Spring 5-19-2017

Mortgage Transition Model Based on LoanPerformance Data

Shuyao Yang
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST.LOUIS

Department of Mathematics

Mortgage Transition Model Based on LoanPerformance Data

By

Shuyao Yang

A thesis presented to

The Graduate School

of Washington University in

partial fulfillment of the

requirements for the degree

of Master of Arts

May 2017

Saint Louis, Missouri

Table of Contents

List of Figures

List of Tables

Acknowledgments

I would like to sincerely thank Prof. Jimin Ding for her kindness and patience. She always provides me a lot of valuable ideas and suggestions when I was in trouble. It would be impossible for me to accomplish this thesis without her help and guidance.

I would also like to thank Prof. Todd Kuffner, Prof. Renato Feres, Prof. N.Mohan Kumar, Prof. Edward Spitznagel, Mary Ann Stenner and other faculty and staff in math department for their kindly help and suggestion to my life, study, thesis and research.

I would also like to thank Yiyao Luo, Xinwei Zhang, Ying Zhuang and other friends for their advice and encouragement to my study and thesis.

Last but not the least, I would like to thank my parents for their love and support. Without their moral and material assistance, I couldn't finish my study.

Shuyao Yang

Washington University in St.Louis

May 2017

ABSTRACT OF THE THESIS

Mortgage Transition Model Based on LoanPerformance Data

By

Shuyao Yang

Master of Arts in Statistics

Washington University in St.Louis, 2017

Professor Jimin Ding, Chair

The unexpected increase in loan default on the mortgage market is widely considered to be one of the main cause behind the economic crisis. To provide some insight on loan delinquency and default, I analyze the mortgage performance data from Fannie Mae website and investigate how economic factors and individual loan and borrower information affect the events of default and prepaid. Various delinquency status including default and prepaid are treated as discrete states of a Markov chain. One-step transition probabilities are estimated via multinomial logistic models. We find that in general current loan-to-value ratio, credit score, unemployment rate, and interest rate significantly affect the transition probabilities to different delinquency states, which lead to further default or prepaid events.

Key words: Loan performance, Markov chain, Mortgage, Multinomial logistic model, Transition probability, Transition Matrix.

Chapter 1: Introduction

Mortgage delinquency and default studies are essential for mortgage loan market. A mortgage goes to default when someone fails to satisfy the terms of a mortgage obligation or fails to pay back the mortgage anymore, and the entity who holds the note may consider take possession of the property to end the loan. For example, mortgage loans default is thought to be the direct cause of subprime mortgage crisis from 2007 to 2009. With the increasing of interest rate and cooling of housing market, a large amount subprime mortgage loans went into delinquencies and then default. Lenders filed for bankruptcy protection, also making influence on investment funds and corresponding corporation. The collapse of the subprime mortgage market in America and the successive impact led to disruption of global finance market.

There are many potential influencing factors for mortgage loans performance, such as unpaid principle balance, current interest rate, property type, borrowers' debt, income, credit score, whether there are mortgage insurances and so on. The economic and financial environment will also affect the decision of borrowers, for example, the employment situation and the market house price. Bank and lender concentrate on the performance of mortgage every month to analyze the character of these loans and predict their short term and long term trend. If there appears a high tendency to go to default of a loan, bank or lender may consider taking measures to stop loss.

An important measure of loan performance is delinquent status. A mortgage loan is considered as delinquent when it fails to receive its' regular payment. Delinquency can be categorized by the duration of past due in months. For example, delinquent status "1" implies that the scheduled mortgage payment is late for more than 30 days, but less than 60 days. There is a strong relationship between default and delinquency. First, if a mortgage loan reaches a high delinquency status, it implies the borrower is in a terrible situation, for instance, he or she is unemployed. On the contrary, a loan with a low delinquency or even not delinquent now shows a larger probability to get regular payment. Second, once the mortgage loan hasn't gotten payment for several payment periods, if the borrow wants to cure the unpaid principle balance and clean the delinquency by paying more than one period, it will cause pressure on the borrower, who may decide to waive the ownership of the house and make the mortgage loan to go to default rather than paying exceed his or her ability. So the investigation of mortgage loan performance plays an important role in observing the mortgage loan market.

Empirical research has showed a lot of work in the field. Deng (2000) presented an option approach to evaluate the competing risks of mortgage termination for prepaid and default, estimating the two hazards jointly. Thierry (2002) introduced a dynamic discrete choice model using nonparametric estimation. Scott (2011) applied a Markov chain model to subprime loans to forecast the probability of moving next month into 'current', 'delinquent' or 'paid-off' states. Patrick (2013) estimated a dynamic structural mode of borrowers' default behavior, building utility function to consider finite horizon optimal decision problem. Deni (2014) investigated the possibility and accuracy of default prediction using logistic regression. Michelle (2005) used a Heckman two-step procedure and bivariate probit model to estimate probabilities of prepaid and default.

In this paper, we propose a mortgage transition model. In this newly proposed mortgage transition model, instead of estimating the probabilities of default and prepaid directly, we consider the Markov transition among delinquency status and estimate these transition probabilities via a multinomial logistic regression. This approach avoids the problem caused by tiny probability of default for a performing loan. Furthermore, constructing Markov transition matrix makes it clear to predict the path of a loan and mean residual loan age.

The data we used, LoanPerformance data, provided by Fannie Mae, focus on a portion of single-family mortgage loans from January 1, 2000 to December 31, 2015 stored quarterly. The data contains two parts. The acquisition part includes static mortgage loan data at the original time of the mortgage loan. The performance part shows the monthly dynamic performance data of each mortgage loan, until the mortgage loan goes to default or prepaid. A unique loan ID is generated to link the two part and to distinguish each loan.

In the next section of this paper, we will introduce data management steps. The miscellaneous and useless information contained in the original data is removed. Additional prognostic covariates are created. In section 3, we will build a mortgage transition model from conditional Markov chain and multinomial logistic regression. In section 4, we estimate the parameters in the mortgage transition model and interpret the coefficients in economic context. Particularly, we express the effect of variables by presenting the probability change under the change of variables. Then we create a transition matrix to give a brief summary of probabilities. To evaluate the proposed mortgage transition model, we compare our one-step prediction with empirical probabilities in both with-in sample and out-of sample. Finally, in section 5, we summarize the major findings, discuss the potential pitfalls, and provide future improvement advice.

Chapter 2: Data Management

2.1 Data Source

In the acquisition part, original loan-to-value (LTV), debt-to-income (DTI) and borrower's initial credit score are recorded, compared with the original interest rate, unpaid principle balance (UPB), date, number of borrowers and units and loan term.

More data are recorded in the performance part, including current unpaid principle balance (UPB), interest rate and loan age. The dataset also includes the indicator for modification and remaining time to mature. An integer is set as the current delinquency status for each loan in each month to represent the time of delinquency. Also, a five-digit Metropolitan Statistical Area (MSA) code is recorded to indicate the location of the property. If the mortgage loan goes into liquidation at that month, a zero balance code will be created to point out the liquidation reason of that loan and the date on which the mortgage loan balance is reduced to zero is also recorded.

As indicators for current economic condition and potential disruptions in repayment over the life of mortgage loan, we use the monthly unemployment rate at the county level to represent employment situation. The monthly unemployment rate data come from the Bureau of Labor Statistics (BLS).

To track the change in price of Collateral house and economic house price condition, we use housing price indices (HPI) also at county level. The HPI data is also from Freddie Mae, reported monthly by Metropolitan statistical area. So we can adjust the price of a house by comparing the HPI at current time and original time. Furthermore, we can modify the current LTV by adjusting the current price of house. Since the house price declines are considered to be one of the main cause of mortgage loan default, and because there can exist a considerable variation in house price even within the same state, it is important to have the HPI data at a fine geographic level. Take these reasons into consideration, we use the MSA level HPI from Fannie Mae.

2.2 Sampling

Due to huge amount of original data, a random sample of data is necessary to take. Fannie Mae publishes and updates the LoanPerformance data quarterly. So for every quarter dataset, 5% of loans are taken randomly together with their performance data are taken to make up the whole dataset for this research. Finally, 1197871 mortgage loans are taken with 54270152 performance data.

2.3 Weight

We are going to estimate one-step transition probability of each state, each with a separate model to take nonlinear effect of delinquency status into consideration. So the dataset will be divided by delinquency status. Every sub dataset covers part of life time of one loan, so a performing loan is expected to live longer, which will also contribute more information to estimation. To reduce the effect of heteroscedasticity, we consider weighted regression due to repetition. The weight is defined as the reciprocal of life time of one loan in one sub dataset.

2.4 Covariate Derivation

Current LTV

The current LTV is computed based on the original and current value of HPI and unpaid principle balance.

$$Current\ LTV = \left(\frac{Current\ UPB}{Original\ UPB}\right) * \left(\frac{Original\ HPI}{Current\ HPI}\right) * Original\ LTV$$

Unemployment Rate and HPI Change

In order to consider the delayed effect of unemployment rate and HPI, we introduce two variables as the change of unemployment rate and HPI in the past 6 months, which are also monthly MSA-level data. At the same time, these variables can also be indicators for the change of economic condition at those time, which can also have influence on the choice of borrower.

Past Delinquency Performance

The LoanPerformance data only publishes credit score at the original time, no further information about borrower's credit condition. So as to get indicator of borrower's current credit performance, a dummy variable indicating the past delinquency performance is introduced. This is an alternative variable of current credit score and credit level. It is defined as 1 if the mortgage loan once reached a delinquency status higher than or equals to 2. So this variable will only be included in lower delinquency status model estimation.

Standardization

For the sake of reducing the effect of dimension and magnitude of different variables, most of the variable are scaled and centered. Unpaid principle balance is centered in 150000, interest rate is centered in 5%, DTI ratio is centered in 30%, current LTV ratio is centered in 60%, credit score is centered in 600, HPI is centered in 1 and unemployment rate is centered in 5%. Furthermore, the unpaid principle balance is divided by 1,000,00, so the new unit is ten thousand; the loan age is divided by 12, so the new unit is one year, credit score is also divided by 100.

Table 1 Summary Statistics for Sampled Data

Variable	Mean	Std.Dev	Median	Minimum	Maximum
Current UPB	158000	88231	138000	0	1090000
Original UPB	180700	96579	160000	8000	1090000
Original LTV	0.7466	0.1522	0.7900	0.2000	0.9700
Current LTV	0.6773	0.2015	0.6916	0	2.49
Loan Age (month)	38.05	33.79	28.00	0	210.00
Remaining to Mature	323.5	35.2	333.0	69.0	483.0
Delinquency Status	0.2364	2.384	0	0	166
Interest rate	0.0560	0.0124	0.0575	0.0200	0.1162
DTI	0.341	0.145	0.340	0.010	0.640
Credit Score	737.9	55.7	749.0	367.0	850.0
HPI*	1.362	0.301	1.297	0.620	2.733
Unemployment Rate*	0.065	0.025	0.058	0.017	0.300

* Some MSA codes have changed or disappeared, so HPI and unemployment rate cannot be created for these loan. This causes 354637 missing loans.

Table 2 Summary of Categorical Variable

Categorical Variable	0	1
Modification	703637	12523
Past Delinquency Performance	31071941	1610008

Categorical Variable	0	1	2	3	4	5	6
Delinquency Status	31752695	557389	156187	74433	55449	46300	39496

2.5 Response Variable Derivation

To define default using zero balance code, we consider prepaid and repurchased loans as defined prepaid loans, and third party sale, short sale and Deed-in-Lieu, REO disposition loans as defined default loans. Take the delay of accounting into consideration, we treat those mortgage loan with delinquency status higher than 6, that is 6 months, also as default loan. Once the mortgage loans get a delinquency status equals to 7 or higher, the loans are considered as default loans which are censored for the left time.

Chapter 3: Methodology

3.1 Markov Chain Model and Transition Probability

A Markov Chain is a stochastic process that has the Markov property, which requires the future events depend only on the present but not the past. More precisely, let $\mathcal{S}$ denote a discrete set with countable states and $\{Y_t : t \in \mathbb{N}\}$ be a $\mathcal{S}$-valued stochastic process. The process $\{Y_t : t \in \mathbb{N}\}$ is called a Markov chain if it satisfies

$$\mathbb{P}(Y_t = y_t | Y_{t-1} = y_{t-1}, Y_{t-2} = y_{t-2}, \dots \dots, Y_0 = y_0) = \mathbb{P}(Y_t = y_t | Y_{t-1} = y_{t-1})$$

where $y_t \in S, \forall t = \mathbb{N}$. Here we use $\mathbb{P}$ to denote the probability measure defined on a measurable space with a filtration that $\{Y_t : t \in \mathbb{N}\}$ is adapted to. The possible values of Y_t form the state space $\mathcal{S}$ of the chain. If a Markov chain is defined on a continuous time index $t \in T$, then it is called a continuous-time Markov chain or Markov process. Furthermore, if the state space, $\mathcal{S}$, is also continuous, it is often referred to as a stochastic process with Markov property (for example, Brownian motion). In our case, we only focus on discrete-time Markov chain with a finite set states, since mortgage data are recorded only monthly.

Although Markov property is a strong assumption and hardly hold in real life economic data, one way to circumvent the problem is to include additional covariates and relax the Markov assumption to conditional Markov assumption. This is called a conditional Markov chain (CMC). Mathematically, let $\boldsymbol{X}_t = \{X_{t1}, \dots \dots, X_{tp}\}$ denote a set of covariates (including both endogenous and exogenous variables) at time t. Then Y_t satisfies

$$\mathbb{P}(Y_t = y_t | Y_{t-1} = y_{t-1}, \dots \dots, Y_0 = y_{0,} \boldsymbol{X}_{t-1}) = \mathbb{P}(Y_t = y_t | Y_{t-1} = y_{t-1}, \boldsymbol{X}_{t-1})$$

Although mathematical properties of CMC have been investigated only recently, the applications of CMC models have often been used in economics and mathematical finance. Another common way to relax the Markov assumption is to allow the dependence up to the

m-th term history. This is called a Markov chain with order m or a Markov chain with memory m:

$$\mathbb{P}(Y_t = y_t | Y_{t-1} = y_{t-1}, \dots\dots, Y_0 = y_0) = \mathbb{P}(Y_t = y_t | Y_{t-1} = y_{t-1}, \dots\dots, Y_{t-m} = y_{t-m}).$$

In my thesis, I will use a conditional Markov chain to model mortgage data, and investigate the relationship between the conditional transition probability and economic covariates. Furthermore, the conditional Markov chain is assumed to be stationary given the set of covariates $t, \boldsymbol{X}$. That is, conditional on covariates, the probability moving from the state j to the state k is homogeneous over time. This probability is called the transition probability:

$$p_{jk}(\boldsymbol{X}) \coloneqq \mathbb{P}(Y_t = k | Y_{t-1} = j, \boldsymbol{X}_{t-1} = \boldsymbol{x}) = \mathbb{P}(Y_{t-1} = k | Y_{t-2} = j, \boldsymbol{X}_{t-2} = \boldsymbol{x})$$

for all $t \in \mathbb{N}$. Putting all these one-step transition probabilities into a matrix, we have the transition matrix:

$$P(\boldsymbol{X}) = \left(p_{jk}(\boldsymbol{X})\right) = \begin{bmatrix} p_{11} & \cdots & p_{1S} \\ \vdots & \ddots & \vdots \\ p_{S1} & \cdots & p_{SS} \end{bmatrix},$$

which is a square matrix of cardinality of S. The state with $p_{jj} = 1$ is called an absorbing state. The transition matrix, conditional on covariates, satisfies the following two properties:

(1) $\sum_{k=1}^{S} p_{jk} = 1$, the row sum is 1;

(2) the probability of transition from j to k in m steps is $(P^m)_{jk}$;

hence one-step transition matrix is essential in conditional Markov chain model, and can be used to characterize the chain together with the state space and initial state.

3.2 Multinomial Logistic Regression

Logistic regression is a special case of the generalized linear model (GLM). GLM is a powerful statistical model to deal with non-normal data, especially popular for binary outcome data. It extends classical linear regression by allowing the linear combination of predictors to be related to the dependent variable, $\mathbf{Y}$, via a link function, g, and specifying the distribution of $\mathbf{Y}$ via variance function. Let $\mathbf{X}$ be the matrix of covariates. We model

$$E(\boldsymbol{Y}) = \boldsymbol{\mu} = g^{-1}(\boldsymbol{X\beta})$$

$$Var(\boldsymbol{Y}) = V(\boldsymbol{\mu}),$$

where the variance function $V(\boldsymbol{\mu})$ is often determined by the specified distribution of $\boldsymbol{Y}$. The link function, g, describes the relationship between the mean of the response variable and the linear predictor, $\boldsymbol{\eta} = \boldsymbol{X\beta}$. Although the link function might be chosen subjectively, the most common choice of link functions is the canonical link functions derived from the canonical parameters of exponential family, due to computational advantage and interpretation ability.

When the response data are binary, Bernoulli distribution is naturally assumed and the variance function is hence $V(\boldsymbol{\mu}) = \boldsymbol{\mu}(1-\boldsymbol{\mu})$. The canonical link function for Bernoulli distribution is the logit function, $g(p) = log\ (\frac{p}{1-p})$. This leads to the logistic regression model:

$$E(\boldsymbol{Y}) = \mathbb{P}(\boldsymbol{Y}=1) = g^{-1}(\boldsymbol{X\beta}) = \frac{\exp(\boldsymbol{X\beta})}{1+\exp(\boldsymbol{X\beta})}$$

$$Var(\boldsymbol{Y}) = \mathbb{P}(\boldsymbol{Y}=1)\mathbb{P}(\boldsymbol{Y}=0) == \frac{\exp(\boldsymbol{X\beta})}{(1+\exp(\boldsymbol{X\beta}))^2}.$$

The coefficient $\boldsymbol{\beta}$ can be interpreted as the logarithm of odds ratios when the covariate increases by 1 unit.

Multinomial logistic regression is a classification method that generalizes logistic regression to multiclass problems, i.e. with more than two possible discrete outcomes. For example, the blood type that a person has given the results of various diagnostic tests, the major that a college student choose given their grades. The model is used to estimate the effect of independent variables and predict the probabilities of the different possible outcomes of a categorically distributed dependent variable, given a set of independent variables. Let n denote the number of all possible outcomes and $\boldsymbol{\beta_k}$ denote the coefficient with respect to outcome k. Then the probability for each outcome can be expressed as:

$$\mathbb{P}(\boldsymbol{Y}=k)=\frac{\exp\left(\boldsymbol{X\beta_k}\right)}{\sum_{k=1}^{k=n}\exp\left(\boldsymbol{X\beta_k}\right)},$$

where $\beta_1 = 0$ for baseline outcome and the first term in the denominator is 1.

3.3 Mortgage Transition Model

In this subsection, we will combine the conditional Markov chain model and multinomial logistic regression to build our mortgage transition model. Our target is to understand the prognostic covariates that may be associated with the delinquency change, and predict the delinquency path of a mortgage loan given the selected covariates.

We first model the transition of loan delinquency as a conditional Markov chain. The state space of this discrete Markov chain $\mathcal{S} = \{-1,0,1,2,3,4,5,6,7\}$, where -1 represents defined prepaid and 7 represents defined default, and the other states represent the corresponding delinquency status. The long-term and short-term dependence is removed via appropriately conditioning on a selection of exogenous economic variables and endogenous loan history summaries. So given all covariates, we assume the transition of delinquency status is a discrete-time and discrete-state Markov chain. For simplicity, we further assume the delinquency status can only change one status at most in one step transition, since this is the most common case in delinquency status transition. Therefore, from a given delinquency, a loan can only move to the two adjacent states: one delinquency higher or one delinquency lower, or stay at the current delinquency. Secondly, we model these three events by a multinomial logistic regression and estimate the conditional transition probability via the selected covariates. The estimated transition probability is then used to form the transition matrix and predict future delinquency status.

Denote the delinquency status of the i-th loan at time t as D_{it}. Let $\boldsymbol{\beta_{d,d^*}}$ be the coefficient vector for transition from state d to state d^*. Let $\boldsymbol{X}_{it} = \{1, X_{1it}, \cdots, X_{pit}\}$ be the covariates for the i-th loan at time t. We propose the following mortgage transition model:

$$P(\boldsymbol{X}_{it}) = \left(p_{jk}\right), \qquad j,k \in \mathcal{S}$$

where

$$p_{jk} := \mathbb{P}(D_{i(t)} = k | (D_{i(t-1)} = j, \boldsymbol{X}_{it})$$

$$= \begin{cases} \dfrac{\exp(\boldsymbol{X_{it}\beta_{jk}})}{\sum_{k=j-1}^{k=j+1} \exp(\boldsymbol{X_{it}\beta_{jk}})} & if\ j = 0,1,2,3,4,5,6\ and\ k = j-1, j, j+1 \\ 1 & if\ j = k = -1\ or\ 7 \\ 0 & otherwise \end{cases}$$

This transition matrix will only have non-zero elements on main and secondary diagonals and 0 anywhere else. In model fitting, we will treat staying in the same state as a baseline outcome. Mathematically, this implies $\beta_{jj} = 0$ for $j = 0,1,2,3,4,5,6$. So the odds ratio is defined as the ratio of probabilities of movement and staying in baseline:

$$\frac{p_{jk}}{p_{jj}} = \exp(\boldsymbol{X_{it}\beta_{jk}})$$

In other words, every unit increase in X_{pit} multiple the relative risk by $e^{\beta_{pjk}}$, and β_{pjk} is the logarithm of the odds ratio.

$X_{it} = \left(1, X_{1it}, X_{2it}, X_{3it}, X_{4it}, X_{5i}, X_{6it}, X_{7i}, X_{8it}, X_{9it}, X_{10it}, X_{11it}, X_{12it} * I_{\{d<2\}}\right)$

X_{1it}: Unpaid principle balance for loan i at time t.

X_{2it}: Loan age, the time since the first payment expressed by month for loan i at time t.

X_{3it}: Interest rate on a mortgage loan in effect for the periodic due for loan i at time t.

X_{4it}: An indicator that denotes if loan i at time t has been modified.

X_{5i}: Debt-to-income ratio for loan i at original time.

X_{6it}: Current loan-to-value ratio for loan i at time t.

X_{7i}: Credit score of the borrower for loan i at original time.

X_{8it}: Unemployment rate of the MSA for loan i at time t.

X_{9it}: House price index of the MSA for loan i at time t.

X_{10it}: Change of unemployment rate of the MSA for loan i at time t rate compared with the unemployment rate at six months ago.

X_{11it}: Change of house price index of the MSA for loan i at time t rate compared with the house price index at six months ago.

X_{12it}: A dummy variable to indicator past delinquency performance for loan i at time t. For those serious delinquent loans, this variable is meaningless, so the effects are only included for model with delinquency status 0 and 1.

$I_{\{d<2\}}$: An indicator function. If the current delinquency status is lower than 2, the value is set to 1, else set to 0.

Chapter 4: Results

4.1 Estimation of Model Parameter

First we report the estimate coefficient with standard error in the parentheses for each prognostic covariant in each delinquency status model. The bold coefficients are significant at $\alpha = 0.05$. The gray lines represent the coefficient $\beta_{d(d-1)}$ and the white lines represent $\beta_{d(d+1)}$. We also put 1 and -1 at second column to indicate the transition tendency.

Table 3 Coefficient and Standard Error From Multinomial Logistic Regression

Delinquency	Tendency	0	1	2	3	4	5	6
Unpaid Balance	-1	**0.257(0.008)**	0.002(0.013)	**-0.124(0.022)**	**-0.148(0.031)**	**-0.143(0.037)**	**-0.128(0.04)**	**-0.178(0.045)**
	1	0.007(0.018)	**0.064(0.014)**	0.013(0.019)	**-0.082(0.025)**	**-0.126(0.03)**	**-0.17(0.033)**	**-0.186(0.037)**
Loan Age	-1	**0.008(0.004)**	**-0.104(0.004)**	**-0.111(0.006)**	**-0.1(0.009)**	**-0.134(0.011)**	**-0.125(0.012)**	**-0.123(0.014)**
	1	**-0.012(0.006)**	**0.04(0.004)**	-0.006(0.005)	**-0.027(0.007)**	**-0.055(0.009)**	**-0.063(0.01)**	**-0.062(0.012)**
Interest Rate	-1	**35.538(0.049)**	**14.177(0.023)**	**17.079(0.043)**	**20.419(0.005)**	**27.138(0.008)**	**23.382(0.008)**	**23.308(0.007)**
	1	**2.923(0.002)**	**-9.298(0.008)**	**-9.036(0.068)**	**-7.045(0.007)**	**-5.282(0.01)**	**-7.795(0.01)**	**-9.019(0.009)**
Modification	-1	-0.134(0.189)	**0.758(0.052)**	**0.385(0.061)**	**0.564(0.083)**	**0.667(0.097)**	**0.835(0.102)**	**0.87(0.113)**
	1	**-0.646(0.068)**	**-0.55(0.05)**	**-1.036(0.052)**	**-0.832(0.068)**	**-0.696(0.078)**	**-0.579(0.085)**	**-0.628(0.094)**
DTI	-1	**0.9(0.063)**	**-0.879(0.079)**	**-0.963(0.135)**	**-0.679(0.2)**	**-0.665(0.242)**	**-0.532(0.268)#**	-0.443(0.306)
	1	**0.837(0.115)**	**0.205(0.089)**	**-0.243(0.119)**	**-0.539(0.168)**	**-0.6(0.202)**	**-0.525(0.222)**	**-0.626(0.253)**
Current LTV	-1	**-0.156(0.044)**	**-1.426(0.055)**	**-1.307(0.087)**	**-1.37(0.12)**	**-1.398(0.14)**	**-1.476(0.151)**	**-1.126(0.169)**
	1	**0.893(0.079)**	**1.431(0.055)**	**1.238(0.072)**	**0.949(0.093)**	**0.773(0.11)**	**0.688(0.119)**	**0.746(0.134)**
Credit Score	-1	**2.4(0.022)**	**1.9(0.018)**	**1.5(0.030)**	**1.4(0.043)**	**1.4(0.053)**	**1.4(0.059)**	**1.4(0.068)**
	1	**-2.4(0.023)**	**-0.9(0.019)**	**-0.8(0.025)**	**-0.9(0.035)**	**-0.9(0.043)**	**-0.9(0.047)**	**-0.9(0.054)**
Unemployment rate	-1	**2.815(0.337)**	**-4.451(0.398)**	**-9.076(0.402)**	**-9.893(0.265)**	**-6.559(0.33)**	**-4.435(0.367)**	**-4.269(0.409)**
	1	**2.221(0.009)**	**5.059(0.387)**	-0.124(0.518)	**-3.041(0.363)**	**-4.524(0.454)**	**-4.062(0.504)**	**-4.549(0.559)**
HPI	-1	**0.05(0.028)**	**-0.097(0.037)**	-0.109(0.067)	**-0.242(0.101)**	**-0.27(0.125)**	**-0.322(0.139)**	0.197(0.161)
	1	**0.154(0.048)**	-0.077(0.042)	**-0.229(0.06)**	**-0.41(0.085)**	**-0.541(0.104)**	**-0.685(0.115)**	**-0.433(0.134)**
Unemployment Rate Change	-1	**-4.864(0.006)**	**-5.376(0.047)**	**-2.034(0.393)**	**-2.213(0.077)**	**-3.173(0.04)**	**-0.443(0.036)**	**-5.278(0.038)**
	1	**5.518(0.008)**	**7.763(0.115)**	**5.633(0.274)**	**1.76(0.097)**	**0.504(0.051)**	**6.791(0.043)**	**3.399(0.041)**
HPI Change	-1	**3.447(0.115)**	**0.807(0.125)**	**1.577(0.215)**	**0.62(0.324)**	**0.41(0.411)#**	-0.601(0.47)	**-1.422(0.542)**
	1	**-0.663(0.168)**	**-0.758(0.131)**	-0.107(0.179)	**-1.079(0.264)**	**-1.768(0.333)**	**-2.549(0.385)**	**-2.37(0.448)**

Figure 1 shows the coefficient compared with its confidence interval for 2 standard errors.

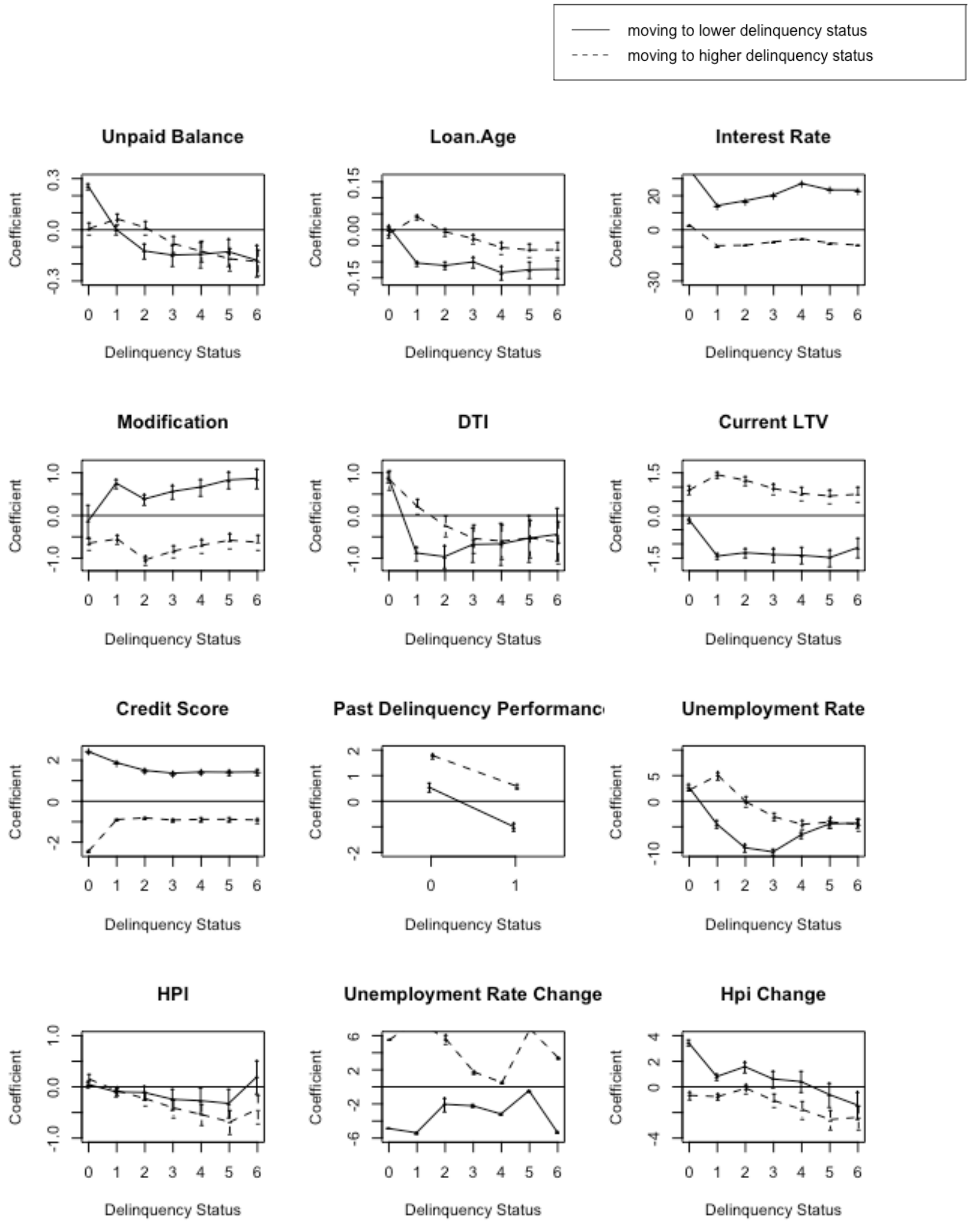

Figure 1 Coefficient Change Among Delinquency Status

The Figure 2 shows the probability of three transitions with the change of delinquency status. Other variables are set to baseline, that is, all $X_{it} = 0$, except that loan age is set to be 15 years $X_{2it} = 15$, and credit score is set to be 650 ($X_{7i} = 50$). This baseline will be used in the following part if no further statement.

As Figure 2 shows, the probability of staying in the same delinquency status declines rapidly as delinquency status increases, which indicates once a loan is delinquent, it becomes much more difficult to stay in the same delinquency status. It shows that this loan has met some barriers and has trouble continuing regular payments, let alone curing the balance by paying more than regular. This cause the lower probability of moving to lower status compared with the probability of staying in current status.

On the contrary, the probability of moving up increases rapidly as delinquency status increases. If the loan reaches delinquency status equals to or higher than 3, the probability of it continuing to move to higher delinquency status next month is larger than 0.6. Once it reaches 6, which is the critical point of defined default, the probability of moving higher, which is also the probability of defined default, reaches up to 0.777. However, once the loan become delinquent, the negative effect of delinquent status tends to be stable with the increase in delinquency status. For example, when the delinquency status change from 0 to 1, the logarithm of odds ratio increases by 3.184 and the moving higher probability change is 0.2456, where the logarithm of odds ratio change is only 0.587, the probability change is 0.09 for delinquency status changing from 3 to 4; this is also showed by the decreasing slope of the probability curve.

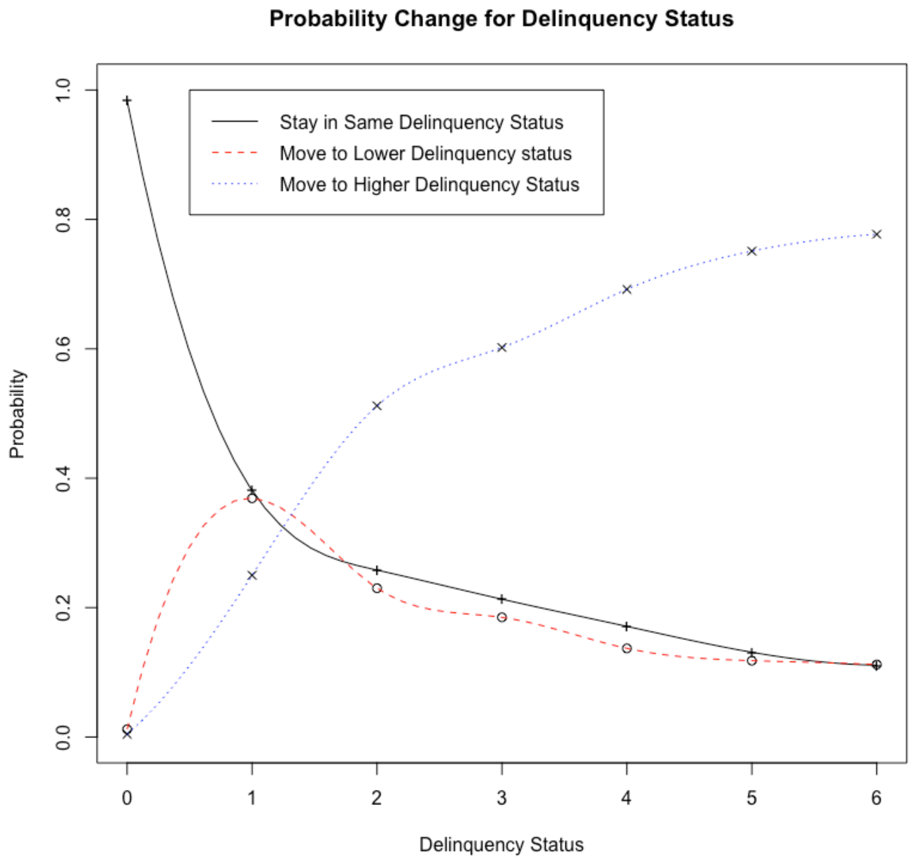

Figure 2 Probability Change for Delinquency Status

Another important covariate is the credit scores. It plays an important role in mortgage loan's next month transition. A borrower with higher credit score shows a higher downward tendency and also a lower upward tendency as expected. Figure 3 shows the change of probability under the change of credit score for different delinquency status.

Figure 3(a) shows that the probability of moving to delinquency status 1 for a performing loan decreases immediately when the credit score increases from 400. The probability of moving to higher delinquency status decreases 0.8233 if the credit score changes from 400 to 600. Once the mortgage loan is delinquent, the probability of continuing to move to higher delinquency status decreases rapidly when the credit score increases from nearly 600. Take delinquency status 3 as an example, the probability of moving to higher status decreases only

0.0687 when credit score changes from 400 to 600, while the change is 0.6055 when credit score changes from 600 to 800.

Figure 3(b) shows that the probability of moving to lower delinquency status is not sensitive to credit score when it is lower than 600. The largest change of probability among all delinquency status when credit score changes from 400 to 600 is only 0.042. But if the credit score is higher than 600, these probabilities for delinquent mortgage loan increase rapidly. For example, for loan with delinquency status 1, the probability of moving to lower status increases by 0.4856 when the credit score increases from 650 to 750. However, the credit score shows a slightly positive effect for those performing loans. The reason is that for a performing loan, the defined downward action means the loan is prepaid, which is weakly related to credit level. The probability increase is only 0.0378 when the credit score increases from 600 to 800.

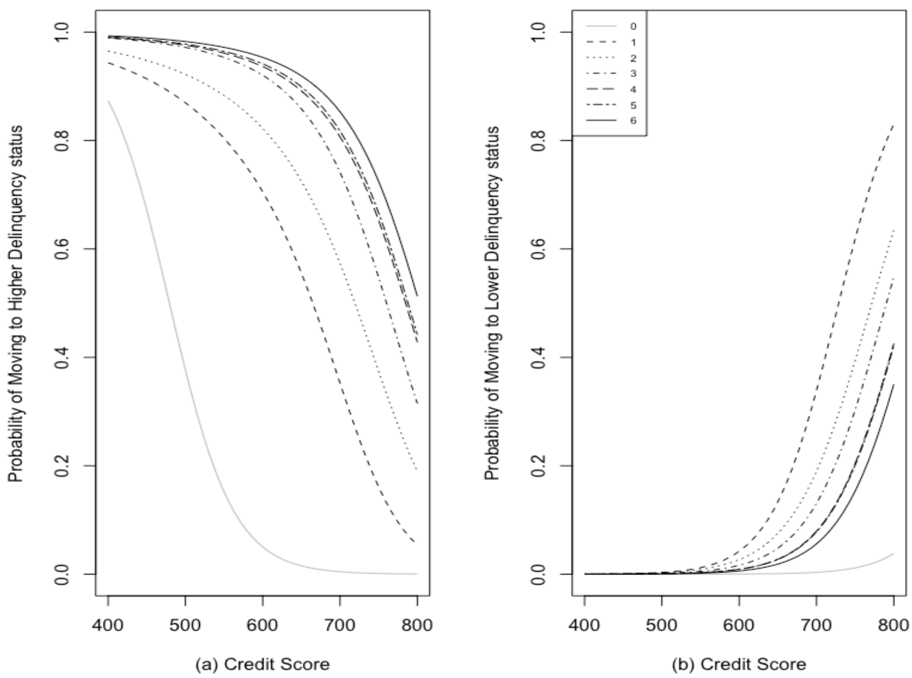

Figure 3 Probability Change for Credit Score

Current LTV is defined as the ratio of unpaid principle balance to the current value of house. We can find upward tendency is positive related to the current LTV and downward tendency is negative related to it. It's obvious that the delinquency status of a mortgage loan with higher current LTV has a higher probability to move to higher status and lower probability to move to lower status correspondingly. When the current LTV is in excess of 100%, which means now the value of the property is less than the mortgage unpaid principle, this negative equity usually leads to strategic default. The probability of default for 100% current LTV reaches up to 0.9526. Figure 4 (a) and (b) reports the probability change under the change of current LTV.

According to Figure 4 (a), the first conclusion is that the current LTV has a strictly positive effect on the probability of moving to higher status and also a strictly negative effect on the probability of moving to lower status. Second, these probabilities are more sensitive to current LTV under low value; the lower delinquency status, the higher effect of current LTV. For instance, with the increase of delinquency status from 1 to 6, the increase of probability of moving to higher delinquency status under the change of current LTV from 60% to 80% are 0.087, 0.578, 0.030, 0.019, 0.017, 0.013 respectively. This conclusion can also be found by the decrease of absolute values of coefficient. Something surprising is that the current LTV shows only a little effect for a performing loan compared with those delinquent loans. Especially when considering the probability of prepaid, the absolute t-value is 3.522, which is the lowest among all situations. In Figure 1, the coefficient and the confidence interval for this coefficient are close to the x-axis.

The interest rate variable is defined as the interest rate on a mortgage loan in effect for the periodic installment due. Figure 4 (c) and (d) reports the probability change of action with the change of interest rate.

Figure 4 (c) shows that with the increase of interest rate, the probability of moving to higher status decreases and the probability of moving to lower status increases. The lower the delinquency status is, the more sensitive the probability for a delinquent loan is to interest rate change. If the interest rate increases from 4% to 8%, the decrease of probability of moving to higher status is 0.1389 for delinquency status 1, but only 0.05339 for status 6. From Figure 4 (d), we can get similar results for moving to lower status of delinquent loans. Although the probability of prepaid for a performing loan seems to have lowest slope, implying that the interest rate has the lowest effect on the probability of prepaid, which is in contradiction to the fact that the corresponding coefficient is the largest. However, if we calculated the rate of probability change as the interest rate changes from 0.02 to 0.1, the rate is only 2.78 for delinquency status 1 but 270.75 for performing loan, showing that interest rate has a significant effect on the probability of prepaid. This result is not caused by the tiny probability, since the rate of probability change of moving up is only 0.216 for performing loan.

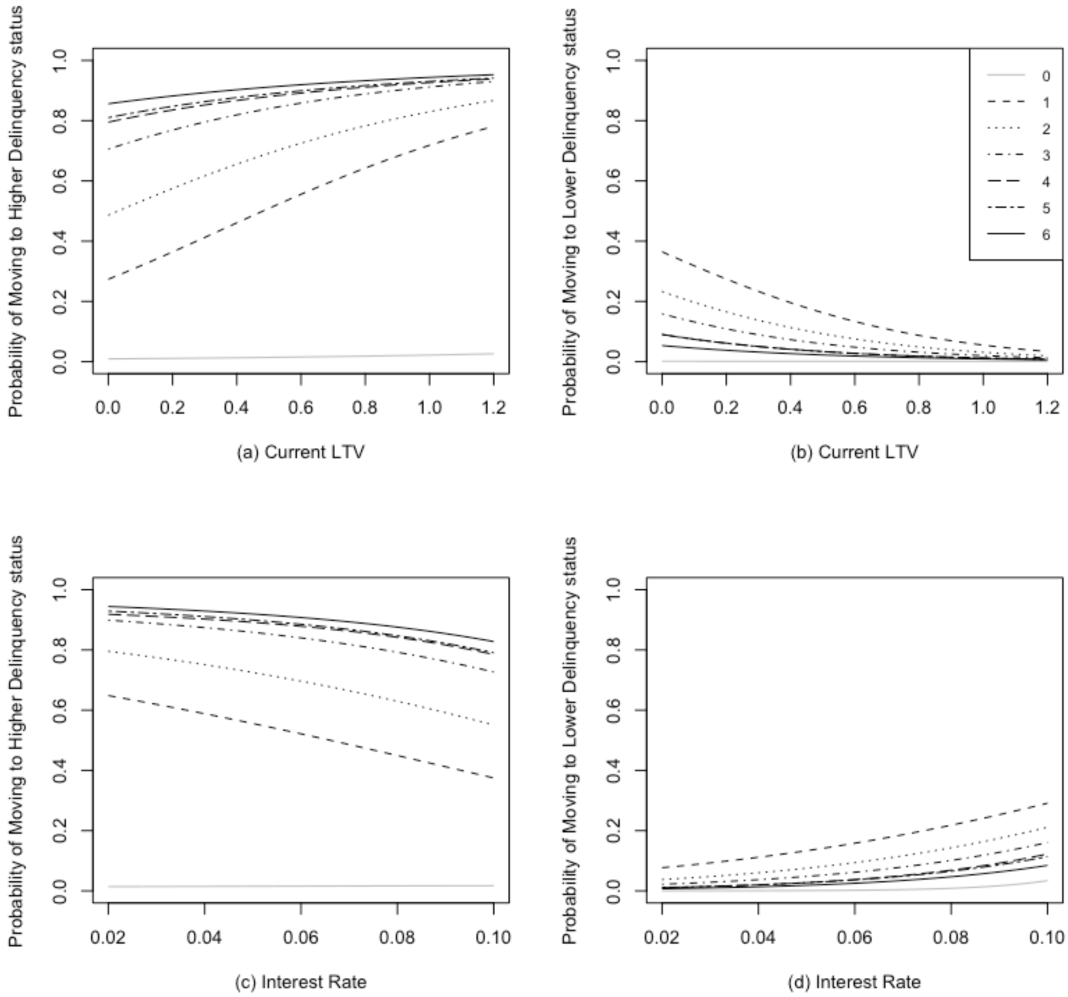

Figure 4 Probability Change for Current LTV and Interest Rate

Unpaid principle balance is correlated to loan age to some extent, but it is not the substitute of loan age. For a mortgage loan with delinquency status higher than or equals to 3, the decrease of unpaid balance will lead to the increase of both moving probability, which implies the decrease probability of staying in the current status. Under the change of unpaid principle balance from 1,000,000 to 0, the increases of the probability of staying in the same status with delinquency status changes from 3 to 6 are 0.097, 0.147, 0.208, 0.214 respectively. The effects under delinquency status 1 and 2 are not statistically significant compared with other effects. Figure 5 (a) and (b) shows the probability change under the change of unpaid balance.

The effect of loan age is strictly negative on the probability of moving to lower delinquency status. The corresponding coefficients are close, but this coefficient will cause the largest decrease probability under delinquency status 1. However, the effect on the probability of moving to higher status is a little complicated. The signs change over different delinquency status. For status 1 and 2, under the increase of loan age, the probability of moving to higher will increase together. Then for delinquency status 0 and 4 to 6, with the increase of loan age, the probability of moving to higher status will decrease. Figure 5 (c) and (d) reports such probability changes.

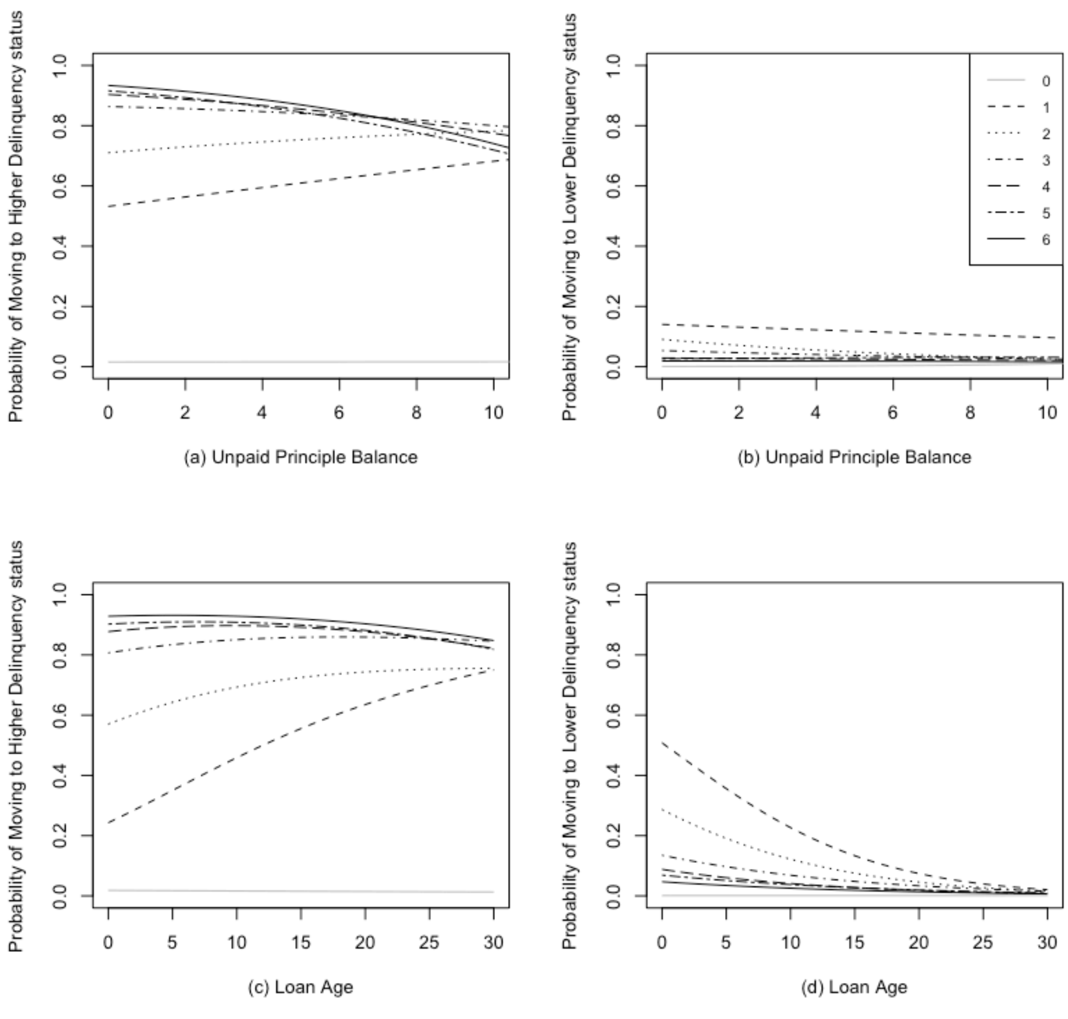

Figure 5 Probability Change for Unpaid Principle Balance and Loan Age

The unemployment rate is always considered to be one of the main influence factors of default. It influences the income of borrower, hence affects the repayment of mortgage loan. Figure 6 shows the probability change with the change of unemployment change. Rising the unemployment rates are supposed to increase the probability of moving to higher delinquency status and decrease the probability of moving to lower status. The latter part of this hypothesis is verified by the decrease curve in Figure 6 (b). However, in our estimated coefficient, the unemployment rate shows a negative effect on the probability of moving to higher status with high current delinquency status. Some curves are decreasing in the Figure 6 (a). So besides estimating the effect of simple unemployment rate, we estimate the effect of unemployment rate change in the past 6 months. The coefficient results show that the rate change is more statistically significant than the simple unemployment rate. The higher the change is, which means the unemployment rate increases a lot, the higher probability of moving to higher status is and the lower probability of moving to lower status is. For instance, for delinquency status 1, if the unemployment rate change ranges from -12% to 19%, the probabilities of moving up and down change up to 0.675 and -0.332.

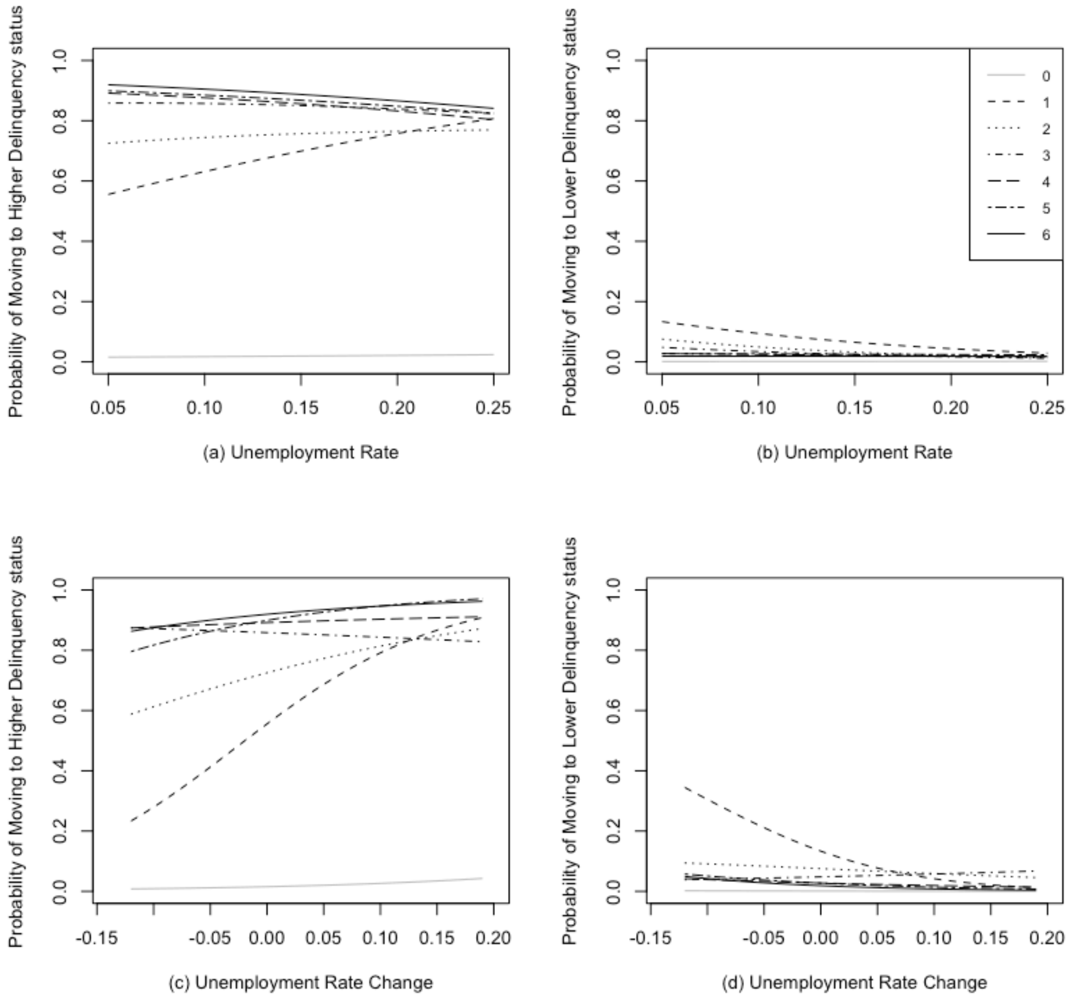

Figure 6 Probability Change for Unemployment Rate

HPI influences the action of borrowers by means of reflecting the current house value. Figure 7 shows the probability change under the change of HPI. We can conclude that HPI itself has a slightly negative effect on the probability of both two transitions. Figure7 shows that if the HPI has a positive change, which means the house becomes more valuable, the probability of moving to higher status will decrease and probability of moving to lower status will corresponding increase. This is not surprising since if the value of house rises, the borrower tends to keep the ownership of the house rather than going to default and losing the house. Having a positive equity position will make the borrower prefer preserving the house instead of letting the house go to foreclosure.

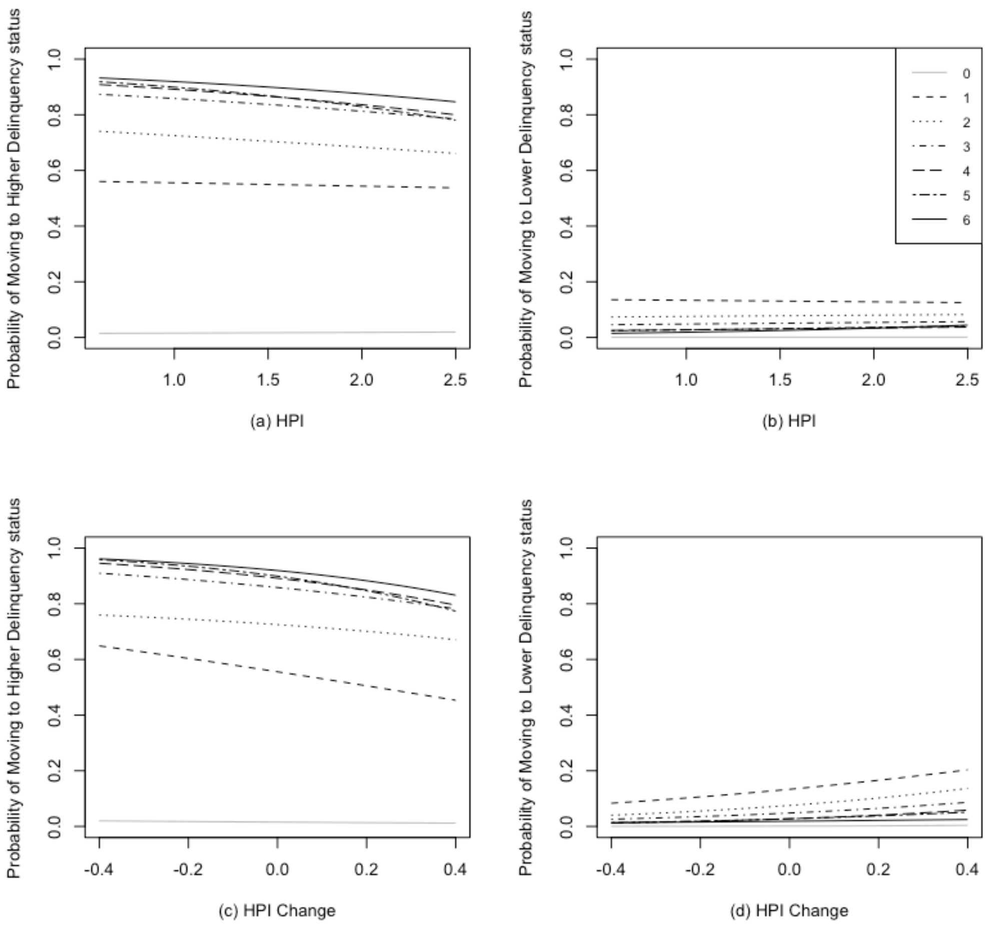

Figure 7 Probability Change for HPI

The debt-to-income (DTI) is a ratio calculated at origination by dividing the borrower's total monthly obligations by monthly income. DTI has negative effect on both the probability of two transitions for a delinquent loan. The delinquent loan with a relatively high debt is supposed to have a higher probability to stay in current situation. However, the data is collected at original time, the original DTI has fewer impact on the current performance of a mortgage loan than the impact of those monthly-update variables. So although most of the coefficients are significant, within the range of original DTI, the probability doesn't get a remarkable change.

The past delinquency indicator is used to imply the borrower's current credit level. This value is set to be 1 if the loan once gets a delinquency status higher or equals to 2. When the current delinquency status is 0, that is a performing loan, the coefficients of two transitions are both positive, showing that if the mortgage loan once had a delinquent status, it will have higher probability of prepaid and moving to status 1, resulting lower probability of staying in status 0. When the current delinquency status is 1, if the mortgage loan once has a higher delinquency, the probability of moving back and staying in status decreases by 0.0465 and 0.054 respectively as expected. While the probability of moving to higher delinquency status increases by 0.1012.

The probability change caused by modification is very clear. A modified mortgage loan will have higher probability of moving to higher delinquency status and lower probability of moving to lower status. These coefficients are also statistically significant.

4.2 Transition Matrix

Under the single movement assumption, we can build a general transition matrix for baseline data defined before. We assume that default and prepaid are terminal actions: Once the mortgage loan is prepaid or default, there is no further action to do and these mortgage loans are treated as censored, that is default or prepaid decision is irreversible. So these two states are absorbing states.

Table 4 One-Step Transition Matrix

j \ k	-1	0	1	2	3	4	5	6	7
-1	1								
0	0.012	0.984	0.004						
1		0.369	0.381	0.250					
2			0.230	0.258	0.512				
3				0.185	0.213	0.602			
4					0.137	0.171	0.692		
5						0.118	0.131	0.751	
6							0.112	0.111	0.777
7									1

with state space $S = \{-1,0,1,2,3,4,5,6,7\}$, where P_{jk} is defined before, satisfying the row sum restriction.

In table 5, we report the standard error of these predictive probabilities approximated by delta method.

Table 5 Standard error of Prediction Probabilities

Standard Error	**Transitions**		
Delinquency status	$\boldsymbol{d^* = d - 1}$	$\boldsymbol{d^* = d}$	$\boldsymbol{d^* = d + 1}$
0	0.00051	0.00015	0.00091
1	0.00200	0.00223	0.00212
2	0.00322	0.00270	0.00304
3	0.00519	0.01355	0.00484
4	0.00488	0.01096	0.00462
5	0.00469	0.00545	0.00450
6	0.00587	0.00535	0.00567

This is not an irreducible matrix since there exist two absorbing states. Finally, all variable in Markov chain will stay in state -1 or 7. This follows the fact that the mortgage loan will always end in prepaid or default. So there doesn't exist corresponding stationary distribution or limiting distribution. However, we can build multistep transition matrix by fixing the covariates as the baseline data defined before, showing the probability of transitions in near future, say after six months or one year.

Table 6 Six Months Transition Matrix

j \ *k*	-1	0	1	2	3	4	5	6	7
-1	1								
0	0.069	0.918	0.008	0.003	0.002	0^*	0^*	0^*	0^*
1	0.032	0.629	0.034	0.041	0.054	0.063	0.065	0.051	0.031
2	0.007	0.215	0.037	0.051	0.077	0.107	0.132	0.140	0.234
3	0.001	0.048	0.018	0.028	0.048	0.075	0.113	0.130	0.539
4	0^*	0.006	0.005	0.009	0.017	0.032	0.054	0.076	0.800
5	0^*	0^*	0^*	0.002	0.004	0.009	0.019	0.026	0.938
6	0^*	0^*	0^*	0^*	0.001	0.002	0.004	0.006	0.986
7									1

0^* mean the corresponding probability is positive but very close to zero. The transition matrix shows that, with time goes by, the delinquent mortgage loan has a higher probability to go to default. For instance, a loan with current delinquency status 6 will go to default after half year with probability reaching up to 0.986.

4.3 Model Fitting

Table 7 shows the within-sample fit probabilities. It shows that the within-sample fit is acceptable.

Table 7 In-sample Fit

Transition	$d^* = d - 1$		$d^* = d$		$d^* = d + 1$	
Delinquency status	**Prediction**	**Data**	**Prediction**	**Data**	**Prediction**	**Data**
0	0.013	0.009	0.959	0.976	0.028	0.015
1	0.154	0.172	0.377	0.392	0.469	0.436
2	0.367	0.358	0.348	0.338	0.285	0.304
3	0.634	0.599	0.134	0.205	0.232	0.196
4	0.679	0.692	0.151	0.161	0.170	0.147
5	0.708	0.720	0.167	0.151	0.125	0.129
6	0.771	0.754	0.143	0.134	0.086	0.112

Because the estimation is fully depended on sample data, one might worry about poor out-of-sample predictions. To check this, we randomly select another sample of the original dataset from LoanPerformance and use this dataset for prediction and validation. Table 8 shows the result of validation. It shows that the out-of sample fit is also acceptable, though it is not surprising that the prediction is slightly worse than the within-sample fit.

Table 8 Out-of-sample Prediction

Transition	$d^* = d - 1$		$d^* = d$		$d^* = d + 1$	
Delinquency status	**Prediction**	**Data**	**Prediction**	**Data**	**Prediction**	**Data**
0	0.019	0.011	0.963	0.975	0.018	0.014
1	0.162	0.178	0.398	0.432	0.440	0.390
2	0.375	0.363	0.347	0.339	0.278	0.298
3	0.612	0.594	0.176	0.206	0.212	0.200
4	0.671	0.683	0.162	0.175	0.167	0.142
5	0.697	0.715	0.161	0.149	0.142	0.136
6	0.795	0.761	0.142	0.127	0.063	0.112

Chapter 5: Conclusion

The unexpectedly high amount of default mortgage loans has led to much plight and many challenges in the loan market. So a better understanding of how and when these mortgage loans terminates is one of the most urgent problems to be solved.

In this paper we propose a mortgage transition model and identify some prognostic factors which are strong associated with the final prepaid and default decisions. We find that in general current loan-to-value ratio, credit score, unemployment rate, and interest rate significantly affect the transition probabilities to different delinquency states, which will lead to prepaid or default. These covariates further affect the probability of a loan to be prepaid and default. The change of some factors gives the opposite effect on prepaid and default while the change of others gives the concordant effect. The effects of some covariates vary with the change of delinquency status while the effects of others are more constant such as interest rate and current LTV.

There also are some drawbacks and inconsistency in the estimation. Some data is only collected once at original time such as debt-to-income ratio and credit scores, which leads to loss of some persuasion. The interpretation of delinquency is not transparent since the effect is expressed by different model. In the future we can consider midterm (3 months) or long term (6 months) prediction. To extend the one-step Markov chain model, one also can build an m order Markov chain to consider dependence effect of past mortgage loan performance and make further predictions.

Reference

Deng Yongheng, John M. Quigley, and Robert Van Order, 2000. Mortgage Terminations, Heterogeneity and the Exercise of Mortgage Options. *Econometrica*, Vol 68, No.2 (March, 2000), 275-307.

Deni Memic, 2015. Assessing Credit Default Using Logistic Regression and Multiple Discriminant Analysis: Empirical Evidence From Bosnia and Herzegovina. *Interdisciplinary Description of Complex Systems* ,13(1), 128-153.

Faraway, Julian James (2016). *Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models*. Boca Raton: CRC/Taylor & Francis Group.

Lawler, Gregory F (2006). *Introduction to Stochastic Processes*. Boca Raton: Chapman & Hall/CRC.

Michelle A. Danis and Anthony Pennington-Cross, 2005. A Dynamic Look at Subprime Loan Performance. *Journal of Fixed Income*, Vol. 15, No. 1 (June 2005): 28-39.

"Multinomial Logistic Regression ", *wikipedia*
https://en.wikipedia.org/wiki/Multinomial_logistic_regression

Patrick Bajari, Chenghuan Sean Chu, Denis Nekipelov and Minjung Park, 2013. A Dynamic Model of Subprime Mortgage Default: Estimation and Policy Implications. National Bureau of Economic Research Working Paper 18850.

Scott D. Grimshaw and William P. Alexander, 2011. Markov chain models for delinquency: Transition matrix estimation and forecasting. *Applied Stochastic Models in Business and Industry* 27(3):267 – 279.

Thierry Magnac and David Thesmar, 2002. Identifying Dynamic Discrete Decision Processes. *Econometrica*, Vol 70, No.2 (March, 2002), 801-816.

T.R. Bielecki, J. Jakubowski, M. Niewgowski (2015). Conditional Markov chainscon-struction and properties. Stochastic Analysis. Special volume in honour of Jerzy Zabczyk, Polish Academy of Sciences, Institute of Mathematics, Warsaw (2015), pp. 3342