UNIVERSITY OF CALIFORNIA
RIVERSIDE

Predicting The Single-Family 30-Year Fixed-Rate Mortgage Default Rate of Principal Residents

A Thesis submitted in partial satisfaction
of the requirements for the degree of

Master of Business Administration

in

Management

by

Yang Liu

September 2019

Thesis Committee:
Prof. Jean Helwege, Chairperson
Prof. Greg Matthew Richey
Prof. Adem Orsdemir

The Thesis of Yang Liu is approved:

Committee Chairperson

University of California, Riverside

Acknowledgment

I am eternally grateful to Professor Jean Helwege. She is a great professor I have ever met. She is beautiful, patient, warmhearted and always puts students in the first place and solves their puzzles and problems in time. During the writing process, she gives me valuable ideas and teaches me how to solve problems. I cannot accomplish the thesis without her help.

I would like to extend my thanks to my other committee members Professor Greg Matthew Richey and Professor Adem Orsdemir. They give me great suggestions for my thesis and study.

I also like to thank Dean Yunzeng Wang, Heather Anderson and other faculty and staff in A. Gary Anderson Graduate School of Management for their help and suggestions for my study and life.

A very specific thanks to my friends and those who give me understanding and support during my life.

Most importantly, I am grateful to my parents. They give me perpetual love and support, teach me how to work hard and how to be a good person. Without their support, I cannot achieve my dream.

ABSTRACT OF THE THESIS

Predicting The Single-Family 30-Year Fixed-Rate Mortgage Default Rate of Principal Residents

by

Yang Liu

Master of Business Administration, Graduate Program in Management
University of California, Riverside, September 2019
Prof. Jean Helwege, Chairperson

Since 1960, researchers have used proxies, such as debt-to-income ratio, loan-to-value ratio, combined-loan-to-value ratio, and the credit score of borrowers to measure mortgage loan performance. These variables provide a static view of mortgage defaults. However, mortgage defaults must be examined in a dynamic framework. Therefore, I used macroeconomic variables, such as real gross domestic product, the consumer price index, real median household income, interest rates, and the national home price index combined with the static variables to measure the relationships of all these variables with the default rate. I find that debt-to-income ratio, loan-to-value ratio, national home price index, and unemployment rate positively associated with the default rate, while the real gross domestic product and the real median household income negatively associated with default rate. I also find that the real median household income is the most critical macroeconomic factor in predicting the default rate.

Key Words: Single-Family, 30-Year Fixed-Rate Mortgages, Default Rate, Logistic Regression Model.

Table of Content

A List of Figures

A List of Tables

Chapter 1: Introduction

Mortgage loans allow people to own real estate as a home, a shelter, or an investing tool. After receiving a loan, borrowers are required to pay scheduled payments to keep the property and avoid defaulting. However, defaults do occur. If a mortgage loan servicer fails to receive the regular payment on a loan, the borrower is delinquent. Delinquent loans can be categorized as less than 30 days, 30 to 59 days, 60 to 89 days, 90 to 119 days, and more than 119 days past the scheduled payment date. A default is usually considered to occur at the same time as delinquency.

Defaults have consequences. Defaults decrease the credit ratings of borrowers and cause banks to distrust them. It also makes it more difficult for borrowers to receive another mortgage loan. Defaults also have social impacts. Defaults reduce local property values, discourage investment, and increase lending risk in communities.

There are two main reasons defaults occur. First, borrowers are sometimes unable to pay the scheduled payments. In this situation, they are aware that they are facing default, but they are also facing financial hardships and struggling to fulfill their commitments. Divorce, sickness, inability to work, loss of a job, failure in business, and relocation are the primary trigger events for defaulting. These events decrease income and make it difficult for borrowers to pay scheduled payments. From 2007 to 2009, the United States experienced a housing bubble burst and mortgage crisis during which many people defaulted on their mortgages because they had no money to pay for their loans.

The second reason that defaults occur is that people are sometimes unwilling to pay their loans. A mortgage loan is a combination of a call option and a put option. A mortgage default is a put option. From a strictly financial perspective, it is reasonable to default when the put option is in the money. The put option is in the money when the house value is less than the mortgage value (when the borrower is in a position of significantly negative equity). Therefore, borrowers are more likely to default when house values fall even if their incomes have not. A defaulting borrower would benefit from living in the house rent-free until the eventual foreclosure.

Researchers have studied mortgage loan performance since the 1960s. One of the primary areas of consideration is the relationship between trigger events and mortgage defaulting. However, it is difficult to obtain household-level information, and the evidence of what causes a default is limited. Consequently, scholars have used proxy measures to link household-level data with loan-level mortgage performance data. For the first factor affecting defaults (inability to pay), income is a fundamental reason for borrowers to default or not default. Hence, scholars use the debt-to-income ratio at the time of the mortgage origination as a proxy for post-origination income. Actual household unemployment experiences are challenging to obtain, so the local unemployment rate is used as a proxy to explain mortgage defaults. For the second reason for defaulting (unwillingness to pay), it is also difficult to obtain information on negative equity, but researchers can obtain loan-to-value (LTV) information at the mortgage origination date. If the LTV ratio is greater than 1, it is more likely for a rational borrower to default later when house prices decline.

The debt-to-income ratio and LTV ratio are recorded when the mortgage originates and included in typical research datasets, so they provide a static view of mortgage defaulting.

However, this static view is only partially accurate because mortgage loans are paid off continuously via monthly payments. Therefore, mortgage loan performance must be examined in a dynamic and forward-looking framework. Macroeconomic variables, such as real gross domestic product (GDP), real median household income, consumer price index (CPI), and the national home price index (HPI), provide a dynamic complement to static variables.

For a borrower who is unable to pay the scheduled payments and is facing the prospect of defaulting, one way to avoid foreclosure is to sell the property and use the proceeds to repay the loan. Real GDP reflects the economic situation of a country. When real GDP is growing, it is usually easier for borrowers to sell a house and avoid defaulting, whereas when real GDP is declining, it is more difficult for borrowers to sell a house and more likely for them to default. Furthermore, defaulting behavior is influenced by an individual's surroundings, such as friends and family members. A borrower who is in a depressed financial situation may not intend to default but is influenced by friends or family members who are in a better financial situation. The house can be sold to them and the borrower can get the proceeds and repay the loan and avoid defaulting. Hence, real GDP influences borrowers' default behavior. If a family has a higher real median household income, the family has more money to spend, which increases their ability to pay and decreases the probability of defaulting. Higher interest rate means higher monthly payment, and borrowers are more likely to default if their income cannot support the high monthly payment.

Consumer price index measures inflation. An increase in CPI means that all house prices increase along with other expenses in the economy. It also reduces the real cost of loans. These effects increase the equity in the house, so it is more likely that a put option will be in the money and less likely that a borrower will default. The national home price index (HPI) calculates single-family house prices using weighted, repeat-sales transactions. It reflects the trend in house price changes in a dynamic way. Hence, the HPI provides a far-sighted view of mortgage loan default.

In addition to potentially adding greater explanatory power in a regression setting, macroeconomic variables have the advantage that they are easier to forecast than household variables. If a researcher desires to predict the likelihood of default, it is more likely that predictions of macroeconomic factors will be available and individual homeowner variables will not, given that LTV and debt-to-income are rarely available after the date of mortgage origination.

In sum, I predicted that a combination of macroeconomic variables and those static variables would provide a dynamic view of the mortgage default decision and supplement existing research work. Based on the reasons for choosing a combination of variables, I have studied the mortgage default rate using a combination of macroeconomic variables and traditional variables, such as debt-to-income ratio, LTV, CLTV, original interest rate, and the borrower's credit score.

The dataset used is from Fannie Mae. Using this dataset, I built a logistic regression model. The original data were too numerous to process. Therefore, I focused on 30-year fixed-rate mortgages of single-family homes that were bought as principal residences. The study period was from January 1, 2000, to January 1, 2012. The dependent variable was the annual default rate, which was used to estimate the likelihood of default. The explanatory variables were composed of three parts. The first part was information on the mortgage loan at the origination date, such as the debt-to-income ratio, the LTV ratio, the borrower's credit score, and the original interest rate. The second part consisted of macroeconomic variables, such as the real GDP, the CPI, the unemployment rate, the real median household income, and HPI. The third part consisted of interest rates, which include the effective federal funds rate, the 10-year Treasury rate, and the current 30-year fixed-rate mortgage (which was compared to the borrower's original interest rate).

This paper discusses the building of five models that examine the role played by these macroeconomic variables. From the results of these models, I found that the debt-to-income ratio, LTV ratio, and credit scores are reliable proxies for predicting the probability of defaulting. Among the macroeconomic variables, the real GDP is more critical than the CPI because the real GDP is significantly negatively related to the mortgage default rate. Although many researchers believe unemployment is a trigger event to defaulting, this research found that real median household income is more important than the unemployment rate. People do not tend to default when they have high real median household income even when they are unemployed. However, if they have a low real median household income and are unemployed, they are more susceptible to defaulting. The paper also found that the HPI has a positive relationship with defaulting.

By comparing the three kinds of interest rates, I found that the effective federal funds rate is the most crucial interest rate to consider as a proxy. It influences the other two interest rates, and it substantially contributes to mortgage performance.

In sum, I have found that the real GDP, real median household income, HPI, and effective federal funds rate are reliable macroeconomic proxies for measuring the mortgage default rate. Real GDP works well as a proxy in a period when good economic situation falls back into a recession. Real median household income measures income level and has a significant correlation with interest rate. I believe it is the most critical macroeconomic proxy for studying the mortgage default rate.

This article is divided into five chapters: Chapter 1 introduces the objective of the topic; Chapter 2 is the literature review; Chapter 3 discusses methodology and data; Chapter 4 discusses the results; Chapter 5 presents the conclusion and discussion.

Chapter 2: Literature Review

From 1969 until the present, many scholars have discussed loan performance behavior: defaulting, prepaying, and refinancing. Their research can be divided into four areas.

2.1 Loan Performance Under Financial Option Theory

Brennan and Schwartz (1985) analyzed the effects of a call option on an underlying asset and inspired the interest of many researchers. They found that the put–call option theory is an effective way to explain default or prepayment. Hence, the theory has become the primary way of discussing loan performance and borrowers' defaulting behavior. The theory states that when a call option is "in the money," borrowers tend to prepay, while when a put option is "in the money," borrowers tend to default. Schwartz and Torous (1989) used the option theory to build a valuation framework to examine the prepayment experience and value the mortgage-backed securities. Stanton (1995) believed that the value of the call option statistically relates to mortgage termination with refinancing. Yongheng, Della, and Changfeng (2005) assessed residential mortgage performance in China and found that the option theory failed in China, while other non-option methods related to financial-economic factors play significant roles in determining default risks in China. They also found that prepayment behavior strongly related to borrowers' characteristics, so mortgage-lending programs can improve market efficiency and enhance household creditability. Ahlawat (2018) used Monte-Carlo simulation to evaluate the joint put–call option embedded in a mortgage contract. He identified the critical difference between ruthless and non-ruthless mortgage defaults and found that non-ruthless mortgage defaulters are more likely to redefault after 90 days.

2.2 The Factors of Mortgage Defaulting

Green and Shoven (1986) discussed the relationship between the interest rate and mortgage performance. Yongheng (1997) adopted a proportional hazard framework to analyze mortgage risk, and the result showed that prepayment and default behavior could be predicted through a stochastic term structure. Calhoun and Yongheng (2002) confirmed the option theory and found that conditional prepayment probability, original LTV ratio, relative loan size, and other economic and demographic factors should be considered in empirical models. Hong (2010) found that people experience substantial financial loss in a credit crisis, and thus the subprime losses cannot be explained by the traditional mortgage model, and HPI and house price appreciation should be considered. Voicu, Jacob, Renger, and Irene (2012) first examined default factors associated with pre-foreclosure outcomes for subprime mortgages and then examined factors related to different results for loans that enter foreclosure. Schmeiser and Gross (2015) discussed the determinants of subprime mortgage performance. They found that a high LTV ratio significantly contributes to redefault and foreclosure, and any modification intended to increase loan principal is most likely to fail and even controls the profit and loss changes. Chao Yue, Quercia, and Riley (2015) believed unemployment is a trigger event for mortgage defaulting.

2.3 Empirical Mortgage Models

Magee (1968) studied multifamily neighborhood default rates using Cambridge data. She captured both lifecycle and property financial variables and built the first formal commercial loan default model. From 1969, Furstenberg (1969, 1970a, 1970b) studied the annual default rate of single-family loans. His work is regarded as the beginning of modern statistical mortgage modeling. Cambell and Dietrich (1983) combined default and prepayment rates into one empirical model. Berkovec, Canner, Stuart, Gabriel, and Hannan (1994) used an empirical model to reject the theory that default rates are relatively lower among minority borrowers and in minority neighborhoods. Kelly and JR (2001) used the pure options-pricing model to analyze prepayment behavior of borrowers and found that the value of delaying prepayment was higher for mortgages with declining rate penalties than mortgages with static-rate penalties and that higher interest rate spread can trigger refinancing. Anthony (2003) examined 30-year fixed-rate mortgages, compared nonprime and prime loan defaults, and found that many differences exist between nonprime and prime loans. Ali (2017) used the K Nearest Neighbor (KNN) model, logistic regression model, tree-based model, and support vector machines (SVM) to predict mortgage loan default and found that the LTV ratio is the most significant factor in predicting loan performance. Shuyao (2017) built a transition model based on the Markov chain model to estimate default transition probabilities.

2.4 Methods of Solving Defaults

Mills and Lubuele (1994) studied the residential mortgage performance in low-and moderate-income neighborhoods and found that lending programs are perfect for building these communities. Hartarska and Gonzalez (2005) studied the effects of the implemented counseling programs in the Midwest of the United States and found that counseling programs affect the behavior of lenders and borrowers and that the net effect should be evaluated under both prepayment and default.

Although scholars use diverse methods to discuss the factors of default, they focus on the loan-to-value ratio, combined loan-to-value ratio, debt-to-income ratio and borrowers' information, such as credit score. They seldom talk about default rate with other macroeconomic data but their research implies the importance of macroeconomic variables, so the paper studies this area and extends the existing research.

Chapter 3: Methodology and Data

3.1Empirical Model

This paper's main objective is to discover the relationship between macroeconomic variables and mortgage default rate, so models with different variables must be compared. The logistic regression model is the basic model used in the paper. By using different variables in the model, we can determine how macroeconomic variables explain the default rate.

Logistic Regression Model

The logistic regression model uses a logistic function to model a binary dependent variable. A logistic regression model must have a dependent variable with two possible outcomes, such as pass/fail, win/loss, or healthy/sick.

The basic logistic model is:

$$Log(y)=log[y/(1-y)] =\beta_0 + \beta_1 X_1 + \beta_2 X_{2+} \ldots\ldots\ldots\ldots\ldots\ldots \beta n X_n$$

The corresponding odds:

$$O=b^{\beta 0 + \beta 1X1+ \beta 2X2+ \ldots\ldots\ldots\ldots\ldots\ldots \beta nXn}$$ (b is the base of the logarithm)

The probability:

$$P= b^{\beta 0 + \beta 1X1+ \beta 2X2+ \ldots\ldots\ldots\ldots\ldots\ldots \beta nXn} / (b^{\beta 0 + \beta 1X1+ \beta 2X2+ \ldots\ldots\ldots\ldots\ldots\ldots \beta nXn}+1)$$

In this paper:

Traditional model:

Log (default rate) = $\beta_0 + \beta_1$ LTV+β_2 CLTV+β_3 Borrower's Credit Score+ β_4 Original Interest Rate +β_5 Debt-to-income-ratio

The Model with Macroeconomic Variables:

Log (default rate) = $\beta_0 + \beta_1$ Real GDP+β_2 CPI+β_3 Real Median Household Income+β_4 HPI+β_5 Unemployment Rate +β_6 LTV +β_7 Debt-to-income-ratio+β_8 Effective Federal Funds Rate +β_9Borrower's Credit Score

The Model with Macroeconomic Variables and 10-year Treasury Rate:

Log (default rate) = β0 + β1Real GDP+β2CPI+β3Real Median Household Income+β4HPI+β5 Unemployment Rate +β6 LTV +β7 Debt-to-income-ratio+β8 10-Year Treasury Rate+β_9Borrower's Credit Score

The Model with Macroeconomic Variables and 30-Year Fixed-Rate Mortgage Minus the Original Interest Rate:

Log (default rate) = β0 + β1Real GDP+β2CPI+β3Real Median Household Income+β4HPI+β5 Unemployment Rate +β6 LTV +β7 Debt-to-income-ratio+β8 30-Year Fixed-Rate Mortgage minus the original interest rate +β_9Borrower's Credit Score

3.2 Data Source

The original data are from Fannie Mae from January 1, 2000, to January 1, 2012. These data are divided into two parts. One is acquisition data. Fannie Mae records the original data of loans, such as LTV ratio, CLTV ratio, debt-to-income ratio, original interest rate, and borrowers' credit score.

The other portion of the data is performance data. Fannie Mae dynamically records the monthly data of loans, such as the monthly reporting period, current delinquency status of the loan, and the location of the borrower.

A stock index loan ID is linked to combine the two datasets.

Because the performance dataset documents cross-panel data, the number of original data is very large. There are more than 600 million monthly observations and more than 28.8 million mortgage loans. Due to the extensive dataset, I decided to focus on data that meet the following five conditions:

First, a mortgage must not have a delinquency status (current delinquency status=0).The number of these data is very large, and they were omitted from this paper.

Second, the data must focus on the single-family property type. Other types, such as condo, manufactured housing, or co-op housing were not included in the paper.

Third, the data must concentrate on fixed-rate mortgages, so adjusted-rate mortgages were excluded from the paper.

Fourth, the data must be limited to 30-year loan terms and must not be 10-year, 15-year, 25-year, or other terms of mortgage.

Finally, the data must focus on mortgage loans of principal residence at origination date, so mortgage loans of secondary homes and investments were omitted.

From these five conditions, there were 85,258,409 monthly observations, but missing values remained. After removing missing values, 36,417,887 observations and 13,498,409 mortgages were used in the paper, and the SAS 9.4 version was used to calculate the sample data.

Table 1 The Number of Single-Family 30-Year Fixed-Rate Mortgages that bought by the Principal Residence

Origination Year	The Number of Mortgages
2000	787,349
2001	1,645,317
2002	1,664,112
2003	2,453,407
2004	878,153
2005	779,861
2006	594,304
2007	679,807
2008	720,207
2009	1,128,360
2010	673,137
2011	573,624
2012	920,771

3.3 Description of Acquisition Data.

From the acquisition data set, we can see the static data at the first time of the mortgage loan. I extracted eight variables from the data set.

Table 2 Description of Acquisition Dataset

Data Element	Description	Data Type	Values Calculation
Loan id	An index to identify specific loan	Numeric	N/A
Original Interest Rate	The interest rate when first create the mortgage	Numeric	N/A
Original Loan Term	Borrow payments scheduled term	Numeric	360
Origination Date	The date when the mortgage was documented	Date	MM/YYYY
Original Loan-To-Value (LTV)	The loan-to-value ratio at the origination date	Numeric	0%-97%
Original combined Loan-To-Value (CLTV)	The ratio of all secured loans on a property to the value of a property	Numeric	0%-200%
Original Debt-to-income-ratio	Borrowers’ total monthly obligations (including housing expense) by monthly income	Numeric	1%-64%
Borrower Credit Score at Origination	Borrower Credit Score at Origination	Numeric	300-850

3.4 Description of Performance Data set

The performance dataset shows the monthly dynamic data of each mortgage loan until the mortgage loan goes to default or is prepaid. I extracted three variables from the dataset. The three variables are loan ID, monthly reporting period, and current loan delinquency status. Current loan delinquency status measures the number of months the obligor was delinquent for. The value 0 means current or less than 30 days past due, 1 means 30–59 days past due, 2 means 60–89 days past due, 3 means 90–119 days past due, and the sequence continues as such for every 30 days. For the sample dataset, 80% of mortgages have no default behavior and 20% mortgages have default behavior. Fannie Mae uses *X* to represent unknown current loan delinquency status. Because the number of unknown data equals the average of known data, I placed *0* next to those mortgages with unknown current loan delinquency status.

Table 3 Description of Performance Dataset

Data Element	Description	Data Type	Values Calculation
Loan id	An index to identify specific loan	Numeric	N/A
Monthly reporting Period	The cut-off period for mortgage loan information	Date	MM/DD/YYYY
Current Loan Delinquency Status	The number of months the obligor is delinquent as determined by the governing mortgage documents	Alpha-numeric	▪0=current, or less than 30 days past due ▪1= 30-59 days ▪2=60-89 ▪3=90-119 Sequence continues thereafter for every 30 days period ▪X=unknown

3.5 Description of Macroeconomic Variables

All macroeconomic variables are from Federal Reserve Economic Data. Real GDP, CPI, unemployment rate, real median household income, national HPI, effective federal funds rate, 10-year Treasury rate, and 30-year fixed-rate mortgage were used in this research. All variables are of annual frequency. The period was from January 1, 2000, to January 1, 2012.

3.5.1 Real Gross Domestic Product

Real GDP measures the value of all goods and services of a given economic entity in a specific year. It adjusts for inflation, so it more accurately reflects the financial situation of a country. High real GDP reflects the faster development of a nation.

3.5.2 Unemployment Rate

The unemployment rate measures the unemployed labor force in the job market. When the market is strong, the rate is low; when the market is weak, the rate is high. From the work of previous researchers, unemployment is a trigger event to mortgage defaulting.

3.5.3 Real Median Household Income

Real median household income divides household income into two parts: One part is higher than the median level, and the other is less than the median level. The higher the real median household income, the better the household economic situation, and the household can spend more money on loans, autos, and other family expenditures.

3.5.4 Consumer Price Index

The CPI measures the weighted price change of a market. Usually, it uses a specific year as a base and then records the price change of other years for comparison to the base year. The CPI is also an indicator of inflation: Under the same income level, if inflation increases, goods and services cost more. Regarding equity position, higher CPI results in a higher equity position.

3.5.5 The National Home Price Index

The national HPI measures the movement of single-family house prices and records house price trends at geographic levels. The HPI has been proven by researchers to be a valid measurement of the mortgage default rate.

3.5.6 Effective Federal Funds Rate

The Federal Reserve has bank accounts and sets reserve requirements for them. Banks with larger end-of-day balance are usually required to lend funds to other banks and charge an interest rate, which is known as the effective federal funds rate. The effective federal funds rate is one of the most influential interest rates in the U.S. economy because it affects the economic situation, such as employment, growth, and inflation. The rate also influences the mortgage interest rate.

3.5.7 10-Year Treasury Rate

The 10-year Treasury rate is used by the Federal Reserve to determine the index of 10-year Treasury securities. This rate is affected by the federal funds rate and usually determines the mortgage rate by affecting mortgage securities. The 1-year Treasury maturity rate is widely used to determine adjustable-rate mortgages. The 10-year Treasury rate influences 30-year fixed-rate mortgages.

3.5.8 30-Year Fixed-Rate Mortgage

A fixed-rate mortgage is a loan with a fixed interest rate and 30-year fixed-rate mortgages mean the interest rate remains the same for 30 years. The 10-year Treasury rate influences 30-year fixed-rate mortgages.

3.6 Explanatory Variables

For a logistic regression model, the dependent variable must be binary, so I introduced a new variable, *delinquent*, to represent the current delinquent status. If the current delinquent status equals 0 or *unknown*, then *delinquent* = 0; otherwise *delinquent* = 1. The number 0 means there is no default. The number 1 means there is a default. The Fannie Mae performance data record monthly delinquent status, but the objective of this paper is to estimate the annual default rate, thus, I changed the monthly delinquent data to annual delinquent data. Because Fannie Mae records the current delinquency status before loan default or prepayment, the time used in the model is not fixed. If a mortgage defaulted in the same year as the origination year, then the default data are only for 1 year. However, if a mortgage defaulted in 2012, then the period of the loan is 13 years.

Table 4 and 5 show the delinquency binary result from those annual observations.

Table 4 Delinquency Binary Result

Delinquent	Frequency	Percent	Cumulative frequency	Cumulative percent
0	28712813	78.84	28712813	78.84
1	7705074	21.16	36417887	100.00

From 2001 to 2003, the default percent was remained around 15% but increased to nearly 23% from 2007 to the 2012 year.

Table 5 Default Frequency by Year

Year	Default Frequency	Percent
2000	5749	19.48
2001	114,602	15.74
2002	305,833	15.93
2003	386,023	15.09
2004	583,018	20.3
2005	593,273	19.69
2006	660,769	20.96
2007	715,498	21.42
2008	832,298	22.89
2009	900,157	22.95
2010	927,779	22.91
2011	873,440	22.85
2012	806,635	23.91

3.7 Independent variables:

3.7.1 Individual's Measurement

3.7.1.1 Loan-to-Value Ratio

Loan-to-value measures the ratio of the loan to the value of the purchased asset. The higher the ratio, the higher the probability of defaulting. The mean LTV in the sample data is 72.07%.

3.7.1.2 Combined-Loan-to-Value Ratio

The CLTV reflects all loan values to a mortgage value. The mean of the CLTV in the sample is 73.04%.

3.7.1.3 Debt-to-Income Ratio

Debt-to-income ratio is debt payment divided by gross income and shows the relationship between debt and income. If the ratio approaches 1, it means the debt proportion is higher in the overall income. The mean of the debt-to-income ratio in the sample is 34.66%.

3.7.1.4 Borrowers' Credit Score

Borrowers' credit score evaluates the credit rating of borrowers. High credit score usually reflects trustworthiness of borrowers. In the paper, the borrowers' credit score refers to the Fair, Isaac and Company (FICO) score. The mean score in the sample is 723.35.

3.7.1.5 Original Interest Rate

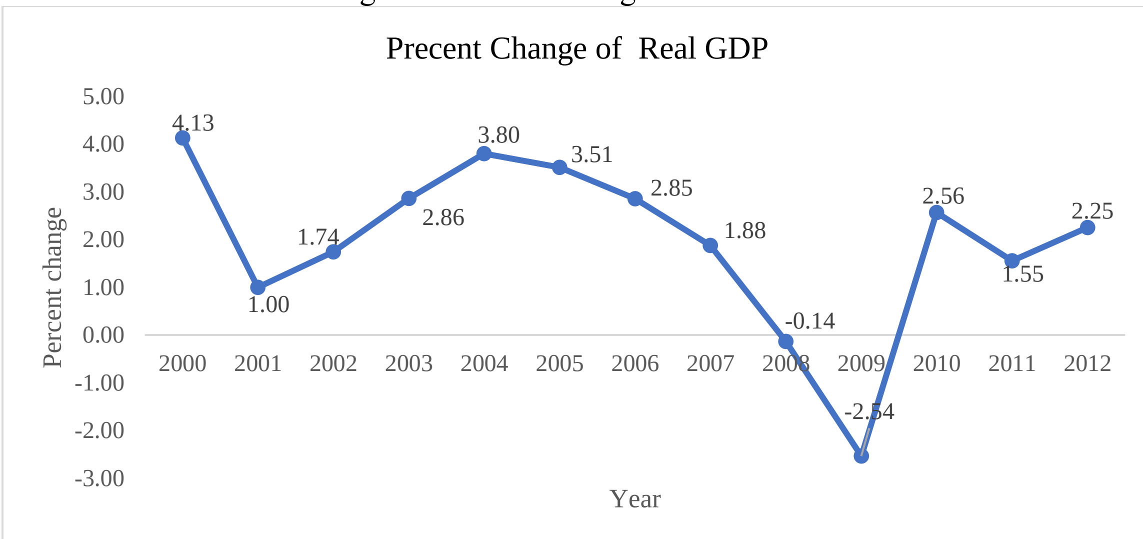
The original interest rate is the interest rate of a mortgage at the origination date. The mean of the original interest rate in the sample is 6.09%.

3.7.2 Macroeconomic Variables

3.7.2.1 Real GDP

The annual percentage change of the real GDP was taken from 2000 to 2012. The mean of the 13 years is 1.96%. The maximum was 4.13% in 2000, and the minimum was -2.54% in 2009.

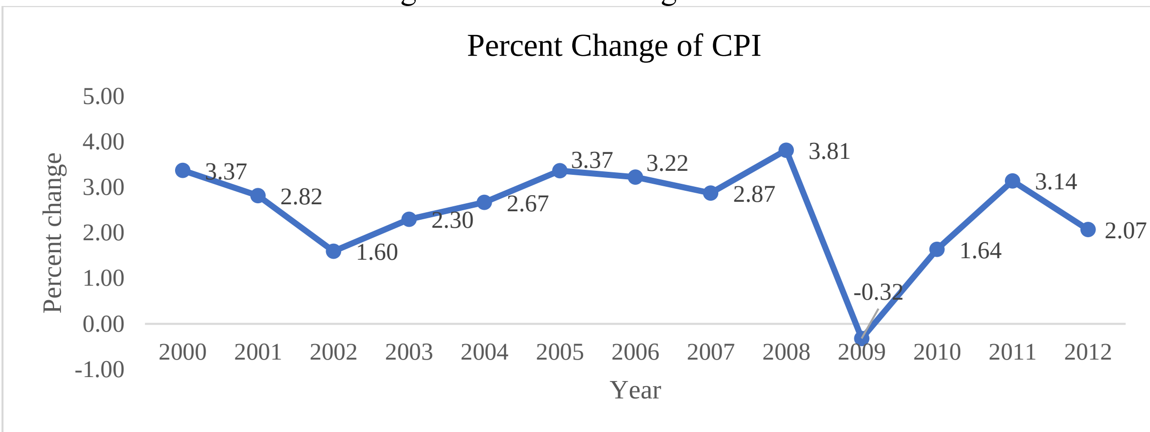
Figure 1 Percent Change of Real GDP



3.7.2.2 Consumer Price Index: Total of All Items for the United States

The annual percentage change of the CPI was also taken from 2000 to 2012. The mean of the 13 years is 2.50%. The maximum was 3.81% in 2008, and the minimum was -0.32% in 2009. After comparing the percentage change of CPI and real GDP in the given sample period, we found that both had dropped to their lowest values in 2009, which confirmed the downward trend of the Great Recession.

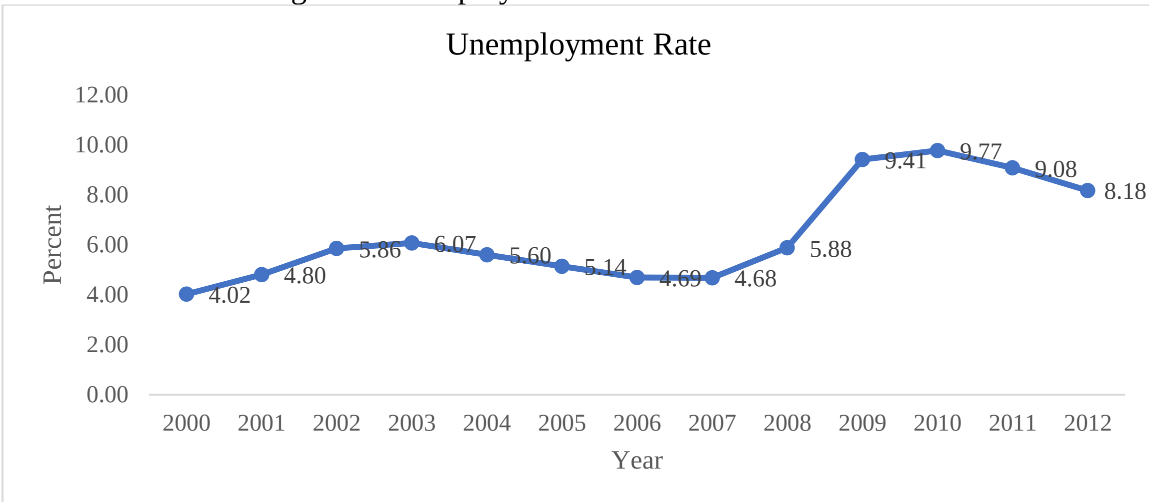
Figure 2 Percent Change of CPI



3.7.2.3 Unemployment Rate: All Persons Aged 15–64 in the United States

Unemployment usually coincides with the business cycle. High economic growth means a healthy labor market and low unemployment rate. Slow economic growth causes a high unemployment rate. From 2008, the unemployment rate increased steadily and peaked in 2010.

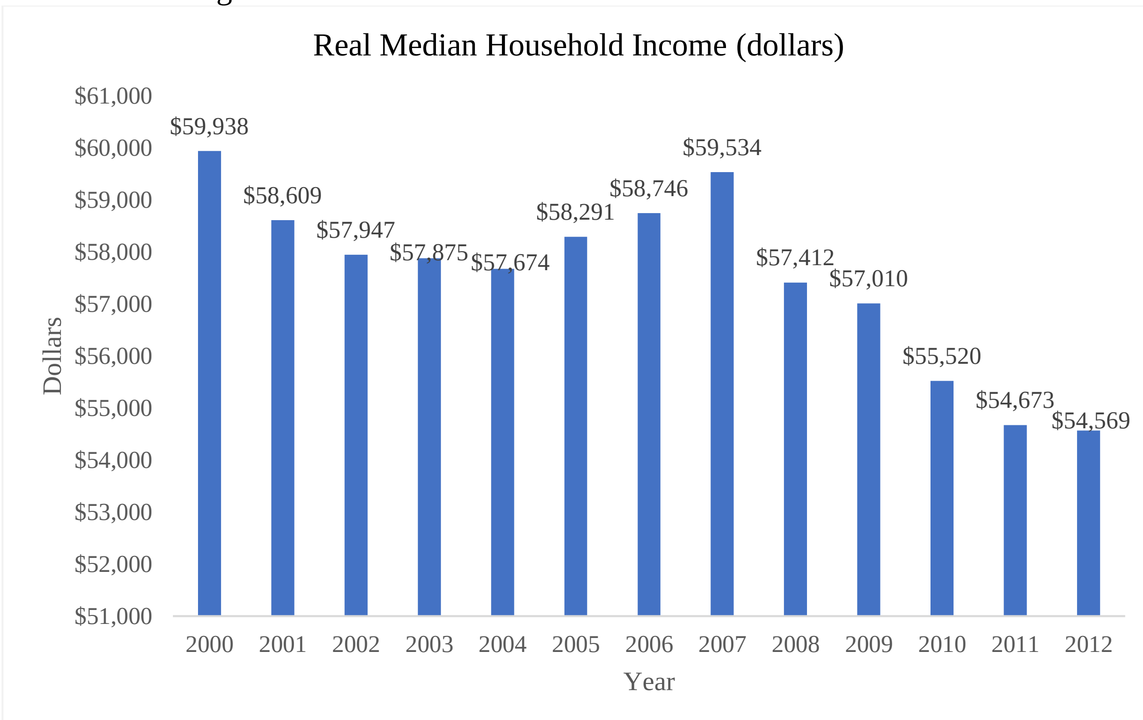
Figure 3 Unemployment Rate from 2000 to 2012



3.7.2.4 Real Median Household Income

The mean of the real median household income from 2000 to 2012 is $57,522.92. From 2000 to 2006, the real median household income was approximately $58,000 but dropped to $54,673 in 2011 and $54,569 in 2012.

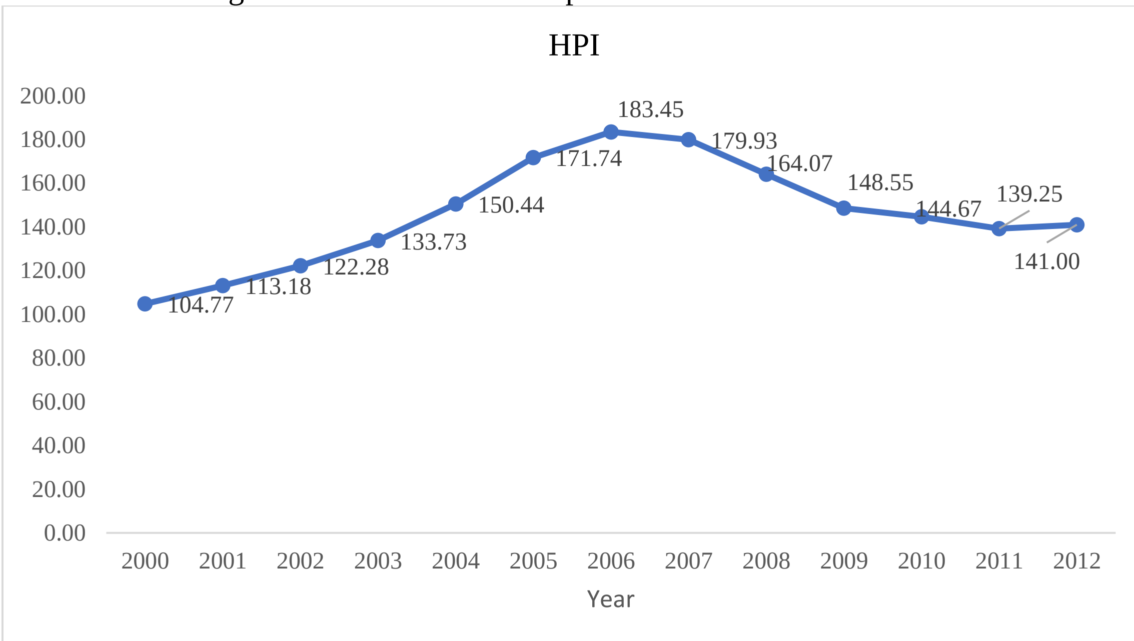
Figure 4 Real Median Household Income from 2000 to 2012



3.7.2.5 National Home Price Index

The S&P/Case–Shiller U.S. National Home Price Index has index value 100 for January 2000. From 2000 to 2006, the HPI continued an increasing trend. The maximum HPI was in 2006 with 183.45.

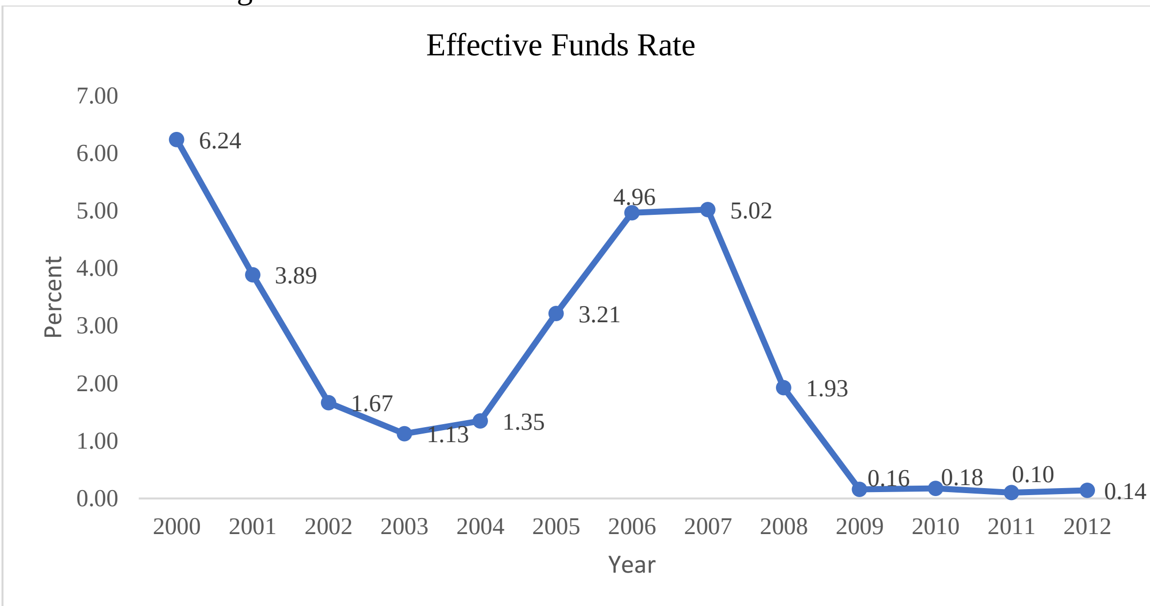
Figure 5 The national home price index from 2000 to 2012



3.7.2.6 Effective Federal Funds Rate

In the year 2000, the effective federal funds rate was high at 6.24% but dropped to 1.13% in 2005 and increased to 5.02% in 2007. Unfortunately, the financial crisis occurred, and the Federal Reserve decreased the rate to 0.16% in 2009 and kept it low until 2012. The low interest rate simulated capital investment and consumption. From 2007 to 2012, inflation was also high. The combination of high inflation with a low-interest-rate environment encouraged consumers to spend money. In this way, the Federal Reserve hoped to stimulate economic recovery.

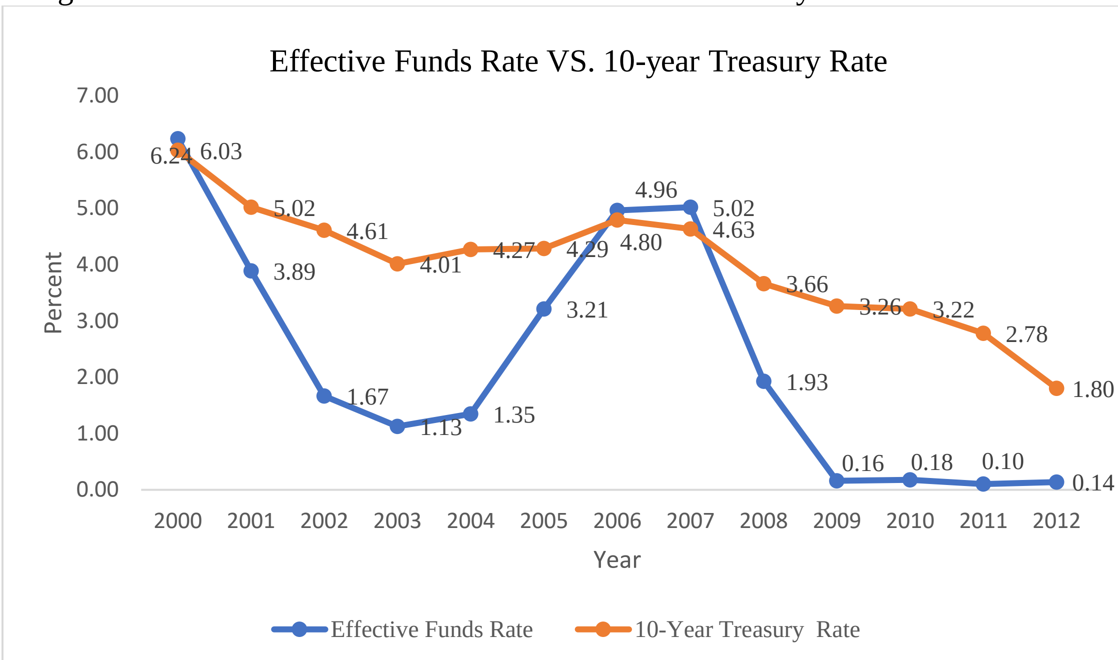
Figure 6 Effective Federal Funds Rate from 2000 to 2012



3.7.2.7 10-Year Treasury Rate

From the comparison of the effective federal funds rate and 10-year Treasury rate, I found that the two rates are positively related. They both decreased between 2000 and 2004, and increased from 2005 to 2007, and then decreased from 2008 to 2012. The effective federal funds rate moved more sharply than the 10-year Treasury rate.

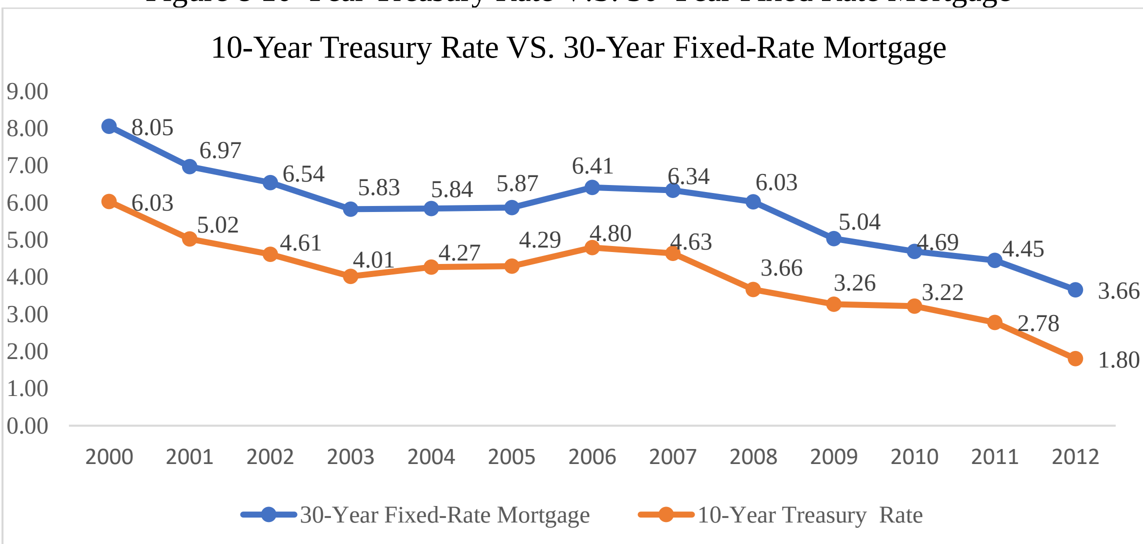
Figure 7 Effective Federal Funds Rate V.S.10-Year Treasury Rate from 2000 to 2012



3.7.2.8 30-Year Fixed-Rate Mortgage

The Figure 8 below shows that the 10-year Treasury rate and 30-year fixed-rate mortgage rate have the same trend, and the 30-year fixed-rate mortgage rate is higher than the 10-year Treasury rate because the 10-year Treasury is a risk-free asset, while the 30-year fixed-rate mortgage is a risky asset and there is a risk premium.

Figure 8 10-Year Treasury Rate V.S. 30-Year Fixed Rate Mortgage



3.8 Descriptive Statistics of Independent Variables

From descriptive statistics of independent variables, I believe that the LTV ratio and CLTV ratio may contribute equally to the logistic model. Furthermore, real GDP and CPI may contribute similarly to the model. In addition, the real GDP, CPI, HPI, and unemployment rate sharply changed during the sample period. I believe that they are good proxies for use in the logistic regression model to predict the probability of defaulting.

.

Table 6 Descriptive Statistics of Independent Variables

	Mean	Standard Error	Minimum	Maximum
Debt-to-income ratio	34.66	12.42	1.00	64.00
Loan-to-value ratio	72.07	16.26	1.00	97.00
Combined-Loan-to value ratio	73.04	16.50	1.00	192.00
Borrowers' score	723.35	57.76	300.00	850.00
Original interest rate	6.09	0.81	1.88	13.50
Real GDP	1.96	0.50	-2.54	4.13
CPI	2.50	0.30	-0.32	3.81
HPI	145.93	6.76	104.77	183.45
Unemployment rate	6.40	0.55	4.02	9.77
Real Median Household Income	57522	471	54569	59938
Effective Funds Rate	2.31	0.59	0.10	6.24
10-Year Treasury Rate	4.03	0.30	1.80	6.03
30-Year Fixed-Rate Mortgage	5.82	0.32	3.66	8.05

Chapter 4 Discussion of result

4.1 Model Comparison

4.1.1 Traditional model

From previous literature, I found that LTV, CLTV, borrowers' credit score, and debt-to-income ratio have been used as proxies to measure mortgage performance. This model keeps the four variables and adds original interest rate to estimate the default rate.

From the traditional model, the likelihood ratio chi-square value of 4,997,748.46 with a p-value of 0.0001 shows that the entire model is significantly better than an empty model. Score and Wald tests were used to test the same hypothesis, and they both show that the model is statistically significant.

Table 7 Validation Test of The Traditional Model

Test	Chi-Square Value	P-value
Likelihood	4997748.46	0.0001
Score	4909347.50	0.0001
Wald	4124182.03	0.0001

By using the logistic model, the paper estimates the situation of *delinquent* =1, which means there exists default behavior. From the result, we can get the model as follows:

Log (default rate) = 7.43+ 0.0188 Debt-to-income ratio +0.00675 LTV+ 0.00323 CLTV +0.0605 original interest rate -0.0148 borrowers' credit score.

From the result, I found that all parameters are significant at p-value=0.0001.

For every unit increase in debt-to-income ratio, the log odds of defaulting increase by 0.0188; for every unit increase in LTV, the log odds of defaulting increase by 0.00675; for every unit increase in borrowers' credit score, the log odds of defaulting decrease by 0.0148; for every unit increase in original interest rate, the log odds of defaulting increase by 0.0605.

In the traditional model, debt-to-income ratio, LTV, CLTV, and original interest rate have a positive relationship with the default rate. The borrowers' credit score negatively associated with mortgage default rate. The higher the borrowers' credit score, the lower the default rate. This result coincides with expectations and outcomes of previous researchers. The original interest rate contributed most to this model, which shows the original interest rate is most important factor to predict default rate. However, this model has a static viewpoint, so I added macroeconomic variables to see how they contribute to the mortgage default rate.

Table 8 Parameter of The Traditional Model

Parameter	**Estimate**	**Standard Error**
Intercept	7.4349***	0.00772
Debt-to-income-ratio	0.0188***	0.000036
LTV	0.00675***	0.000114
CLTV	0.00323***	0.000112
Original Interest Rate	0.0605***	0.000585
Borrower's credit score	-0.0148***	8.186E-6

Note: ***represent significant levels of 1%

4.1.2 The Model with Macroeconomic Variables

In the traditional model, the LTV and CLTV have the same relationship with default rate, but the CLTV has a smaller parameter, so I used LTV instead of CLTV in the next models. In the first model with macroeconomic variables, I used debt-to-income ratio, LTV, borrower's credit score, real GDP, CPI, unemployment rate, HPI, real median household income, and effective federal funds rate.

Table 9 Validation Test of The First Model with Macroeconomic Variables

Test	Chi-Square Value	P-value
Likelihood	5345829.14	0.0001
Score	5231273.59	0.0001
Wald	4353678.36	0.0001

For the first model with macroeconomic variables, the paper also estimates the situation of *delinquent* = 1, which means there exist defaults. The model is significant, and the parameter is calculated as follows:

Log (default rate) = 17.86 + 0.0175 Debt-to-income-ratio + 0.0121 LTV - 0.0155 borrowers' credit score - 0.0332 Real GDP - 0.0268 CPI + 0.0155 Unemployment Rate + 0.0114 HPI - 0.00020 Real Median Household Income + 0.0480 Effective Federal Funds Rate

For every unit increase in debt-to-income ratio, the log odds of defaulting increase by 0.0175; for every unit increase in LTV, the log odds of defaulting increase by 0.0121; for every unit increase in borrowers' credit score, the log odds of defaulting decrease by 0.0155; for every percentage increase in real GDP, the log odds of defaulting decrease by

0.0332; for every percentage increase in CPI, the log odds of defaulting decrease by 0.0268; for a unit increase in the unemployment rate, the log odds of defaulting increase by 0.0155; for a unit increase in HPI, the log odds of defaulting increase by 0.0114; for a unit increase in real median household income, the log odds of defaulting decrease by 0.00020; for a unit increase in effective federal funds rate, the log odds of defaulting increase by 0.0480.

Comparing the first model that includes macroeconomic variables with the traditional model, I found that debt-to-income ratio, LTV, and borrowers' credit score are in accordance with the traditional model, which means the three variables are significant indicators in estimating default rate under a macroenvironment.

The real GDP, CPI, and real median household income have a negative relationship with the mortgage default rate. When the economy is growing, borrowers are inclined to refinance and avoid default. A high CPI shows that the inflation is also high, and as a result, the value of equity in the home increases and decreases the probability of becoming negative equity. As real median household income increases, borrowers have more money to spend and a higher likelihood of paying the monthly scheduled payment on time and avoiding default.

The unemployment rate, HPI, and effective federal funds rate have a positive relationship with the mortgage default rate. The unemployment rate is composed of two main aspects: being laid off and being unable to work. In a good economic situation, the unemployment rate is low. When the economy experiences a recession, companies and other institutions must lay off employees to continue operating. The laid-off workers lose income, and it becomes difficult for them to find new jobs due to economic recession. Under this situation, borrowers must default because they have no money to pay the mortgage. Inability to work is the other aspect of unemployment. When borrowers begin a mortgage loan, they may be sure that they can pay the loan payments on time. However, accidents happen. Borrowers lose the ability to work and must default due to a shortage in money. The HPI is a predictor of house prices. A higher house index causes borrowers to pay more to the loan and causes a higher probability of defaulting. A low interest rate means a cheaper cost to borrowers, which then increases demand. The high demand increases house prices, improves the household equity position, allows people to borrow more, and increases the debt amount. A low interest rate usually follows a high interest rate in a business cycle. The high interest rate boosts house prices, decreases demand, and lowers the equity position, resulting in positive equity becoming negative. The combination of low income or no income, high monthly payment, and a negative equity position causes a default.

Table 10 Parameter of The First Model with Macroeconomic Variables

Parameter	Estimate	Standard Error
Intercept	17.8629***	0.0579
Debt-to-income-ratio	0.0175***	0.000037
LTV	0.0121***	0.000030
Borrower’s credit score	-0.0155***	8.02E-6
Real GDP	-0.0332***	0.000318
CPI	-0.0268***	0.000772
Unemployment Rate	0.0155***	0.000807
HPI	0.0114***	0.000043
Real Median Household Income	-0.00020***	9.224E-7
Effective Federal Funds Rate	0.0480***	0.000736

Note: ***represent significant levels of 1%

4.1.3 The Model with Macroeconomic Variables and 10-year Treasury Rate

In the first model with macroeconomic variables, the real GDP and CPI have a negative relationship with the default rate and a similar estimated parameter, so I created a correlation table to see the relationship among these macroeconomic variables.

Table 11 Correlation of All Macroeconomic Variables

	Real GDP	CPI	Unemployment rate	HPI	Real Median Household Income	Effective Funds Rate	10-Year Treasury Rate	30-Year Fixed-Rate Mortgage
Real GDP	1.00							
CPI	0.55	1.00						
Unemployment Rate	-0.46	-0.62	1.00					
HPI	-0.06	0.18	-0.04	1.00				
Real Median Household Income	0.23	0.31	-0.88	0.04	1.00			
Effective Federal Funds Rate	0.38	0.55	-0.86	0.07	0.87	1.00		
10-Year Treasury Rate	0.35	0.36	-0.83	-0.19	0.93	0.85	1.00	
30-Year Fixed-Rate Mortgage	0.26	0.39	-0.84	-0.24	0.92	0.84	0.98	1.00

The correlation table of all macroeconomic variables shows that real GDP and CPI have a significant positive relationship. Real median household income negatively associated with the unemployment rate. This means when real median household income is higher, the unemployment rate is lower, and vice versa. The effective federal funds rate significantly relates to CPI and real median household income. Because the 10-year Treasury rate and 30-year fixed-rate mortgage are influenced by the effective federal funds rate, it also has a significant relationship with real median household income.

In the second macroeconomic variables model, I used a 10-year Treasury rate instead of the effective federal funds rate and omitted real median household income due to the high correlation, and I left other variables unchanged.

The validation test shows that this model is also significant.

Table 12 Validation Test of The Second Model with Macroeconomic Variables

Test	Chi-Square Value	P-value
Likelihood	5332886.41	0.0001
Score	5218790.49	0.0001
Wald	4345873.56	0.0001

The logistical model is estimated as follows:

Log(default) = 6.39+ 0.0175 Debt-to-income-ratio +0.0121LTV -0.0154 borrowers' credit score-0.00847 Real GDP +0.0512 CPI+ 0.0898 Unemployment Rate+ 0.0104 HPI -0.1644 10-Year Treasury Rate

For every unit increase in debt-to-income-ratio, the log odds of defaulting increases by 0.0175; for every unit increase in LTV, the log odds of defaulting increases by 0.0121; for every unit increase in borrowers' credit score, the log odds of defaulting decreases by 0.01554; for every percentage increase in real GDP, the log odds of defaulting decrease by 0.00847; For every percentage increase in CPI, the log odds of defaulting increase by 0.0512; for a unit increase in the unemployment rate, the log odds of defaulting increase by 0.0898; for a unit increases in HPI, the log odds of defaulting increases by 0.0104; for a unit increase in 10-year Treasury rate, the log odds of defaulting decrease by 0.1644.

Table 13 Parameter of The Second Model with Macroeconomic Variables (no real median household income)

Parameter	Estimate	Standard Error
Intercept	6.3922***	0.0105
Debt-to-income-ratio	0.0175***	0.000037
LTV	0.0121***	0.000030
Borrower's credit score	-0.0154***	8.012E-6
Real GDP	-0.00847***	0.000302
CPI	0.0512***	0.000587
Unemployment Rate	0.0898***	0.000488
HPI	0.0104***	0.000032
10-Year Treasury Rate	-0.1644***	0.000872

Note: ***represent significant levels of 1%

In the second macroeconomics model, the debt-to-income ratio, LTV, unemployment rate, and HPI have the same trend as the first model and positively relate to the default rate. The newly added variable, the 10-year Treasury rate, has a negative relationship with the mortgage default rate. However, the result is somewhat strange because the 10-year Treasury rate is influenced by the effective federal funds rate, and the two variables should have the same trend with the default rate. Subsequently, I added the real median household income back into the model, and the logistic regression result is the following:

Log (default rate) = 16.63 + 0.0176 Debt-to-income-ratio + 0.0121LTV - 0.0154 borrowers' credit score - 0.0295 Real GDP - 0.0219CPI + 0.004 Unemployment Rate + 0.0131HPI - 0.00018 Real Median Household Income + 0.00504 10-Year Treasury rate

In the modified model, the real median household income has a negative relationship with the mortgage default rate, and the 10-year Treasury rate shows a positive relationship. The result shows that real median household income is a significant macroeconomic variable in predicting the default rate. The result also shows the estimated parameter of unemployment rate increases to 0.09 after deleting real median household income but decreases to 0.004 after returning the real median household income, which proves that the real median household income is more important than the unemployment rate in measuring household income ability and predicting the default rate.

Table 14 Parameter of The Second Model with Macroeconomic Variables

Parameter	Estimate	Standard Error
Intercept	16.6300***	0.1103
Debt-to-income-ratio	0.0176***	0.000037
LTV	0.0121***	0.000030
Borrower's credit score	-0.0154***	8.015E-6
Real GDP	-0.0295***	0.000377
CPI	-0.0219***	0.000977
Unemployment Rate	0.00400***	0.00104
HPI	0.0131***	0.000043
Real Median Household Income	-0.00018***	1.968E-6
10-Year Treasury Rate	0.00504***	0.00201

Note: ***represent significant levels of 1%

4.1.4 The Model with Macroeconomic Variables and 30-Year Fixed-Rate Mortgage Minus the Original Interest Rate

The leading sellers of the 30-year fixed-rate mortgage are Wells Fargo, J.P. Morgan Chase, and Bank of America. The 30-year fixed-rate mortgage is often chosen because the scheduled monthly payment is fixed and less than that for 10-year mortgages, 15-year mortgages, and other fixed mortgages. Also, less monthly payment enables borrowers to obtain a more expensive mortgage. One way for borrowers to avoid defaulting is being willing to refinance. The net present value of mortgages is a rule for refinancing that is recognized by most scholars. Most borrowers prefer to refinance with a lower interest rate than a higher interest rate. However, researchers believe borrowers should refinance if differences in interest rates is higher. In the next model, I used the difference between 30-year fixed-rate mortgage and original interest rate to measure the interest-rate volatility and to see if it is a useful proxy to study the default rate.

Table 15 Validation Test of The Third Model with Macroeconomic Variables

Test	Chi-Square Value	P-value
Likelihood	5570714.26	0.0001
Score	5474008.69	0.0001
Wald	4507338.90	0.0001

The Likelihood test with Chi-Square Value of 5570714.26 and P-value of 0.0001, shows that the model is significant than an empty model.

The logistical model is estimated as follows:

Log (default rate) = 5.59+ 0.0168 Debt-to-income-ratio +0.01103 LTV -0.0145 borrowers' credit score-0.0385 Real GDP+0.0781CPI+0.0648 Unemployment Rate+0.0168 HPI -0.00002 Real Median Household Income -0.3299 30-year fixed rate mortgage minus original interest rate

For every unit increase in debt-to-income-ratio, the log odds of defaulting increases by 0.0168; for every unit increase in LTV, the log odds of defaulting increase by 0. 01103; for every unit increase in borrowers' credit score, the log odds of defaulting decreases by 0.0145; for every percentage increase in real GDP, the log odds of defaulting decrease by 0.0385; for every percentage increase in CPI, the log odds of defaulting increase by 0.0781; for a unit increase in the unemployment rate, the log odds of defaulting increase by 0.0648; for a unit increases in HPI, the log odds of defaulting increase by 0.0168; for a unit increase in real median household income, the log odds of defaulting decrease by 0.00002; for a unit increases in the difference between 30-year fixed-rate mortgage and original interest rate, the log odds of defaulting decreases by 0.3299.

Table 16 Parameter of The Third Model with Macroeconomic Variables

Parameter	Estimate	Standard Error
Intercept	5.5936***	0.0581
Debt-to-income-ratio	0.0168***	0.000037
LTV	0.0103***	0.000030
Borrower's credit score	-0.0145***	8.205E-6
Real GDP	-0.0385***	0.000313
CPI	0.0781***	0.000799
Unemployment Rate	0.0648***	0.000811
HPI	0.0168***	0.000036
Real Median Household Income	-0.00002***	9.127E-7
The difference between 30-year fixed-rate mortgage and original interest rate	-0.3299***	0.000690

Note: ***represent significant levels of 1%

In the model, debt-to-income ratio, LTV, borrower's credit score, real GDP, unemployment rate, HPI, and real median household income have the same relationship as in the previous models, while CPI changed from a negative correlation to a positive correlation to mortgage defaulting. The difference between 30-year fixed-rate mortgage and the original interest rate shows a negative association with default rate. The negative relationship shows that difference in interest rates can trigger borrowers to refinance and avoid default. The combination with CPI and the interest volatility can explain the change in CPI. Higher inflation causes equity prices increase and borrowers tend to borrow more to obtain the loan. However, once the inflation decreases, this positive equity becomes negative equity, and it is difficult for borrowers to refinance in a poor economic situation, resulting in an increased default rate.

4.1.5 The Model with Macroeconomic Variables by Period

The subprime mortgage crisis occurred in 2007 and developed into a worldwide financial crisis and a global economic downturn. I divided the sample period into two periods: before the financial crisis and after the financial crisis.

Phase 1: 2000 to 2006

In March 2000, the burst of the stock market bubble occurred in NASDAQ, decreasing real GDP to the lowest level in 2001. To recover the economy, the Federal Reserve dramatically dropped the effective federal funds rate and kept it low until the year 2004. Economic recovery began in 2001 and increased until 2005. After the economy declined in 2001, the unemployment rate increased until 2003 and then began decreasing until 2006. In this period, the whole economy was in recovery, and many people borrowed mortgage loans in the period because of the low effective federal funds rate and relatively small 30-year fixed-rate mortgages.

In the correlation table, the real GDP and CPI significantly correlates, so I omitted CPI from the following models and kept other variables.

The validation test shows that that the model is significant.

Table 17 Validation Test of The Fourth Model with Macroeconomic Variables

Test	Chi-Square Value	P-value
Likelihood	1511188.54	0.0001
Score	1516541.75	0.0001
Wald	1304453.89	0.0001

The logistical model is estimated as follows:

Log (default rate) = 39.17+ 0.0096 Debt-to-income-ratio +0.0087 LTV -0.0136 Borrowers' credit score-0.0312 Real GDP-0.0473 Unemployment Rate+0.00933 HPI -0.00053 Real Median Household Income -0.1214 30-year fixed rate mortgage minus original interest rate

For every unit increase in debt-to-income-ratio, the log odds of defaulting increase by 0.01096; For every unit increase in LTV, the log odds of defaulting increase by 0. 0087; For every unit increase in borrowers' credit score, the log odds of defaulting decrease by 0.0136; for every percentage increase in real GDP, the log odds of defaulting decrease by 0.0312; for a unit increase in unemployment rate, the log odds of defaulting decrease by 0.0473; for a unit increase in HPI, the log odds of defaulting increase by 0.00933; for a unit increase in real median household income, the log odds of defaulting decrease by 0.00053; for a unit increase in the difference between 30-year fixed rate mortgage and original interest rate, the log odds of default decrease by 0.1214.

Table 18 Parameter of The Fourth Model with Macroeconomic Variables

Parameter	Estimate	Standard Error
Intercept	39.1726***	0.3860
Debt-to-income-ratio	0.00957***	0.000060
LTV	0.00867***	0.000050
Borrower's credit score	-0.0136***	0.000013
Real GDP	-0.0312***	0.00250
Unemployment Rate	-0.4773***	0.00391
HPI	0.00933***	0.000099
Real Median Household Income	-0.00053***	6.445E-6
The difference between the 30-year fixed-rate mortgage and original interest rate	-0.1214***	0.00112

Note: ***represent significant levels of 1%

In the model of 2000 to 2006, the debt-to-income ratio, LTV, borrowers' credit score, real GDP, HPI, and real median household income maintain the same relationship with default rate, while unemployment rate changes from a positive to a negative association with default rate. In this phase, the real median household income parameter increases from -0.00002 to -0.00053, which shows that the real median household income has a more significant contribution to this model than previous models. The change in unemployment rate from the default relationship can be explained by the following reasons.

First, the economy of the United States has grown from 1992 to 2006, and people had relatively high real median household income during this period. Although a decline happened in 2001, the economy soon recovered. Since people have relatively high real median household income, they can spend more on monthly payments, and decrease the probability of defaulting.

Second, although many people were laid off in this period, they received relatively high compensation, which supported them in paying loans on time.

The model demonstrates that unemployment is a trigger event to defaulting. However, if an income can support a borrower in paying a loan, the borrower is more likely to avoid a default even in an unemployment situation.

Phase 2: 2007 to 2012

From 2002 to 2004, the Federal Reserve had lowered the effective federal funds rates and made mortgage loans more attractive to borrowers, especially in the subprime mortgage market, making mortgage loans available to people who previously could not pay. From 2005, the Federal Reserve continued increasing the effective federal funds rates and sparked a subprime mortgage default crisis, which developed into a global financial crisis. During the period, the real GDP sharply dropped to the lowest level in 2008 and increased the unemployment rate. The unemployment rate peaked in 2010. The Federal Reserve kept the effective federal funds rate at a low level, approximately 0.1 to 0.2 from 2009 to 2012, to simulate the economy.

In this model, I used the same variables as in the previous model, and the results show that the model is significant.

Table 19 Validation Test of The Fifth Model with Macroeconomic Variables

Test	Chi-Square Value	P-value
Likelihood	4093395.17	0.0001
Score	3954215.27	0.0001
Wald	3170757.29	0.0001

The logistical model is estimated as follows:

Log (default rate) = 20.88+ 0.0206 Debt-to-income-ratio +0.0115 LTV -0.0150 borrowers' credit score-0.0957 Real GDP+0.2334 Unemployment Rate+0.1088 HPI -0.00056 Real Median Household Income -0.4560 30-year fixed rate mortgage minus original interest rate

For every unit increase in debt-to-income-ratio, the log odds of defaulting increase by 0.0206; for every unit increase in LTV, the log odds of defaulting increase by 0. 0115; for every unit increase in borrowers' credit score, the log odds of defaulting decrease by 0.0150; for every percent increase in Real GDP, the log odds of defaulting decrease by 0.0957; for a unit increase in unemployment rate, the log odds of defaulting increase by 0.2334; for a unit increase in HPI, the log odds of defaulting increase by 0.1088; for a unit increase in real median household income, the log odds of defaulting decrease by 0.00056; for a unit increase in the difference between 30-year fixed rate mortgage and original interest rate, the log odds of defaulting decrease by 0.4560.

Table 20 Parameter of The Fifth Model with Macroeconomic Variables

Parameter	Estimate	Standard Error
Intercept	20.8813***	0.1285
Debt-to-income-ratio	0.0206***	0.000047
LTV	0.0115***	0.000038
Borrower's credit score	-0.0150***	0.000011
Real GDP	-0.0957***	0.000799
Unemployment Rate	0.2334***	0.00197
HPI	0.1088***	0.000747
Real Median Household Income	-0.00056***	4.484E-6
The difference between the 30-year fixed-rate mortgage and original interest rate	-0.4560***	0.000901

Note: ***represent significant levels of 1%

In this model, the debt-to-income ratio, LTV, unemployment rate, and HPI positively relate to the default rate. Borrowers' credit score, real GDP, real median household income, and the difference between average 30-year fixed-rate mortgage and the original interest rate have a negative relationship with default rate.

In this period, real GDP had a significant contribution to defaulting compared with the previous models. This is because the real GDP showed a considerable change in the period, and the change significantly affected the model's result. The unemployment rate had a positive relationship with defaulting in this period. The real median household income decreased from 2007, and the severe financial recession made it more challenging to find jobs during this time. Without a reliable income, borrowers are more likely to default.

4.2 Results

In the first model, I used five static variables to build the model and found that they indeed contribute to the default rate. In the second model, I combined the traditional model's variables with macroeconomic variables. In the third model, I used the 10-year Treasury rate instead of the effective federal funds rate and left other variables unchanged. In the fourth model, I used the difference between the average 30-year fixed-rate mortgage and the original interest rate instead of the 10-year Treasury rate and left other variables unchanged. Then, I divided the period into two periods to see the contribution of macroeconomic variables in different periods.

The results show that all models are significantly better than an empty model, which means the findings of the entire paper are valid. By comparing the five models, I found that debt-to-income ratio, LTV, and HPI have a positive relationship with the default rate. Borrowers' credit score, real GDP, and real median household income have a negative correlation with the default rate.

Chapter 5 Conclusion and Discussion

5.1 Conclusion

Debt-to-income ratio measures the proportion of debt to income. The larger the ratio, the higher the probability of defaulting. The LTV calculates the loan for a purchased asset. The higher the ratio, the higher the possibility of defaulting. The HPI is a measure of the house price. As a house price increases, borrowers are more likely to default, and vice versa. A borrower's credit score reflects the trustworthiness of the borrower. The higher the borrower's credit score, the lower the default rate.

Real GDP has significant negative relationship with default rate. As the fifth model shows, one percentage change increase in real GDP, the default rate decrease by 0.1%. Real median household income seldom appeared in the previous work of scholars, but this paper shows that it is an excellent proxy to measure the default rate. When borrowers have a higher real median household income, they are more likely to keep a property even when they are facing unemployment situations. However, if the real median household income is low, and borrowers have no other income sources, then unemployment can be a significant event to trigger a default. The CPI is an indicator of inflation, and it works similarly to real GDP in most situations, but it is a proxy that must be combined with other variables to demonstrate its role in predicting the mortgage default rate.

I also used three kinds of interest rates to measure their relationship with the default rate in a dynamic way. The first rate is the effective federal funds rate, and it is an excellent proxy to predicting defaults. The rate has a positive relationship with mortgage default. A low rate encourages people to consume and spend money. From 2002 to 2004, the rate was low, and the number of mortgages was high, which shows that a flat rate encourages more people to buy houses. However, after the Federal Reserve increased the rate, house prices and monthly payments increased, and borrowers defaulted due to having no money for monthly payments. The 10-year Treasury rate also has a negative relationship with default rate. The difference between 30-year fixed-rate mortgage and the original interest rate has a negative relationship with the default rate. The negative relationship shows that interest rate volatility is a useful proxy to predict default rate and it is also a useful policy tool for controlling mortgage performance.

Based on these models, I found that real GDP, HPI, real median household income, and the effective federal funds rate are good proxies in predicting the mortgage default rate. Although the estimated parameter of real median household income is small compared with the other three variables, the real median household income always has a negative relationship with the default rate. Moreover, the real median household income also has a high correlation with interest rate, so I believe it is an essential proxy among other macroeconomic variables in predicting the default rate.

5.2 Further Research

5.2.1 Because Fannie Mae records the location of mortgage borrowers in the performance dataset, I should gather 20 city HPIs to explore the default rate of the leading 20 cities and the relationship between the macroeconomies of these 20 cities in the future.

5.2.2 In the paper, I used only three variables from the performance data, but I hope to use more variables under a reasonable limited storage environment.

5.2.3 In the models discussed in this paper, I used only the logistic regression model to predict the default rate, but there are many other methodologies that can be used to predict the default rate, and I should use diverse methods to study this rate.

References

[1] Brennan, M. J., & Schwartz, E. S. (1985). Determinants of GNMA Mortgage Prices. Real Estate Economics, 13(3), 209–228.

[2] Schwartz, E. S., & TOROUS, W. N. (1989). Prepayment and the Valuation of Mortgage-Backed Securities. The Journal of Finance, 44(2), 375–392

[3] Stanton, R. (1995). Rational Prepayment and the Valuation of Mortgage-Backed Securities, The Review of Financial Studies 8(3), 677-708.

[4] Yongheng D., Della Z., & Changfeng L. (2005) An Early Assessment of Residential Mortgage Performance in China. The Journal of Real Estate Finance and Economics, 31(2), 117–136.

[5] Ahlawat, S. (2018b). Evaluation of Mortgage Default Characteristics Using Fannie Mae's Loan Performance Data. The Journal of Real Estate Finance and Economics,1-28.

[6] Voicu, I., Jacob, M., Rengert, K., & Fang, I. (2011). Subprime Loan Default Resolutions: Do They Vary Across Mortgage Products and Borrower Demographic Groups? The Journal of Real Estate Finance and Economics, 45(4), 939–964.

[7] Schmeiser, M. D., & Gross, M. B. (2015). The Determinants of Subprime Mortgage Performance Following a Loan Modification. The Journal of Real Estate Finance and Economics, 52(1), 1–27.

[8] Tian, C. Y., Quercia, R. G., & Riley, S. (2015). Unemployment as an Adverse Trigger Event for Mortgage Default. The Journal of Real Estate Finance and Economics, 52(1), 28–49.

[9] Hong, C. H. T. (2010). Dynamic Econometric Loss Model: A Default Study of US Subprime Markets. Handbook of Quantitative Finance and Risk Management, 779–805.

[10] Calhoun, C. A., & Deng, Y. (2002). A Dynamic Analysis of Fixed- and Adjustable-Rate Mortgage Terminations. The Journal of Real Estate Finance and Economics, 24(1/2), 9–33.

[11] Green, J., & Shoven, J. B. (1986). The Effects of Interest Rates on Mortgage Prepayments. Journal of Money, Credit and Banking, 18(1), 41-50.

[12] DENG, Y. (1997). Mortgage Termination: An Empirical Hazard Model with Stochastic Term Structure. The Journal of Real Estate Finance and Economics, 14(3), 309–331.

[13] Von Furstenberg, G. M. (1969). Default Risk On FHA-Insured Home Mortgages as a Function of the Terms of Financing: A Quantitative Analysis. The Journal of Finance, 24(3), 459–477.

[14] Von Furstenberg, G. M. (1970a). Quantitative Analysis of Financial Decisions. The Journal of Finance 25(3):715

[15] Von Furstenberg, G. M. (1970b). Interstate Differences in Mortgage Lending Risks: An Analysis of The Causes. Journal of Financial and Quantitative Analysis 5(02):229-242

[16] Pennington-Cross, A. (2003). Credit History and the Performance of Prime and Nonprime Mortgages. The Journal of Real Estate Finance and Economics, 27(3), 279–301.

[17] Campbell, T. S., & Dietrich, J. K. (1983). The Determinants of Default on Insured Conventional Residential Mortgage Loans. The Journal of Finance, 38(5), 1569-1581.

[18] Berkovec, J. A., Canner, G. B., Gabriel, S. A., & Hannan, T. H. (1994). Race, redlining, and residential mortgage loan performance. The Journal of Real Estate Finance and Economics, 9(3), 263–294.

[19] Kelly, A., & Slawson, Jr., V. C. (2001). Time-Varying Mortgage Prepayment Penalties. The Journal of Real Estate Finance and Economics, 23(2), 235–254.

[20] Bagherpour A. (2017). Predicting Mortgage Loan Default with Machine Learning Methods. University of California, riverside.

[21] Shuyao Y. (2017). Mortgage Transition Model Based on Loan Performance Data. Washington University in ST. LOUIS. Department of Mathematics.

[22] Mills, E. S., & Lubuele, L. S. (1994). Performance of residential mortgages in low and moderate-income neighborhoods. The Journal of Real Estate Finance and Economics, 9(3), 245–260.

[23] Schmeiser, Maximilian D., and Matthew B. Gross. (2015). The Determinants of Subprime Mortgage Performance Following a Loan Modification. Finance and Economics Discussion Series, 1-37.