratio test between the non-ruthless model with (Model 1) and without controls for appraisal bias (Model 1' in Appendix C) returns a chi-square value of 355.79 with degree of freedom of 4, which favors the non-ruthless model with controls for appraisal bias.

value. Prepayment by selling the house involves agreement on the current property price from the seller and the buyer, which can be assumed to be independent from the origination appraisal process. Prepayment by refinance is more influenced by the second appraisal at refinance rather than the first appraisal at origination. For example, a higher appraisal value at refinance might lead to a better refinance offer and a higher chance of approval from mortgage lenders. This value is more influenced by current housing market conditions than prior appraisal value at the first mortgage's origination.

Therefore, it is sensible that the origination property valuation bias is found to be insignificant in determining prepayment behaviors. Nevertheless, one should expect that the second appraisal at refinance would impact this new loan's default probability in the future, just as the above model exhibits.

The coefficients on appraisal bias and its squared term show that default probability is significantly influenced by origination appraisal bias. The partial derivative of the default hazard function with respect to appraisal bias evaluated at its minimum, 25th percentile, mean level, median, 75th percentile and 99th percentile are all positive, indicating a higher origination appraisal bias leads to a higher likelihood of default. The parameter on the squared term of the appraisal bias indicates that this effect is increasing in a decreasing fashion. The turning point for the effect to become negative is when appraisal bias exceeds 54 percent; only less than half of one percent of loans in the entire dataset has bias exceeding this value.

The importance of appraisal bias on default requires it be controlled in the default estimation. The mis-measured negative equity based on lender reported house value by itself is insufficient to explain the likelihood of default. Buyers on the secondary mortgage market are gamed if they rely solely on lender reported property value to determine default probabilities and provide pricing schedules. This result calls for reevaluation of the magnitude of equity related variables' impact on default probability and stringent rules on appraisal practice. Default risk will remain underestimated, which results in uncompensated extra default losses born by creditors. This in turn will lead to reduction in consumer credit extension if such appraisal behavior is allowed to continue. Effect of appraisal bias on default risk must be carefully evaluated before one begins investigating the impact of equity on default probability.

Model 1 also addresses asymmetric information by including borrower characteristics, property characteristics and trigger event variables. In addition to the financial option variables and measures for appraisal bias, it also includes consumer credit score provided by Fair Isaac Corp. (e.g. FICO score) at the time of loan origination, borrower's debt to income ratio at loan origination, origination LTV ratio, origination loan amount and quarterly state level unemployment rate where the borrower resides. It attempts to test the trigger event hypothesis that borrower's willingness to exercise financial options may be triggered or hindered by variables other than the put and call options. The trigger event hypothesis states that unexpected loss of job, divorce or death in the family might induce the borrower to take a termination action that otherwise would not be considered utility maximizing.

For example, borrowers with weak financial conditions are more prone to unexpected financial disturbance than those with sound financial conditions. Borrowers' with less creditworthiness, by definition, are less financially trustworthy and responsible; a small amount of financial chaos will likely to incur big changes in their financial behaviors. These borrower characteristics can be approximately reflected by consumer credit score at the time of the loan origination, and their debt to income ratio at loan originations. The macro economic conditions can be approximated by quarterly state level unemployment rate.

There also exists asymmetric information between mortgage borrowers and lenders that is not captured in the financial option variables. For example, borrowers might know more about the volatility of their financial constraints than lenders do and hence choose mortgage contracts accordingly. Such asymmetry might be reflected by origination loan-to-value (LTV) ratio, where borrowers who anticipate fluctuations in incomes might put in few down payments and elect mortgage contracts with high LTV ratio. Similarly, borrowers who are more financially constrained might select a small mortgage with a less value property, reflected by the size of the loan amount and LTV ratios. On the contrary, borrowers with no concerns on either the volatility or the soundness of their financial affordability might apply for big mortgages and/or large down payments and properties with high value and reside in areas with fast house price appreciation. Other asymmetric information might include whether the borrower expects to move out sooner than average, or if the lender charges higher rate for risky borrowers. Information asymmetry

can result in biased estimates of the put and call option variables if they are not controlled for.

Results in Model 1 shows that borrower characteristics, loan characteristics and economic conditions are all significant in determining one or both of the termination outcomes.

The result shows that default hazard monotonically increases as origination credit rating deteriorates. Credit score at origination summarizes the borrower's historical attitude toward credit. Not surprisingly, a borrower with a habit of delinquency and default on his other credit lines also has a tendency to default on his mortgage. It does not have deterministic power in prepayment hazard.

Debt-to-income ratio is found to be positively and significantly impacting default risk; the higher the debt-to-income ratio at origination, the higher the default risk is. Debt-to-income ratio at origination measures the borrower's financial conditions. The higher the ratio is, the less reserves there is should the borrower be forced to readjust his financial plans. Once the reserves runs out, the borrower might resort to defaulting on his mortgage payments to reduce financial burdens. Borrowers with sound financial conditions do not have to exercise default option if faced with the same level of financial disturbance.

Therefore, one should expect a positive relationship between debt-to-income ratio and default probability, which is confirmed by the model results. Debt-to-income ratio does not appear to influence prepayment option.

Origination loan to value (LTV) ratio shows significant and positive impact on default probability. Default probability monotonically increases with origination LTV. As explained earlier, loan to value ratio at origination measures aspects of individual asymmetry that is not captured by financial option variables. Borrowers choose the optimal mortgage contracts to maximize their utilities. High LTV mortgage can be viewed as good options for borrowers who have less liquid assets for the down payment or who want to save for expect big fluctuations in future income flow. Both of the two populations of borrower are less financially stable and are more likely to exercise default option if disturbance occurs than people who are financially stable. Therefore, one should expect high original LTV ratio would lead to high probability of default. The model results suggest that information between borrowers and lender at origination is asymmetric and that there exists positive relationship between LTV ratios and default probability.

Origination loan amount is found to be positively influencing prepayment risk and negatively influencing default risk. This finding once again indicates that information asymmetry exists between borrowers and lenders. Borrowers with big loan amount must have high collateral values, controlling for LTV ratio. And properties with high value are more likely to be offered refinance opportunities because they are less likely to default and incur costs on lenders. This explains why origination loan amount and prepayment probability are positively connected. Mortgages with small loan amount, given LTV ratio are those that are based on low collateral values and hence are deemed more risky to default. In addition, low loan amount also might reflect a borrower's financial soundness

and credit worthiness, which impact a borrower's default risk. Therefore, one should not be surprised to see a high loan amount leads to high probability of prepayment and low probability of default.

Lastly, parameters on state level quarterly unemployment rate suggest that unemployment rate significantly increases default probability. When economic conditions worsen, reflected by an increase in the state level unemployment rate, the mortgage default probability also increases. This indicates that controlling the put option value, a high unemployment rate triggers the borrower to exercise the default option that would not have been exercised in a low unemployment rate environment. All of the above results suggest that borrower characteristics, mortgage characteristics, property characteristics and economic conditions provide explanatory power to default and prepayment behaviors. Option related variables alone are not sufficient to answer mortgage termination behaviors even if termination behaviors are considered financial options³⁵.

In summary, the model result suggests that option related variables are significant in determining mortgage termination outcomes. However, a non-ruthless model specification where borrower characteristics, property characteristics and economic conditions are included in addition to option variables provides a better model fit. In

³⁵ Results from ruthless model specifications are similar to non-ruthless specifications, with likelihood ratio test favoring the non-ruthless model with appraisal bias specification. Please refer to Appendix C for details on estimation results for other specifications.

addition, appraisal bias is found to be significantly impacting default decision, indicating the importance to control for it in mortgage termination studies³⁶.

3.5.2 Unobserved Heterogeneity

This section presents estimation results where borrower heterogeneity is introduced to the specification.

Unobserved heterogeneity arises when there are unobserved individual characteristics that affects the outcome of the interest. It is an important factor because heterogeneity is correlated with the error term and will result in biased estimates for the variables of interest if not controlled for. In the mortgage termination analysis, unobserved heterogeneity may reflect the financial astuteness of borrowers and help explain why we observe different prepayment speeds across borrowers. If borrowers differ in their astuteness toward interest rate, then those who are financially savvy are more likely to exercise prepayment option and those who are financially unsophisticated are more likely to remain in the at risk population for default. For example, Deng et al. point out in their paper that there appears to be three types of people in their dataset: one with the highest likelihood of exercising the prepayment option, the other the lowest, and another group in the middle. Their results suggest that given the same market and economic environment, the high-risk group is about three times riskier than the intermediate group, and about

³⁶ Several versions of different starting values are tested in all the model specifications and the parameters estimated remain to be very close, indicating stable estimates. In addition, similar results have also been obtained in multinomial discrete choice model specifications where default and prepayment are jointly estimated in a conditional probability framework. The multinomial results confirm what have been established in the competing risk proportional hazard models with similar parameters estimates and same conclusions.

twenty times riskier than the low risk group in terms of prepayment. They didn't find heterogeneity in terms of exercising the default option (Deng, Quigley and Van Order 2000). Huang and Ondrich (2002) adopt Deng's approach and conclude that failing to control for unobserved heterogeneity can lead to severe biases in the coefficient estimates of key covariates. The claim that borrowers' difference in financial awareness is picked up by the unobserved heterogeneity is one of the plausible interpretations; there exists other interpretations such as the heterogeneity may be attributed to unmeasured house-specific factors, e.g., unexpected depreciation or appreciation of property values, as well as borrower tastes or abilities. I will use the difference in borrowers' attitude interpretation throughout the paper.

As illustrated in the previous section, the approach taken here is different from what Deng et al. employed because theirs restricts the heterogeneity between prepayment and default to be monotonically correlated. Because they assign the same probability to default and prepayment heterogeneity, the two heterogeneities move in the same direction. The approach taken here is a different implementation of the Heckman-Singer mass point distribution where I allow default and prepayment to move independently and have independent probabilities.

Specifically, I assume a 2x2 discrete support for the distribution of default and prepayment heterogeneity. I assume there are two types of individuals for each risk; their joint distribution and associated probabilities are presented in the following table.

θ_p / θ_d	θ_d^1	θ_d^2	Marginal Probability of θ_p
θ_p^1	p_{11}	p_{12}	$p_{11} + p_{12}$
θ_p^2	p_{21}	p_{22}	$p_{21} + p_{22}$
Marginal Probability of θ_d	$p_{11} + p_{21}$	$p_{12} + p_{22}$	1

This setup does not restrict default and prepayment heterogeneity location parameters to move in the same direction with the same probability. One freely estimates five of the eight parameters with restrictions that all probabilities sum to one, the heterogeneity is distributed mean one (in the exponential specification, it can be approximated by mean zero) and all probabilities must range between zero and one.

The above restrictions can be translated into the following expressions:

$$p_{11} = \frac{\exp(\rho_{11})}{1 + \exp(\rho_{11}) + \exp(\rho_{12}) + \exp(\rho_{21})},$$

$$p_{12} = \frac{\exp(\rho_{12})}{1 + \exp(\rho_{11}) + \exp(\rho_{12}) + \exp(\rho_{21})},$$

$$p_{21} = \frac{\exp(\rho_{21})}{1 + \exp(\rho_{11}) + \exp(\rho_{12}) + \exp(\rho_{21})},$$

$$p_{22} = \frac{1}{1 + \exp(\rho_{11}) + \exp(\rho_{12}) + \exp(\rho_{21})}$$

$$\theta_d^2 = -\frac{P_{11} + P_{21}}{P_{12} + P_{22}} \cdot \theta_d^1,$$

$$\theta_p^2 = -\frac{P_{11} + P_{12}}{P_{21} + P_{22}} \cdot \theta_p^1$$

The five free parameters to be estimated are

$$(\rho_{11}, \rho_{12}, \rho_{21}, \theta_d^1, \theta_p^1)$$

Likelihood ratio tests suggest that models that controlled for unobserved heterogeneity generate better fit than its counterparts that do not control for heterogeneity. The likelihood ratio test before and after controlling for heterogeneity returns a chi-square value of 2152.42 with 5 degrees of freedom for the non-ruthless model with controls for appraisal bias, favoring the specification with heterogeneity³⁷.

Between the two models that controlled for unobserved heterogeneity, the likelihood ratio test suggests that the best model fitness is generated by the non-ruthless model with bias measures, followed by the non-ruthless model without appraisal bias. The chi-square values generated by the likelihood ratio tests between the non-ruthless model with bias measures and its respective restricted versions are 432.57 with 4 degrees of freedom.

³⁷ The likelihood ratio test for the non-ruthless model without controls for appraisal bias before and after controlling for heterogeneity returns a chi-square value of 2075.64 with 5 degree of freedom, favoring the heterogeneity specification.

The following Model 2 presents the non-ruthless model specification with appraisal bias where unobserved heterogeneity is controlled³⁸.

Table V. Maximum Likelihood Estimation of Competing Risks Proportional Hazard Model – Model 2: Non-Ruthless Model With Unobserved Heterogeneity and Controls for Appraisal Bias

³⁸ Model 2' in Appendix C presents the non-ruthless model specification with unobserved heterogeneity without controls for appraisal bias.

	Prepayment	Default
baseline hazard parameters		
2 years of age	-0.8324 (0.2367)	-4.0908 (0.1926)
3 years of age	-0.6595 (0.2369)	-3.6619 (0.2343)
4 years of age	-0.4662 (0.2382)	-3.5446 (0.2712)
5 years of age	-0.0957 (0.2420)	-3.4131 (0.3066)
6 years of age	-0.0779 (0.2644)	-3.6288 (0.3906)
>6 years of age	0.0429 (0.2584)	-2.6978 (0.3650)
Intercept	-8.6928 (1.1772)	-2.3670 (3.2246)
Credit Score	0.1393 (0.6868)	-12.1071 (1.7988)
Debt to Income Ratio	0.2545 (0.2818)	1.9957 (0.6033)
Loan to Value Ratio	0.0333 (0.3173)	4.3569 (0.9048)
Loan Amount	4.5712 (0.9486)	-1.9428 (2.1787)
Appraisal Bias	-0.7063 (0.6470)	8.7681 (1.4168)
Squared Term of Appraisal Bias	1.6886 (1.4938)	-8.6199 (2.7429)
Put Option	-1.6059 (0.6486)	8.5607 (1.6587)
Call Option	8.1473 (0.7608)	0.0379 (1.2597)
Squared Term of Put Option	2.9953 (1.8601)	-7.5961 (2.8785)
Squared Term of Call Option	1.2369 (3.7953)	3.2613 (6.3509)
State Unemployment Rate	0.0670 (0.0351)	0.2390 (0.0741)
vd1		2.6229 (2.0265)
vp1		-6.9287 (0.8083)
rho11		-6.8112 (8.1958)
rho12		-0.9404 (0.8640)
rho21		1.9208 (0.9536)
Loglikelihood		-477,220.84

Note: Standard errors reported in parentheses.

Risks are estimated jointly with maximum likelihood approach.

The inclusion of the unobserved heterogeneity does not change the fundamental relationship between financial variables and prepayment and default hazards. Prepayment hazard remains to be positively influenced by call option and negatively influenced by put option and default hazard remains to be positively influenced by put option. And appraisal bias continues to positively impact default risk.

The non-ruthless model with controls for appraisal bias suggests that there exist groups of people whose attitudes toward prepayment are different, but their attitudes toward default are the same.

θ_p / θ_d	0	0	Marginal Probability of θ_p
-6.929	0.102	0.102	0.204
1.771	0.695	0.102	0.797
Marginal Probability of θ_d	0.797	0.204	1

The above matrix shows the prepayment and default heterogeneity locations and its respective marginal probability and joint density. The result indicates that there exist two groups of borrowers who possess different trends of prepayment risks: one high and one low. According to the heterogeneity parameters, 79.7 percent of people have a high tendency toward prepayment risk, whereas 20.4 percent of the people have low astuteness toward prepayment. There does not appear to be any heterogeneity in the default space³⁹.

³⁹ I further expand the prepayment heterogeneity into 3, 4 and 5 mass points and reduce the default heterogeneity into none. The log likelihood functions are unchanged but several mass points become insignificant. The coefficients of other parameters remain unchanged.

Literature on mortgage terminations establishes that majority of mortgage borrowers are financially savvy toward prepayment opportunities and resistant toward default behaviors (Huang and Ondrich 2002). Their findings are based on analyses without measures for appraisal bias. Hence I estimate the non-ruthless specification without controls for appraisal bias to compare with their results. The result confirms previous findings that majority of people have a high tendency toward prepayment and low tendency toward default. In addition, prepayment heterogeneity and default heterogeneity are negatively correlated: people who are more likely to prepay are also very unlikely to default and vice versa⁴⁰.

As Model 2 exhibits, the heterogeneity in default space disappears after controlling for appraisal bias. A plausible explanation for the disappearance of default heterogeneity is that the heterogeneity in default may be captured by appraisal bias and hence become insignificant in the full model. There still exist different types of people whose attitude toward prepayment is different and their differences are significant in explaining prepayment behavior. The difference in default is fully captured in the full specification model, whereas there is less difference in prepayment that are yet to be captured. The heterogeneity results confirm that there exist different trends of prepayment or default behaviors among borrowers when we fail to incorporate all variables that governing mortgage terminations into the models. In this case, heterogeneity needs to be controlled in order to obtain correct estimates of the variables of interest.

⁴⁰ Please refer to Model 2' in Appendix C.

The approach employed here is a general implementation of Heckman-Singer's mass point heterogeneity distribution. The mass point distribution is utilized to reduce unnecessary and incorrect assumptions on the distribution in order to achieve optimal estimation results. Specifically, the specification applied here assumes there exists a 2X2 distribution of default and prepayment heterogeneities. Default and prepayment heterogeneities can move freely with no restrictions on the directions and relative magnitude.

The specification further eliminates the correlation imposed on prepayment and default heterogeneities. It does not suffer from the limited applications that previous researchers relied. Namely, the specification designed in this paper relaxes the dependency and monotonic relationship that previous studies imposed on the default and prepayment heterogeneity. The results, unlike Huang and Ondrich (2002) which shows prepayment and default awareness move in the same direction, suggest that the majority people who exhibit a high probability of prepayment also have a low probability of default. This result is confirmed by many empirical studies.

For completeness, Model 3 in Appendix C presents the single risk proportional hazard model with non-ruthless specification to compare with the competing risk specification. Due to the flexible single risk specification, variables that are not predictive to prepayment risk are removed from the prepayment model. A comparison of the coefficient estimates across the single risk and the competing risk without heterogeneity model suggests that the results of them are virtually identical. For this data set, explicitly

controlling for the possibility that default and prepayment occur in the same quarter makes almost no difference in the coefficient estimates. This result should not be surprising given the relative infrequency of the default events. However, in the model that controls for unobserved heterogeneity, the magnitudes of several coefficient estimates changed. For example, the coefficient of put option has decreased from 9.21 to 8.56; the coefficient of loan amount has decreased from -6.1 to insignificant. These results suggest that failing to control for unobserved heterogeneity may lead to severe biases in the coefficient estimates of key covariates.

In order to demonstrate that McCall's $\frac{1}{2}$ - $\frac{1}{2}$ weight assumption is inappropriate for mortgage termination studies, I present the estimation results based on McCall's assumption in Model 4 in Appendix C. Comparing the estimation results from Model 1 and Model 4, one concludes that allowing the data to freely estimate the weight not only provides a better model fit than the $\frac{1}{2}$ - $\frac{1}{2}$ arbitrary assumption as represented by the likelihood ratio test, but also provides meaningful estimates of the covariates. It turns out that under the arbitrary weight assumption, most of the covariates become insignificant, contrary to theory predictions. The $\frac{1}{2}$ - $\frac{1}{2}$ assumption on the last period hazard rate on competing risks with grouped duration data may be appropriate in part-time vs. full-time unemployment studies because the two risks are very alike. It is inappropriate for mortgage termination studies because default risk and prepayment risk are drastically different. While prepayments occur with more than 50% likelihood, defaults are only a fraction of a percent. As demonstrated by the model results, one needs to estimate the weight jointly with the model covariates in order to achieve meaning outcome.

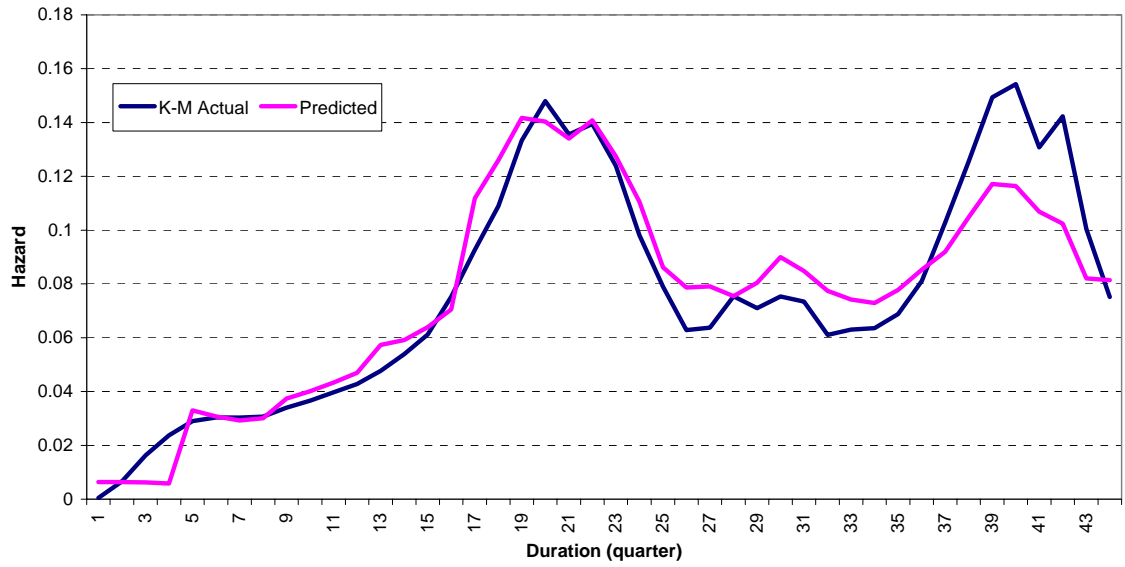
3.5.3 Predicted Hazards

This section provides graphical comparison between empirical hazards and model predicted hazards for prepayment and default. It also presents the relationship between prepayment hazard by calendar time and market interest rate.

Figure 2 compares the Kaplan-Meier empirical termination hazards for prepayment (Panel A) and default (Panel B) with the average predicted hazards from the non-ruthless model with heterogeneity (Model 2). The prediction does a good job tracking the empirical data for prepayment hazard for 25 or less quarter durations. Although still tracking the movement of the actual prepayment, the prediction does not capture the rapid up rise after 36 quarters. This is primarily due to small sample size. Only less than 5 percent of the loan observations were originated prior to 1995, making potential at risk loans with 36 or more quarter durations very few. The actual default hazard presents larger volatility than actual prepayment hazard. Default is a rare event, with the highest observed per quarter hazard less than 0.04 percent. According to the empirical hazards, prepayment hazard can be as high as 1000 times greater than default hazard. Coupled with the significant reduction in surviving sample due to termination by prepay after 20 quarter of durations, default becomes hard to track by the prediction. Nevertheless, the predicted default hazard tracks the general trend of actual default with less accuracy than its counterpart for prepayment.

Figure 2. Actual vs. Predicted Hazards

Panel A. Prepayment



Panel B. Default

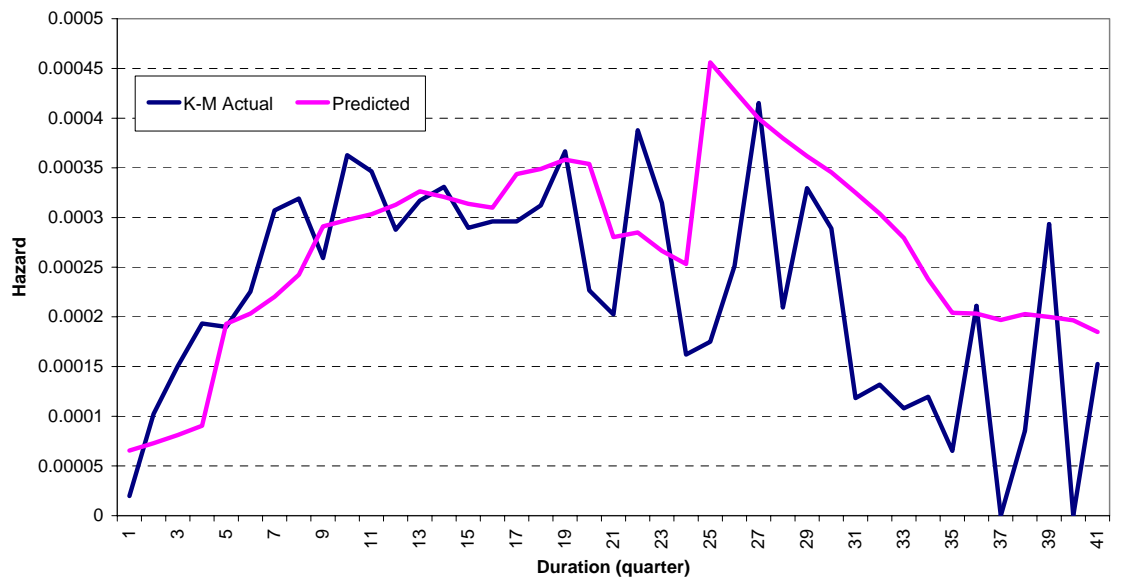


Figure 3 presents the relationship between predicted prepayment hazard by calendar time and market interest rate. Because call option moves in the opposite direction to interest rate, one should expect that calendar specific prepayment hazard is negatively correlated with market interest rate. The market interest rate is obtained from Freddie Mac's 30-year fixed rate mortgage interest rate survey (www.freddiemac.com). Predicted prepayment hazard is negatively correlated with market interest rate; the local highest prepayment usually corresponds to the local lowest interest rate and vice versa. The correlation coefficient and its statistical significance are included in Figure 3. It indicates that prepayment hazard by calendar time closely tracks the movement of market interest rate and they are negatively correlated.

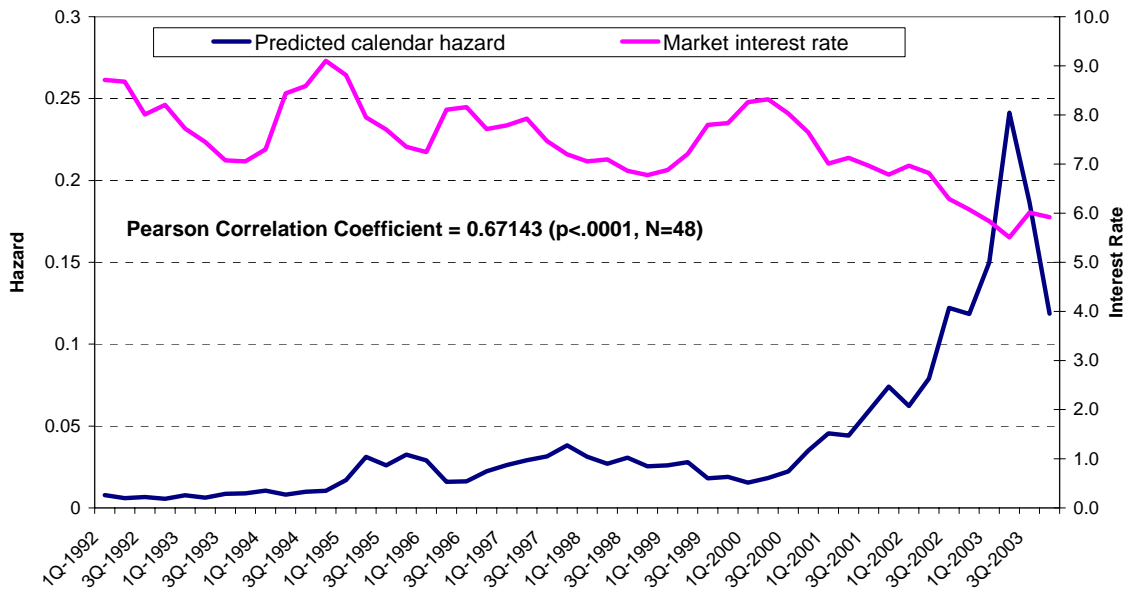


Figure 3. Prepayment Hazard and Market Interest Rate

3.6 Conclusion

This paper analyzes mortgage terminations in a competing risk proportional hazard with grouped duration data framework. It treats default and prepayment as interdependent risks and jointly estimates both termination outcomes in maximum likelihood estimation.

The paper verifies the importance of financial option variables in mortgage terminations. It suggests that the financial value of the call option is strongly associated with the prepayment action and the probability of negative equity is strongly associated with the default option. The results also provide support for the interdependence between prepayment and default behaviors. It indicates that borrowers whose put option is in the money, which translates to a high default hazard, also have low prepayment hazards.

This paper also tests borrower characteristics, loan characteristics and trigger events such as economic conditions in addition to the financial option variables and concludes that variable that measure individual specific information provides improvement to the model fit. For example, borrowers with low credit score are more likely to default than those with high credit scores; borrowers with small financial buffers are more likely to default than those with big buffers. Also, borrowers tend to choose mortgage contract according to their preferences for risk and expected future behaviors. Lastly the state level quarterly unemployment rate is predictive in default behaviors, suggesting the impact of trigger event exists. All the above findings indicate that financial option variables themselves are not sufficient to govern borrowers' mortgage termination behaviors.

This paper also controls for unobserved heterogeneity in the population and estimates the heterogeneity simultaneously with the parameters and the baseline hazards associated with prepayment and default functions. The paper contributes to the literature by applying a general specification of the Heckman-Singer mass point non-parametric heterogeneity distribution. The result generated by this specification produces fruitful results about the relationship between default and prepayment, which is confirmed by many empirical studies. The heterogeneity results suggest that there exist differences among mortgage borrowers, particularly regarding prepayment. If not controlled, the unobserved heterogeneity will lead to biased estimation of the parameters of interest.

The most important feature of this paper is the inclusion of the measures for origination appraisal bias.

Appraisal bias is the natural product of the secondary mortgage market structure where mortgage buyers and mortgage lenders face information asymmetry. Mortgage lenders know more about the quality of the mortgage than mortgage buyers do and it is relatively costly for mortgage buyers to verify the truthfulness of the information provided by mortgage lenders. The asymmetric information structure creates incentives for mortgage lender to maximize his utility function at the expense of mortgage buyers.

Appraisal bias refers to actions taken by mortgage lenders who intentionally inflate collateral property values to under report default risks in order to maximize their profit at the expense of mortgage buyers. This practice brings revenue to lenders either through

expanded business or customer satisfaction or increased bonuses. This practice, on the other hand, affects the measure of equity variables and leads to miscalculated default risk of a borrower. Therefore, mortgage buyers who bear the default risk of mortgages will be adversely impacted and incur unnecessary credit loss imposed by the unmeasured bias. If the credit loss incurred by mortgage buyers keeps increasing, one would expect that honest lenders be penalized with a high fee or secondary mortgage market buyers start tightening credit extension, significantly reducing credit supplies for borrowers.

Realizing all the adverse impact by appraisal bias, one concludes that it shall be controlled for in order to reach a correct and unbiased measure of default risk related variables.

This paper uses a unique dataset that allows me to quantify measures of appraisal bias. The results suggest that appraisal bias significantly increase default hazards across all model specifications. The higher the appraisal bias is, the higher the default risk. In addition, inclusion of measures of appraisal bias appears to alleviate the unobserved heterogeneity existed on the default side and provides superior estimation results. This suggests that one needs to update his belief on the magnitude of equity related variables on default and infer the correct default risk and price accordingly.

Future research should continue on identifying variables that can explain the prepayment behaviors, suggested by the heterogeneity existed on the prepayment side.

In addition, researchers should continue improving the measures on appraisal bias. The bias measured in the paper is based on a unique dataset where a property enters the dataset twice with the first transaction a purchase transaction. The data reflects properties that had been purchased at least once within a certain time limit, which may not reflect the entire mortgage market in the present. In the future, researchers should try expanding the dataset to contain as much properties and transactions occurred across a larger horizon to capture the entire market condition.

Another approach of controlling for appraisal bias can be realized by implicitly correct for appraisal bias in the put option calculation instead of explicitly entering it into the model. Since the put option is computed based on house value, which in turn is determined by appraisal bias, one can correct the appraisal bias by correcting the put option value directly. Therefore, the model comparisons could be one that focus on measurement error and based on non-nested statistical model inference.

Non-nested model and hypotheses are called for where there is a need for statistical procedures for testing non-nested, or separate, parametric families of hypotheses. In these situations, one model cannot be obtained from the other by imposing appropriate restrictions or as a limiting suitable approximation. And the conventional model performance comparison tests cannot in general be used for hypotheses of the non-nested type. In my example, a put option variable based on lender reported house value and a put option variable based on appraisal bias adjusted house value create two models that are non-nested and one cannot conclude on the model fitness by the usual tests. These tests of

separate models are specification tests that use information about a specific alternative and test whether the null can predict the performance of the alternative. The non-nested model test results can give inference on how to best correct for appraisal bias.

Theoretical research can also be done on solving for an optimal game structure such that the incentive for appraisal bias disappears. Eliminating the incentive for appraisal bias can completely erase the existence of appraisal bias and avoid unnecessary losses to participants in the residential mortgage market.

Appendix A: Specification of Call and Put Option Variables

Appendix A and B are borrowed extensively from Deng et. al. (2000). Following directly from Deng, Quigley, and Van Order (1996, 2000), the variables measuring the value of the call and put options are defined by the initial terms of the mortgage and current market conditions.

A fixed-rate mortgage i with an original loan amount of O_i , a mortgage note rate of r_i , and a monthly payment of P_i in principal and interest and mortgage term TM_i in quarters. At each quarter k after origination at time τ_i , the local market interest rate is m_{j,τ_i+k} , where j indexes the local region. The call option variable is defined as the difference in the present value of the payment stream at the current market interest rate and mortgage note rate:

$$CO_{i,k_i} = \frac{\sum_{t=1}^{TM_i-k_i} \frac{P_i \times 3}{\left(1 + \frac{m_{j,\tau_i+k_i}}{4 \times 100}\right)^t} - \sum_{t=1}^{TM_i-k_i} \frac{P_i \times 3}{\left(1 + \frac{r_i}{4 \times 100}\right)^t}}{\sum_{t=1}^{TM_i-k_i} \frac{P_i \times 3}{\left(1 + \frac{r_i}{4 \times 100}\right)^t}} = \frac{V_{i,m_{j,\tau_i+k_i}} - V_{i,r_i}}{V_{i,r_i}}$$

Since interest rate is always measured annually and in percentage, and the observation windows are in quarterly, we apply 4, the number of quarters in a year and 100, to change it from its normal form to a decimal value measured by quarters. 3 is the number of months in a quarter and we use it to compute the quarterly principal and interest payments.

The call option value variable changes each quarter due to quarterly changes in market interest rate. It measured the difference between a borrower's present value of mortgage payment evaluated at her contract note rate and that evaluated at the current market rate. A positive call option indicates prepayment option is valuable to exercise.

The put option variable is defined as the probability that the ratio of contemporaneous equity to market value of the mortgage is negative. The contemporaneous property value M_{i,k_i} is constructed as mapping the origination property value to current by the quarterly zip code level weighted repeated sale house price index; the contemporaneous mortgage value is simply V_{i,m_j,τ_i+k_i} listed above. The ratio of equity to market value of the mortgage is:

$$E_{i,k_i} = \frac{M_{i,k_i} - V_{i,m_j,\tau_i+k_i}}{V_{i,m_j,\tau_i+k_i}}$$

The put option is defined as the probability that equity is negative:

$$PO_{i,k_i} = \Pr(E_{i,k_i} < 0) = \Phi\left(\frac{\log V_{i,m_j,\tau_i+k_i} - \log M_{i,k_i}}{\sqrt{\omega^2}}\right)$$

where $\Phi(\cdot)$ is cumulative standard normal distribution function, and ω^2 is the estimated variance defined in Appendix B. The more negative the put option value is, the more likely the default option is exercised.

Appendix B: The Weighted Repeated Sales House Price Index

The weighted repeated sales (WRS) method was first proposed by Bailey, Muth and Nourse (1963) and implemented empirically in the 1980s by Case and Shiller (1987; 1989). This method uses two consecutive property transaction prices to estimate the average appreciation rate for the location.

Since the time between the transaction pairs can vary within the sample, this method suggests a three-stage estimation procedure to address it. First, time dummies are used to generate average value changes for cohorts of transactions. Secondly, the variance is assumed to be a function of time difference between two transactions and its square term. Hence the error term from the 1st stage is regressed on time difference and its square term. Lastly, a generalized least square regression is employed with the weights equal the fitted residuals from the 2nd stage.

Following Bailey, Muth, and Nourse (1963) exactly and borrowing from Deng et. al. (2000), the model assumes that the natural logarithm of the price of house i at time t can be expressed in terms of a market price index, a Gaussian random walk and white noise, such that

$$P_{it} = I_t + H_{it} + N_{it}$$

where I_t is the logarithm of the market house price level, H_{it} is a Gaussian random walk with the following assumption:

$$E(H_{i,t+k} - H_{it}) = 0,$$
$$E(H_{i,t+k} - H_{it})^2 = k\sigma_{\eta_1}^2 + k^2\sigma_{\eta_2}^2;$$

and N_{it} is white noise with the following assumption:

$$E(N_{it}) = 0,$$

$$E(N_{it})^2 = \frac{1}{2}\sigma_v^2$$

In addition, H_{it} and N_{it} are assumed to be independent.

This specification allows us to express the percent change in transaction prices as a function of an average appreciation rate over time. The Gaussian random walk describes how variation in individual house price appreciation rates around the rate of change in the market index causes house prices to vary over time. The white noise represents cross-sectional dispersion in housing values arising from purely idiosyncratic differences in how the individual properties are valued at any given point in time.

The model is estimated on paired sales properties. In the first stage, the log price of the second sale minus the log price of the first sale is regressed on a set of dummy variables, one for each time period in the sample except the first period. The dummy variables have values of 1 for the quarter of the second sale, -1 for the quarter of the first sale and zeros otherwise.

In the second stage, the squared residuals ω^2 from each observation in the first stage are regressed on k and k^2

$$\omega^2 = A + Bk + Ck^2$$

where k is the interval between the first and second sale. The coefficients A , B , and C are estimates of σ_v^2 , $\sigma_{\eta_1}^2$, and $\sigma_{\eta_2}^2$ respectively.

In the third stage, the stage one regression is re-estimated by GLS with weights

$(A + Bk + Ck^2)^{1/2}$. The estimated log price level difference $(\hat{I}_{t+k} - \hat{I}_t)$ is distributed with

mean $(I_{t+k} - I_t)$, and variance $(k\sigma_{\eta_1}^2 + k^2\sigma_{\eta_2}^2 + \sigma_v^2)$. Denote $zip_t = \exp(\hat{I}_t)$ as the

estimated zip code level house price index, in k th quarter subsequent to purchase,

$\log\left(\frac{zip_{t+k}}{zip_t}\right)$ is then normally distributed with mean $(I_{t+k} - I_t)$ and variance

$(k\sigma_{\eta_1}^2 + k^2\sigma_{\eta_2}^2 + \sigma_v^2)$.

Appendix C: Other Specification Estimations

Model 1'. Maximum Likelihood Estimation Of Competing Risk Proportional Hazard Model
 - Ruthless Model With Controls For Appraisal Bias

	Prepayment	Default
baseline hazard parameters		
2 years of age	-2.8048 (0.2352)	-6.0333 (0.1805)
3 years of age	-2.6302 (0.2334)	-5.3693 (0.1979)
4 years of age	-2.4760 (0.2316)	-5.0901 (0.2199)
5 years of age	-2.0906 (0.2317)	-4.9936 (0.2400)
6 years of age	-2.1754 (0.2495)	-5.2020 (0.3064)
>6 years of age	-2.2818 (0.2484)	-4.8190 (0.2946)
Intercept	-0.5110 (0.2194)	-3.8511 (0.2265)
Appraisal Bias	-0.3025 (0.4761)	8.4178 (1.3021)
Squared Term of Appraisal Bias	0.2381 (1.4535)	-7.7983 (2.3362)
Put Option	-1.2066 (0.6054)	14.1848 (1.3888)
Call Option	8.6958 (0.7084)	0.0134 (1.0802)
Squared Term of Put Option	1.9046 (1.6105)	-13.1606 (2.4858)
Squared Term of Call Option	-5.9875 (3.0321)	-0.6570 (5.0358)
Loglikelihood	-481,474.04	

Note: Standard errors reported in parentheses.

Risks are estimated jointly with maximum likelihood approach.

Model 1". Maximum Likelihood Estimation Of Competing Risk Proportional Hazard Model
- Non-Ruthless Model Without Controls For Appraisal Bias

	Prepayment	Default
baseline hazard parameters		
2 years of age	-0.7986 (0.2358)	-4.2309 (0.1878)
3 years of age	-0.6352 (0.2345)	-3.7599 (0.2150)
4 years of age	-0.4902 (0.2348)	-3.7299 (0.2439)
5 years of age	-0.1009 (0.2349)	-3.6463 (0.2679)
6 years of age	-0.1410 (0.2518)	-3.8413 (0.3355)
>6 years of age	-0.2489 (0.2510)	-3.1738 (0.3257)
Intercept	-8.7036 (1.0010)	5.1007 (2.3413)
Credit Score	0.4315 (0.5507)	-11.7820 (1.5082)
Debt to Income Ratio	0.0975 (0.2321)	1.7068 (0.5142)
Loan to Value Ratio	-0.1576 (0.2675)	3.3629 (0.7115)
Loan Amount	4.8573 (0.7934)	-5.3766 (1.9191)
Put Option	-1.2428 (0.6076)	8.2501 (1.5533)
Call Option	8.2040 (0.7043)	-0.8837 (1.1385)
Squared Term of Put Option	2.3828 (1.5967)	-8.0380 (2.7349)
Squared Term of Call Option	-4.9707 (2.9930)	1.4358 (4.9642)
State Unemployment Rate	0.0595 (0.0305)	0.2686 (0.0654)
Loglikelihood	-478,474.95	

Note: Standard errors reported in parentheses.

Risks are estimated jointly with maximum likelihood approach.

Model 1". Maximum Likelihood Estimation Of Competing Risk Proportional Hazard Model
- Ruthless Model Without Controls For Appraisal Bias

	Prepayment	Default
baseline hazard parameters		
2 years of age	-2.8053 (0.2349)	-6.0600 (0.1765)
3 years of age	-2.6293 (0.2331)	-5.4741 (0.1919)
4 years of age	-2.4742 (0.2312)	-5.2850 (0.2106)
5 years of age	-2.0870 (0.2311)	-5.2318 (0.2267)
6 years of age	-2.1726 (0.2490)	-5.4337 (0.2923)
>6 years of age	-2.2810 (0.2480)	-4.9967 (0.2793)
Intercept	-0.5173 (0.2184)	-3.3699 (0.1981)
Put Option	-1.1848 (0.5929)	12.8848 (1.2563)
Call Option	8.6655 (0.7036)	1.2152 (1.0128)
Squared Term of Put Option	1.8964 (1.6085)	-12.7325 (2.3527)
Squared Term of Call Option	-5.9203 (3.0250)	-1.0794 (4.5281)
Loglikelihood	-481,631.22	

Note: Standard errors reported in parentheses.

Risks are estimated jointly with maximum likelihood approach.

Model 2'. Maximum Likelihood Estimation Of Competing Risk Proportional Hazard Model
- Non-Ruthless Model With Unobserved Heterogeneity, Without Controls For Appraisal Bias

	Prepayment	Default
baseline hazard parameters		
2 years of age	-0.8316 (0.2365)	-4.1484 (0.1778)
3 years of age	-0.6566 (0.2366)	-3.5015 (0.1949)
4 years of age	-0.4656 (0.2378)	-3.2590 (0.2266)
5 years of age	-0.0954 (0.2414)	-3.1345 (0.2582)
6 years of age	-0.0823 (0.2639)	-3.2867 (0.3298)
>6 years of age	0.0341 (0.2582)	-2.5222 (0.3067)
Intercept	-8.9173 (1.1669)	6.1348 (2.6388)
Credit Score	0.2340 (0.6705)	-15.2867 (1.4107)
Debt to Income Ratio	0.2501 (0.2762)	2.0708 (0.6425)
Loan to Value Ratio	-0.0051 (0.3126)	3.8461 (0.7995)
Loan Amount	4.5115 (0.9416)	-7.0982 (2.1174)
Put Option	-1.5115 (0.6426)	9.0167 (1.5232)
Call Option	8.1190 (0.7564)	-0.9131 (1.0982)
Squared Term of Put Option	2.8931 (1.8966)	-8.0856 (2.4256)
Squared Term of Call Option	1.1513 (3.8317)	10.7734 (5.4462)
State Unemployment Rate	0.0672 (0.0349)	0.1887 (0.0621)
vd1		5.6134 (1.1588)
vp1		-12.4279 (0.6745)
rho11		-7.4191 (1.5515)
rho12		-2.9806 (0.2361)
rho21		-2.6274 (0.3833)
Loglikelihood		-477,437.13

Note: Standard errors reported in parentheses.

Risks are estimated jointly with maximum likelihood approach.

Model 3. Maximum Likelihood Estimation of Single Risk Proportional Hazard Model
- Non-Ruthless Model

	Prepayment	Default
baseline hazard parameters		
2 years of age	-0.8025 (0.2346)	-4.1829 (0.1895)
3 years of age	-0.6428 (0.2329)	-3.6481 (0.2178)
4 years of age	-0.5024 (0.2324)	-3.5176 (0.2474)
5 years of age	-0.1138 (0.2318)	-3.3501 (0.2781)
6 years of age	-0.1551 (0.2486)	-3.5753 (0.3481)
>6 years of age	-0.2611 (0.2474)	-2.9627 (0.3442)
Intercept	-8.3100 (0.8923)	3.9370 (2.5198)
Credit Score	.	-11.4915 (1.6281)
Debt to Income Ratio	.	1.9761 (0.5561)
Loan to Value Ratio	.	4.9386 (0.8639)
Loan Amount	4.7154 (0.7500)	-6.0626 (2.0323)
Appraisal Bias	.	8.8814 (1.3061)
Squared Term of Appraisal Bias	.	-8.6491 (2.4927)
Put Option	-1.4891 (0.7972)	9.2090 (1.5910)
Call Option	8.1863 (0.6990)	-1.7433 (1.2152)
Squared Term of Put Option	2.6667 (1.5168)	-8.4970 (2.7219)
Squared Term of Call Option	-4.8917 (2.9780)	3.5119 (5.0555)
State Unemployment Rate	0.0622 (0.0302)	0.2486 (0.0671)
Loglikelihood	-473,121.20	-5,240.41

Note: Standard errors reported in parentheses.

Risks are estimated jointly with maximum likelihood approach.

Model 4. Maximum Likelihood Estimation Of Competing Risk Proportional Hazard Model
 - Using McCall's 1/2 - 1/2 Assumption on the Weight instead of Estimating the Weight

	Prepayment	Default
baseline hazard parameters		
2 years of age	-0.8020 (0.2334)	-9.3644 (0.8766)
3 years of age	-0.6377 (0.2321)	-7.8722 (2.1269)
4 years of age	-0.4940 (0.2327)	-8.3301 (0.9455)
5 years of age	-0.1055 (0.2331)	-9.3884 (0.5304)
6 years of age	-0.1447 (0.2500)	-6.2023 (2.5338)
>6 years of age	-0.2477 (0.2490)	-5.9131 (2.5228)
Intercept	-8.6045 (0.9997)	-2.1034 (5.6193)
Credit Score	0.3077 (0.5566)	-12.1277 (6.0774)
Debt to Income Ratio	0.1036 (0.2310)	1.5661 (4.5524)
Loan to Value Ratio	-0.1396 (0.2690)	3.5560 (6.9674)
Loan Amount	4.8435 (0.7899)	-0.7064 (3.3823)
Appraisal Bias	-0.3993 (0.5037)	9.0292 (5.4584)
Squared Term of Appraisal Bias	0.9322 (1.3606)	-8.9700 (13.3389)
Put Option	-1.2080 (0.8909)	9.1640 (12.8536)
Call Option	8.1939 (0.7043)	-0.8390 (9.5221)
Squared Term of Put Option	2.3614 (1.5820)	-7.5192 (16.0577)
Squared Term of Call Option	-5.0767 (2.9858)	2.9676 (2.7376)
State Unemployment Rate	0.0596 (0.0304)	-1.6188 (0.8341)
Loglikelihood	-569,270.70	

Note: Standard errors reported in parentheses.

Risks are estimated jointly with maximum likelihood approach.

Bibliography

- Akerlof, G. (1970): "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism," *The Quarterly Journal of Economics* 84 (3), 488-500.
- Amemiya, T (1985): *Advanced Econometrics*. Harvard University Press, Cambridge, MA.
- An, M. (2001): "Maximum Likelihood Estimation of a Proportional-hazard, Competing-Risk Model with Grouped Duration Data," manuscript.
- An, M. (2004): "Likelihood-Based Estimation of a Proportional-Hazard, Competing Risk Model with Grouped Duration Data," forthcoming in *Journal of Econometrics*, 2004
- Anderson, P. (1982): "Testing the Goodness of Fit of Cox's Regression and Life Model," *Biometrics*, 38, 66-77.
- Archer, W., and D. Ling (1993): "Pricing Mortgage-Backed Securities: Integrating Optimal Call and Empirical Models of Prepayment," *Journal of The American Real Estate and Urban Economics Association*, 21(4), 373-404.
- Archer, W., D. Ling, and G. McGill (1996): "The Effect of Income and Collateral Constraints on Residential Mortgage Terminations," *Regional Science and Urban Economics*, 24(3-4), 235-261.
- Austen-Smith, D. and Banks J. (1989): *Electoral Accountability and Incumbency*. In Ordeshook, P. (Ed.) *Models of Strategic Choice in Politics*. Ann Arbor: University of Michigan Press.
- Bailey, M., R. Muth, and H. Nourse (1963): "A Regression Method for Real Estate Price Index Construction," *Journal of the American Statistics Association*, 58, 933-942.
- Bigelow, J. R. Cooper, and T. Ross (1993): "Warranties without Commitment to Market Participation," *International Economic Review*, 34 (1) 85-100.
- Black, F., and M. Scholes (1973): "The Pricing of Options and Corporate Liabilities," *Journal of Political Economy*, 81, 637-654.
- Buser, S., and P. Hendershott (1984): "Pricing Default Free Mortgages," *Housing Finance Review*, 3, 405-429.
- Campbell, T., and J. Dietrich (1983): "The Determinants of Default on Conventional Residential Mortgages," *Journal of Finance*, 38 (5), 1569-1581.
- Case, K., and R. Shiller (1987): "Prices of Single-Family Homes Since 1970: New Indexes for Four Cities," *New England Economic Review*, 45-56.

- Champsaur, P. and J. C. Rochet (1989): "Multiproduct Duopolists," *Econometrica*, 57 (3), 533-557.
- Chinloy, P., M. Cho, and I. Megbolugbe (1997): "Appraisals, Transaction Incentives, and Smoothing," *Journal of Real Estate Finance and Economics*, 14 (1), 89-112.
- Cho, M. and I. Megbolugbe (1996): "An Empirical Analysis of Property Appraisal and Mortgage Redlining," *Journal of Real Estate Finance and Economics*, 7(2), 145-172.
- Ciochetti, B., Y. Deng, B. Gao, and R. Yao (2002): "The Termination of Mortgage Contracts through Prepayment and Default in the Commercial Mortgage Markets: A Proportional Hazard Approach with Competing Risks," *Real Estate Economics*, 30 (4), 595-633.
- Ciochetti, B., Y. Deng, Lee, Shilling and R. Yao (2003): "A Proportional Hazards Model of Commercial Mortgage Default with Origination Bias," *Journal of Real Estate Finance and Economics*, 27 (1), 5-23.
- Cox, D. (1972): "Regression Models and Life Tables (With Discussion)," *Journal of the Royal Statistical Society*, 34(2), 18-220.
- Cox, D. (1975): "Partial Likelihood," *Biometrika*, 62(2), 269-276.
- Cox, D. and D. Oakes (1984): "Analysis of Survival Data," Monographs on Statistics and Applied Probability, Chapman and Hall, London.
- Cox, D., J. Ingersoll, and S. Ross (1985): "A Theory of the Term Structure of Interest Rates," *Econometrica*, 53, 385-467.
- Crawford, V. and J. Sobel (1982): "Strategic Information Transmission," *Econometrica*, 50, 1431-51.
- Cunningham, D. and C. Capone (1990): "The Relative Termination Experience of Adjustable to Fixed-Rate Mortgages," *The Journal of Finances* 45 (5), 1687-1703.
- Cunningham, D. and P. Hendershott (1984): "Pricing FHA Mortgage Default Insurance," *Housing Finance Review*, 13, 373-392.
- Deng, Y. (1997): "Mortgage Termination: An Empirical Hazard Model with a Stochastic Term Structure," *Journal of Real Estate Finance and Economics*, 14(3), 309-320.
- Deng, Y., and J. Quigley and R. Van Order (2000): "Mortgage Terminations, Heterogeneity, and the Exercise of Mortgage Options," *Econometrica*, 68(2), 275-307.

- Dunn, K. and J. McConnell (1981): "Valuation of Mortgage-Backed Securities," *The Journal of Finance*, 36, 599-617.
- Epperson, J., J. Kau, D. Keenan and W. Muller (1985): "Pricing Default Risk in Mortgages," *Journal of The American Real Estate and Urban Economics Association*, 13, 261-272.
- Findley, M., and D. Capozza (1977): "The Variable Rate Mortgage: An Option Theory Perspective," *Journal of Money, Credit and Banking*, 9(2), 356-364.
- Foster, C., and R. Van Order (1984): "An Option-Based Model of Mortgage Default," *Housing Finance Review*, 3(4), 351-372.
- Foster, C., and R. Van Order (1985): "FHA Terminations: A Prelude to Rational Mortgage Pricing," *Journal of The American Real Estate and Urban Economics Association*, 13(3), 273-291.
- Giliberto, S. M., and A. L. Houston (1989): "Relocation Opportunities and Mortgage Default," *Journal of The American Real Estate and Urban Economics Association*, 17(1), 55-69.
- Green, J., and J. Shoven (1986): "The Effect of Interest Rates on Mortgage Prepayments," *Journal of Money, Credit and Banking*, 18, 41-50.
- Grossman, S.J. and O.D. Hart (1983): "An Analysis of the Principal-Agent Problem," *Econometrica* 51, 7-45.
- Han, A., and J. Hausman (1990): "Flexible Parametric Estimation of Duration and Competing Risk Models," *Journal of Applied Econometrics*, 5, 1-28.
- Hansz, H., and J. Diaz (2001): "Valuation Bias in Commercial Appraisal: A Transaction Price Feedback Experiment," *Real Estate Economics*, 29(4), 553-565.
- Heckman, J., and B. Singer (1984): "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data," *Econometrica*, 52(2), 271-320.
- Heal, G. (1977): "Guarantees and Risk-Sharing," *The Review of Economic Studies*, 44 (3), 549-560.
- Hendershott, P., and R. Van Order (1987): "Pricing Mortgages: An Interpretation of the Models and Results," *Journal of Financial Services Research*, 1, 77-111.
- Herzog, J.P., and J.S. Earley (1970): "Home Mortgage Delinquency and Foreclosure," New York: National Bureau of Economic Research.

- Holmstrom, B. (1979): "Moral Hazard and Observability," *Bell Journal of Economics*, 10, 74-91.
- Holmstrom, B., and P. Milgrom (1987): "Aggregation and Linearity in the Provision of Intertemporal Incentives," *Econometrica*, 55, 303-328.
- Huang, W., and J. Ondrich (2002): "Stay, Pay, or Walk Away: A Hazard Rate Analysis of Federal Housing Administration-Insured Multifamily Mortgage Terminations," 13 (1), 85-117.
- Jackson, J. and D. Kasserman (1980): "Default Risk on Home Mortgage Loans: A Test of Competing Hypotheses," *Journal of Risk and Insurance* 3, 678-690.
- Jaffee, D. T. Russell (1976): "Imperfect Information, Uncertainty, and Credit Rationing," *Quarterly Journal of Economics* 90 (4) 651-666.
- Judge, G., W. Griffiths, R. Hill, H. Lutkepohl, and T. Lee (1985): *The Theory and Practice of Econometrics, Second Edition*, Wiley Series in Probability and Mathematics Statistics.
- June, A. (1962): "Terms on Conventional Mortgage Loans on Existing Homes," *Journal of Finance* 17, 432-443.
- Kalbfleisch, J., and R. Prentice (1980): *The Statistical Analysis of Failure Time Data*, New York Wiley.
- Katz, L. (1986): "Layoff, Recall, and the Duration of Unemployment," Working Paper 1825. Cambridge, MA: National Bureau of Economic Research.
- Katz, L., and B. Meyer (1990): "Unemployment Insurance, Recall Expectations, and Unemployment Outcomes," *Quarterly Journal of Economics*, 105, 975-1002.
- 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., and D. Keenan (1996): "Patterns of Rational Default," Working Paper, The University of Georgia, Athens GA.
- Kau, J., D. Keenan, W. Muller, and J. Epperson (1992): "A Generalized Valuation Model for Fixed-Rate Residential Mortgages," *Journal of Money, Credit and Banking*, 24(3), 279-299.
- Kau, J., D. Keenan, W. Muller, and J. Epperson (1995): "The Valuation at Origination of Fixed Rate Mortgages with Default and Prepayment," *The Journal of Real Estate Finance and Economics*, 11(1), 5-36.

- Kennan, J. (1985): "The Duration of Contract Strikes in U. S. Manufacturing," *Journal of Econometrics*, 28, 5-28.
- Kiefer, N. (1988): "Economic Duration Data and Hazard Functions," *Journal of Economic Literature*, 26, 646-679.
- Kinnard, W., M. Lenk, and E. Worzala (1997): "Client Pressure in the Commercial Appraisal Industry: How Prevalent is it?" *Journal of Property Valuation and Investment*, 15(3), 233-244.
- Klein, B. and K. Leffler (1981): "The Role of Market Forces in Assuring Contractual Performance," *Journal of Political Economy*, 89 (4) 615-641.
- Lancaster, T. (1979): "Econometric Methods for the Duration of Unemployment," *Econometrica*, 47, 646-679.
- Lancaster, T. (1990): *The Econometric Analysis of Transition Data*. Cambridge: Cambridge University Press.
- Levy, D., and E. Schuck (1999): "The Influence of Clients on Valuation," *Journal of Property Valuation and Investment*, 17(4), 380-400.
- Maskin, E. and J. Riley (1984): "Monopoly with Incomplete Information," *Rand Journal of Economics*, 15, 171-196.
- McCall, B. (1992): "A Note on the Identifiability of a Dynamic Binary Choice Model with State Dependence," *Economic Letter*, 40, 273-280.
- McCall, B. (1993): "The Non-Parametric Identifiability of a Competing Risks with Regressors," University of Minnesota, mimeo.
- McCall, B. (1994): "Testing the Proportional Hazards Assumption in the Presence of Unmeasured Heterogeneity," *Journal of Applied Econometrics*, 9(3), 321-334.
- McCall, B. (1996): "Unemployment Insurance Rules, Joblessness, and Part-time Work," *Econometrica*, 64 (3), 647-682.
- McCall B. (1997): "The Determinants of Full-Time Versus Part-Time Reemployment Following Job Displacement," *Journal of Labor Economics*, 15(4), 714-734.
- McFadden, D. (1973): "Conditional Logit Analysis of Qualitative Choice Behavior," *Frontiers in Econometrics*, edited by P. Zarembka. New York: Academy Press.
- Merton, R. (1973): "Theory of Rational Option Pricing," *Bell Journal of Economics and Management Science*, 4, 141-183.

- Meyer, B. (1986): "Semiparametric Estimation of Hazard Models," MIT Working Paper.
- Meyer, B. (1990): "Unemployment Insurance and Unemployment Spells," *Econometrica*, 58, 757-782.
- Mirrlees, James A. (1986): "The Theory of Optimal Taxation," Chapter 24 in K.J. Arrow and M.D. Intriligator (eds.), *Handbook of Mathematical Economics, Volume III*, 1197-1249. New York: Elsevier Science Publishers.
- Noordewier, T., D. Harrison, and K. Ramagopal (2001): "Semivariance of Property Value Estimates As A Determinant of Default Risk," *Real Estate Economics*, 29(1), 127-159.
- Page, A. (1964): "The Variation of Mortgage Interest Rates," *Journal of Business*, 7 (3), 280-294.
- Prentice, R., and L. Gloeckler (1978): "Regression Analysis of Grouped Survival Data with an Application to Breast Cancer Data," *Biometrics*, 34, 57-67.
- Quercia, R., and M. Stegman (1992): "Residential Mortgage Default: A Review of the Literature," *Journal of Housing Research*, 3, 341-379.
- Quigley, J., and R. Van Order (1990): "Efficiency in the Mortgage Market: The Borrower's Perspective," *Journal of The American Real Estate and Urban Economics Association*, 18(3), 237-252.
- Quigley, J., and R. Van Order (1995): "Explicit Tests of Contingent Claims Models of Mortgage Default," *The Journal of Real Estate Finance and Economics*, 11(2), 99-117.
- Rothschild, M., and J. E. Stiglitz (1976): "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information," *Quarterly Journal of Economics*, 90, 626-649.
- Salop, J. and S. Salop (1976): "Self-Selection and Turnover in the Labor Market," *Quarterly Journal of Economics*, 90, 619-627.
- Sandor, R., and H. Sosin (1975): "The Determinants of Mortgage Risk Premiums: A Case Study of the Portfolio of A Savings and Loan Association," *The Journal of Business*, 48 (1), 27-38.
- Schwartz, E., and W. Torous (1989): "Prepayment and The Valuation of Mortgage-Backed Securities," *The Journal of Finance*, 44(2), 375-392.
- Schwartz, E., and W. Torous (1993): "Mortgage Prepayment and Default Decisions: A Poisson Regression Approach," *Journal of The American Real Estate and Urban Economics Association*, 21(4), 431-449.

- Shapiro, S. C. (1982): "Generalized augmented transition network grammars for generation from semantic networks," *The American Journal of Computational Linguistics*, 8 (1), 12–25.
- Shapiro, S. C. (1983): "Premiums for High Quality Products as Returns to Reputations," *Quarterly Journal of Economics*, 98 (4), 659-679.
- Spence, M. Market Signaling: Informational Transfer in Hiring and Related Processes, Cambridge, Harvard University Press, 1974.
- Spence, M. (1973): "Job Market Signaling," *Quarterly Journal of Economics*, August 1973.
- Spence, M. (1974): "Competitive and Optimal Responses to Signals: An Analysis of Efficiency and Distribution," *Journal of Economic Theory*, March 1974.
- Spence, M. (1976): "Competition in Salaries, Credentials, and Signaling Prerequisites for Jobs," *Quarterly Journal of Economics*, 90 (1), 51-74.
- Spence, M. (1976): "Informational Aspects of Market Structure: An Introduction," *Quarterly Journal of Economics*, 90 (4), 591-597.
- Stanton, R. (1995): "Rational Prepayment and the Valuation of Mortgage-Backed Securities," *The Review of Financial Studies*, 8(3), 677-708.
- Stiglitz, J. and A. Weiss (1981): "Credit Rationing in Markets with Imperfect Information," *The American Economic Review*, 71 (3), 393-410.
- Stiglitz, J. E. (1976): "Monopoly and rate of extraction of exhaustible resources," *The American Economic Review*, 66 (4), 655—661.
- Sueyoshi, G. (1992): "Semiparametric Proportional Hazards Estimation of Competing Risks Models with Time-Varying Regressors," *Journal of Econometrics*, 51, 25-58.
- Titman, S., and W. Torous (1989): "Valuing Commercial Mortgages: An Empirical Investigation of the Contingent Claims Approach to Pricing Risky Debt," *The Journal of Finance*, 44, 345-373.
- Vandell, K. (1978): "Default Risk Under Alternative Mortgage Instruments," *Journal of Finance* 33 (5), 1279-1296.
- Vandell, K. (1992): "Predicting Commercial Mortgage Foreclosure Experience," *Journal of the American Real Estate and Urban Economics Association*, 20(1), 319-359.

- Vandell, K. (1993): "Handing Over the Keys: A Perspective on Mortgage Default Research," *Journal of The American Real Estate and Urban Economics Association*, 21(3), 211-246.
- Vandell, K. (1995): "How Ruthless is Mortgage Default? A Review and Synthesis of the Evidence," *Journal of Housing Research*, 6(2), 245-264.
- 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.
- Von Furstenberg, G. (1969): "Default Risk on FHA-Insured Home Mortgages as a Function of the Term of Financing: A Quantitative Analysis," *Journal of Finance*, 24 459-477.
- Von Furstenberg, G. (1970): "Interstate Differences in Mortgage Lending Risks: An Analysis of Causes," *Journal of Financial and Quantitative Analysis*, 5, 229-242.
- Von Furstenberg, G. (1970): "The Investment Quality of Home Mortgages," *Journal of Risk and Insurance*, 37 (3), 437-445.
- Von Furstenberg, G. and R. J. Green (1974): "Home Mortgages Delinquency: A Cohort Analysis," *Journal of Finance*, 29, 1545-1548.
- Williams, A.O., W. Beranek, and J. Kenkel (1974): "Default Risk in Urban Mortgages: A Pittsburgh Prototype Analysis," *Journal of the American Real Estate and Urban Economics Association*, 2(2): 101-112.
- Wolinsky, A. (1983): "Retail Trade Concentration Due to Consumers' Imperfect Information," *Bell Journal of Economics*, 14 (1), 275-282.
- Zorn, P. and M. Lea (1989): "Mortgage Borrower Repayment Behavior: A Microeconomic Analysis with Canadian Adjustable Rate Mortgage Data," *Journal of the American Real Estate and Urban Economics Association*, 17 (1), 118-136.