AN APPLICATION OF LOGIT MODEL TO CREDIT SCORING AND ITS IMPLICATIONS TO FINANCIAL MARKETS

WANG WENDI

Master by Research. Econ, NUS

A THESIS SUBMITTED

FOR THE DEGREE OF MASTER BY RESEARCH

DEPARTMENT OF ECONOMICS

NATIONAL UNIVERSITY OF SINGAPORE

2014

Declaration

I hereby declare that the thesis is my original work and it has been written by me in its entirety. I have duly acknowledged all the sources of information which have been used in the thesis.

This thesis has also not been submitted for any degree in any university previously.

Wang Wendi

1 Jul 2014

Wang Wendi

01 Jul 2014

Acknowledgment

I would like to extend my grateful thanks to all the people who always encourage me, support me, and criticize me during my research and thesis writing. No matter how strange my belief or silly mind is, they give me strength and love me all along.

First and foremost, I would like to show the greatest appreciation to my respectable and resourceful supervisor, Professor Zeng Jinli, who gave me comprehensive instructions and considerable support, as well as broadened my horizon on doing the research. Due to the difficulty in collecting Chinese credit data, we both made great efforts in conducting the empirical work.

Moreover, I shall extend my thanks to Jianguang, Songtao, and other patient and nice seniors who shared me with their precious experience on research. I would also like to thank my lovely friends and kind fellows, especially Wanyu, Wang Yi, Jiajie, Xuyao and Yifan who accompanied me for the happiest two years. Without their encouragement and supervision, I could hardly finish my work on time.

Last but not the least, I give my sincere thanks to my parents, who firmly support me, always love me wherever I am and whoever I will be.

Table of Contents

Summary

Nowadays credit scoring becomes increasingly important in financial services. Unlike the situation in developed countries where credit scoring has been widely used for years and the fact that credit scoring benefits the banking market from reducing default risk to gain profits, most of the banks in China are still using a judgmental system which largely depends on the individual experts and lacks the efficiency. The emergence of big data leads the world to speak through numbers, making the establishing of an effective credit scoring system in China necessary. Compared to the traditional judgmental system and another frequently used technique, linear regression models, logit models are successful and commonly used technique. We did a tentative work by collecting loan data in 2012 from Freddie Mac mortgage company and make 90 days past due as a proxy to estimate the applicants' probability of default. Additionally, we analyzed the economic benefits by using credit scoring method and give implications to China's financial markets. We suggest China take such a method with available Chinese-characteristic variables into account, thus be able to predict the default probability and manage the credit risk in China.

Keywords: logit model, credit scoring, financial markets

List of Tables

List of Figures

Chapter 1 Introduction

Credit scoring is widely used after it was introduced in 1950s, mainly for credit cards, mortgage lending and other banking services. Credit evaluation is an essential process in bank credit management decision. Banks need to identify customers risk forecasting techniques which can minimize the risk of default as well as maximize the profit from particular customers. They classify loan applicants into "good", "bad" or "indeterminate" categories according to their profitability decisions. The process includes collecting previous customers' information, analyzing and classifying various credit elements and variables to assess the decision (Abou and Pointon, 2011). Naturally, the definition of credit scoring comes as the mechanism used by financial institutions to predict the probability that loan applicants will default or become delinquent.

Traditional ways of deciding whether to grant credit to individuals depend on the experiences and the common sense of analysts, which is called judgmental systems. Analysts compared the characteristics of the loan applicants with previous ones. If the features of the present customer are similar with the past ones who have been granted loans and have paid back on time, the applicants will be approved on a large chance. Otherwise, if their features closely resemble those who have been granted loans but defaulted in the end, they may be rejected normally. The strengths of the judgmental method include considering qualitative characteristics of customers and keeping a good track of records in analyzing previous credits thus accumulate experts' experience in analyzing new ones. However, it seems that the method is subjective and inaccurate depending on different individual analysts' abilities.

As fast development both in financial markets and the computer technology, increasing demand of credit in the market and intense competition between banks force bankers to change the schemes to sophisticated statistical ones to convenience the process of granting decisions. In a credit scoring model, analysts input the past customers' data with the results of being approved or not and derive the quantitative model. By using the model, all applications are evaluated equally and consistently. Then the analysts set up a cut-off point to separate applicants from unacceptable ones according to the past applicants who defaulted and who did not, and lend money to the approved ones under different interest rates. Therefore, getting a good (high) score in an applicant' credit report suggests a high probability of being approved the grant of loans. Conversely, applicants who get a low score are recommended for rejection.

Compared to traditional human judgment systems, credit scoring evaluates the risk of default efficiently, accurately and fairly. It provides a method of quantifying the relative risks of different groups of borrowers (Loretta, 1997). A credit score in the US is a 3-digit number that represents a 'snapshot of that individual's risk level' based on a person's history at a particular point in time (TransUnion, 2007). Unlike traditional manual underwriting methods, credit scores eliminate the risk from human error and provide a neutral and objective base of decision. First, credit scoring models reduce the time spent in the loan granted process thus reducing the cost of banks in approving loans. Moreover, credit scoring benefits the lenders by ensuring that they are applying under the same criteria regardless of gender, race and other factors that may be prohibited by law. Further, it also provides a relatively accurate result by using statistical techniques and predicting the performance of the customers and thus greatly helps the financial institutions in measuring the risks and profits.

Classical statistical methods used in developing score-cards include linear regression, linear discriminant analysis, logistic regression, probit analysis and Markov Chain (Hand and Henley, 1997). Besides, advanced non-statistical methods such as neural networks, expert system and mathematical programming have also been applied in credit scoring process in recent years. Among these statistical models, one of the most commonly used and successful methods in the industry is a logit model, based on the findings by Boyle, Crook, Hamilton, and Thomas (1992), Desai, Crook, and Overstreet (1996, 1997), Henley (1995), Srinivasan and Kim (1987), Yobas, Crook, and Ross (2000)[1].

Logit models popularized in a majority of developed countries since they were introduced from 1980s. During the last few decades, the market of credit products increased enormously, and most of the institutions analyze consumers' data to give credit offers by using logit models. A logit model is a form of generalized linear model characterized by a linear index and a logistic link function (Glennon et al., 2007). The maximum likelihood method is required for the estimation. Its straightforward advantage is that it has a binary outcome. The dependent variables can be conveniently interpreted and applicants can be easily classified into "good" or "bad" groups.

---

[1] For complete survey see Thomas (2000).

Although a linear regression model is frequently used in a scoring model by estimating the coefficients of characteristic variables and giving the weight of score, it has some drawbacks that:

"……it implicitly assumes that the attribute measurements arise from multivariate normal populations such that the classes have identical covariance matrices, differing only in the value of their mean vectors. In the present context, it is unlikely that the covariance matrices are equal. The hypothesis of equality may be tested, however, and if found to be rejected, a quadratic weighting function may be used in computing the parameter estimates (Eisenbeis and Avery, 1972). The second, potentially more damaging problem is that many of the attributes used as independent variables are discrete, and thus would tend inherently to follow a multinomial distribution……" (Wiginton, 1980)

The assumption of the linear regression model that variables have linear relationships, however, usually does not hold and is deviated from the multivariate normality assumption, i.e. the data is independent and normally distributed. In comparison, the advantage of the logit model is that it predicts dichotomous outcomes and linear relationships between variables in the logistic function, without the necessary requirement of multivariate normality assumption, which allows for some parametric distribution data. There are many studies showed that most of the consumer credit scoring datasets are only weakly nonlinear and because of that linear regression and logistic regression both gave good performance (Baesens et al., 2003b). However, even when the assumptions of linear regression models are satisfied, the logit model is almost as efficient as linear regression model (Harrell and Lee, 1985). Moreover, linear regression has better classification ability but a worse prediction ability, whereas a logit model has a relatively better prediction capability (Liang, 2003). Additionally, as computer science develops in a high speed today, the requirement of using the maximum likelihood method for a logit model, instead of ordinary least square, is not considered difficult to meet.

In general, no overall best statistical technique is used in building credit scoring models, for what best depends on the problem details, the data structure, the customers' characteristics, the extent to which it is possible to segregate the groups by using the characteristics, and the objective of classification (Hand and Henley, 1997). Recent studies that compared the new advanced techniques such as neutral networks and sophisticate algorithms with classical statistical methods found that newer techniques perform better in having a higher average correct classification rate; but no evidence reveals that the simple classical methods most widely used in practice do not perform well, they work not statistically different from other techniques (Baesens et al., 2003).

In China, most of the banks are still using traditional ways – a judgmental system in the evaluation process. Credit scoring models are on trial in a limited number of financial services in China, whereas the techniques in modern banks have become mature and advanced. We filled in the gap by acknowledging and implementing such techniques using the empirical study from developed countries to guide the immature financial markets in China.

The rest of the thesis is organized as follows: in section 2 we review the existing literature on both the techniques and the economic benefits of credit scoring; section 3 introduces the current credit scoring application in developed countries; in section 4, we estimate the logit model by using the mortgage data from the US; section 5 presents the current situation in China and some implications to China's credit development; section 6 discusses the benefits of credit scoring in economic development; section 7 concludes.

Chapter 2 Literature Review

A large number of studies on personal and enterprises credit scoring have been conducted by using various methods. Many of them applied two or three scoring methods to one practical issue. Wiginton (1980) proposed the maximum likelihood estimation of the logit model as an alternative method for the linear regression model. After comparing the two models in an actual data experiment, the paper comes to the conclusion that logit function is preferred in developing credit scoring model than the linear regression function. Abdou et al. (2007) used three statistical techniques (linear regression, probit analysis, logistic regression) in evaluating an Egyptian bank's personal loan data-set. They compared the predicting ability of these models and came to the results that the ranking of models varies according to the bank's decision criterion. Abdou et al. (2008) compared neural network and the other three conventional statistical techniques (as stated above) in an Egyptian bank's personal loan data-set. The conclusion is the same that the chosing of models depends on the bank's view. Aumeboonsuke and Dryver (2012) compared the performance of a linear regression model with a logit model by using three sets of simulated population data. After the comparison of horizontal and vertical analysis as well as the selection of different levels of cut-off point, the paper concluded that no single solution to credit scoring could be made. Samreen and Zaidi (2012) evaluated the credit risk in commercial banks of Pakistan by using linear regression model and logit models and compared their performances with the newly created credit scoring model to assess individuals' creditworthiness. The results show that regarding to the accuracy of classifying good loans from the bad, the newly created one is better than the logit model and linear regression model.

Others documented and compared the classical scoring methods with the advanced ones, and evaluated their performance by empirical evidence. Glennon et al. (2007) developed a new credit scoring model, validated and compared the performance of traditional parametric, semi-parametric and non-parametric models. They found little difference between these models, and concluded that to rank the individuals by creditworthiness is easier than to predict actual default rates by models. Hand and Henley (1997) reviewed the statistical methods and issues; Mester (1997) summarized the scoring methods and the applications in banking sectors; Nick (2000) gave an overview of credit scoring techniques; Thomas (2000) surveyed the statistical and operational research techniques; Vojtek (2006) introduced various credit scoring methods; Crook et al. (2007) surveyed recent studies of scoring methods and some concerns on consumer credit risk assessment; Abdou and Pointon (2011) reviewed 214 studies on credit scoring techniques applied in various areas and concluded that no overall best techniques for all circumstances. Desai et al. (1996) explored the ability of neural networks by comparing the methods with classical statistical ones and found that “if the measure of performance is percentage of good and bad loans correctly classified, logistic regression models are comparable to the neural network approach”; Desai et al. (1997) investigated the predictive power of scoring models and found that the new techniques are not outperforming the traditional ones.

Some studies have been done on applying the logistic method in country's banking sectors. Dinh and Kleimeier (2007) proposed a logit model for Vietnam's retail banking market to manage the credit risk. Steenackers and Goovaerts (1989) used a logistic regression model to develop a numerical scoring system for personal loans in a Belgian credit company. The result shows that the company can adjust the cut-off point which depends on the percentage of loans they want to accept. And Lawrence and Arshadi (1995) used actual problem loan files from 52 banks in 25 states in the US to design a multinomial logit model.

Besides, the literature on credit scoring discussed the economic benefits of credit scoring to economic development. Mester (1997) introduced that credit scoring is encouraged in the US mortgage market for underwriting consistency and for cost-effectiveness. Additionally, credit scoring makes small-business lending more popular due to more loans could be approved without increasing default risk as well as it is more feasible to make securitization of the loans. Blöchlinger et al. (2006) studied the profit-maximizing cut-off and the price curve from using credit scoring. Thus the commercial banks can improve the profits and gain economic benefits. TransUnion (2007) summarized the credit scoring's economic benefits that it reduces the decision costs, reduces the moral hazard rate and expands applicants' access to credit. Parisi (2010) claimed that one of the biggest barriers for many companies is their credit constraints. He proposed that today's techniques can solve the three main problems (accuracy, cost and technology) in promoting credit scoring, thus increasing the sales and attracting more customers by making the loan approval process efficient and by loosening the credit limit.

Chapter 3 Credit scoring in developed countries

Credit scoring system has been functioning in modern banks of developed countries for years and benefits them from default risk and in gaining profits. After nearly 200 years' development, the credit service in these countries has formed their own unique pattern. The management for credit in these countries can be divided into three major forms:

The first is the enterprises credit management system which is established by the central bank. The credit registration system includes registration information of enterprises and individual credit information, which is mainly for banks' internal use to prevent the risk of loans and to help the financial supervision and monetary policy decisions. The second is the business credit system which is market-oriented and operated by private credit companies, which is represented by the US and U.K. For example, the American Management Association, Dun & Bradstreet and other well-known companies, which operated as the main US credit management system, provided comprehensive and compensable services including credit investigation, credit rating, credit consulting, commercial credit, educational seminars and publications to the society. The third is the membership mechanism of credit institutions established by bank association. In this mode, credit institutions provide credit service to the membership banks, and these banks have to report customers' information to the institutions.

Take the United States as the example of the market-oriented credit system first. US is both the birthplace of the credit card and the country that has the most developed personal credit scoring system. Twenty years ago, the three agencies (Experian, Equifax and TransUnion) in California designed the credit scoring system which is based on personal credit reports to predict the probability of payment of consumers, and to score them. This method and system developed into the current system of personal credit history database and credit score report. So far, they have collected about 450 million credit history data of customers. In general, when Americans talks about "your score," it usually refers to your current FICO score. FICO credit scoring system is established in US in 1956. and derived scores range between 300 and 850 points. The higher the score, the smaller the default risk, thus a loan applicant is more probable to be approved with a higher score. Besides, the credit scores will also affect an applicant's loan interest rate. Applicant who has a higher credit score will probably get a lower interest rate on his/her loan. For example, the average data of Experian reveals that applicants with a high score of 740 points will get the car loan rate for approximately 3.2%; while the rate for those with 680-739 points will rise to 4.5%; those with lower credit score will have to pay an interest rate of 6.5%-12.9%.

The FICO scoring model considers five main factors and the percentages are based on the importance of the five factors for the general population presented in **Figure 1** below. The customer's credit payment history constitutes 35%, the current level of indebtedness makes up 30%, the length of credit history constitutes 15%, the types of credit in use makes up 10%, and the newly opened credit account accounts for 10%.

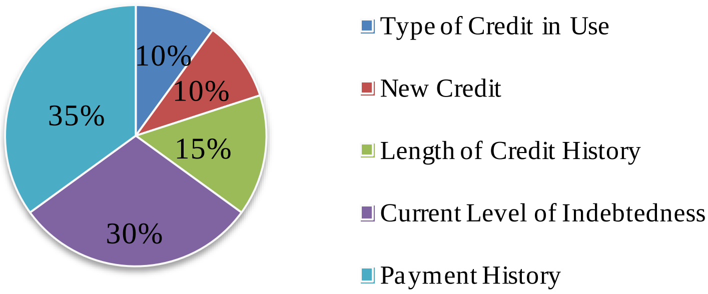

Figure 1 The breakdowns of weights of the five factors in FICO credit score

(Source: http://www.my.co.com/CreditEducation/WhatsInYourScore.aspx.)

In 2012, 75% of the US loan application decisions are based on credit score reports. US consumers reported 11 loan obligations to the credit bureaus: 40% of the lenders are overdue for 30 days, 20% of them for 60 days, more than 85% people are not overdue for three months. The longest age of account in average is 12 years; the average rate of loans on lines of credit is amounted to 34%. The default rate of lenders with 760 points' credit score is about 1%; the ones with less than 550 points will default in a probability of 24%; the default rate of borrowers with 550 to 599 points is 12%; and the ones with 600 to 619 points has a default rate of 8%. A well-designed personal credit system promotes the development of the US personal financial services. US consumer spending is accounts for about two-thirds of the GDP. In 2012, the amount of consumer expenditure and mortgage loans increased by about 1.1 trillion US dollars. From 2001 till now, the contribution of consumer expenditure to the economy reached 85%. In the first quarter of 2005, the balance of 7598 commercial banks' consumer loans reached 2.45 trillion US dollars, accounting for 50 percent of US bank loans (the bank loan assets account for 58% of total assets, the securities assets account for 22% of the total assets). Among them, the rate of consumption loans over the balance of loans in 445 banks with more than $ 1 billion's assets is 52%. In average, every household owns 13 credit cards; the amount of overdraft is up to 9205 US dollars. Alan Greenspan, the chairman of the Federal Reserve of US stated that the significance of the use of credit scoring technology has gone far beyond the original credit risk assessment; they can be used to evaluate the risk-based customers' profitability, to develop the initial and ongoing lines of credit, and to help detect fraud and reduce losses. This will improve the efficiency of the loan granting system, increase the willingness of granting loans, thus will play a significant role in attracting profitable customers.

Take the credit scoring applications in small business loan (which is similar to personal loans) as examples. Unlike traditional ways of borrowing in the neighborhood for good knowledge, credit scoring changes the way banks make small business loans. Large banks enter the market using credit scoring and automated centralized system (Allen and Scott, 2007). Automated small-business lending allows banks to profitably make loans and able to extend more loans than judgmental systems without increasing their default rates (Asch, 1995). Credit scoring may also encourage more lending because it gives banks a tool of accurate risk pricing. For example, the break-even loan size at Hibernia Bank was about $200,000 before automation, but now it has a large portfolio of loans under $50,000 (Zuckerman, 1996). In 2007, a Pennsylvania's regional bank made a mail campaign to 50,000 current and prospective customers by using credit scoring model. A simple application form with no financial statements was used, and up to $35,000 loans were approved based solely on credit scores. Additionally, PNC Bank opened an automated loan center in suburban Philadelphia. It processed 25,000 small-business loan applications in a single year, the process of which is automated and using credit score methods from across the nation (Oppenheim, 1997). As we can see, the spread of credit scoring leads to an increase in small business lending in the US.

Germany is the representative of the first mode, where the credit system is mainly funded by the government to establish a national database, to organize a national research network, and to form a central bank credit registration system. The German government concerns more about the protection of personal privacy with more stringent personal data protection laws. The differences between Germany's government-led mode and the US market-oriented credit system are reflected in three aspects: firstly, the credit system is served as a department established by the central bank, rather than set up by private companies; secondly, banks are required to provide customers' credit information to the government credit bureau; thirdly, the central bank assumes the major supervision role.

In Germany, there are two types of credit institutions. One is similar to associations or clubs which are jointly built by major financial institutions and other information providers but have no entity connection with banks. The members of the organization determine the manner and type of sharing information. Any credit institution who wants to get the credit information from it must become a member of the organization first. Another type is funded by the major credit information providers, which are also the customers of these credit institutions. For example, the SCHUFA (German: Schutzgemeinschaft für allgemeine Kreditsicherung; English: Protection company for general creditworthiness) in Germany is established by its main information providers. In this company, 95% of the data comes from the customers, only 5% of the data comes from the courts, post offices and other public institutions. Besides, 85.3% of the shares are held by banks and other financial institutions, the remaining are held by trade or mail and other companies. The company's major customers are also shareholders. In 1997, this company introduced credit ratings services which designed a different scoring system according to different requirements of customers. It provided comprehensive information on consumers; the system will automatically score the consumers when they inquire. The advantage of the mutual-type credit institutions is that they can easily get support from banks which hold great information of customers. In all, German credit market will grow steadily and rapidly with the growing number of credit agencies and the maturing of database.

Chapter 4 Model estimation in US mortgage market

We applied the logit model in a practical real credit dataset and tried to find the variables that influence the default rate as well as to get the credit scoring model.

4.1 Data Sources

The data is collected from the official website of Freddie Mac which is the Federal Home Loan Mortgage Corporation for secondary market mortgage in US. The Single Family Loan-Level Dataset we used to apply in the model is the sample dataset of loan-level credit performance data master-serviced by Freddie Mac in 2012. The sample dataset is a random sample of 37,500 loans selected from the full Single Family Loan-Level Dataset in 2012. In general, it is an unrepresentative sample: only approved and closed workouts (e.g., short sales, modifications, and deeds-in-lieu of foreclosure) prior to the performance-cutoff-date are included in the dataset. It includes loan-level origination and monthly loan performance data on a portion of the fully amortizing 30-year fixed-rate Single Family mortgages: the origination data file contains loan-level origination information for all the loans originated during the quarter; the monthly performance data file contains monthly loan-level credit performance information for each loan, starting from the time of loan acquisition by Freddie Mac until the earlier of a termination event or the Performance Cutoff Date. We dropped 40 loan files in the origination file but not in the performance file, which are the cases that the loan gets paid off in the month of origination or before first cycle begins, and we match the performance data with the origination ones according to the loan sequence number. The delinquent status is the value corresponding to the number of days the borrower is delinquent, based on the due date of last paid installment reported by servicers to Freddie Mac.

STATA 11.2 is used for our estimation. **Table 1** shows the description of dependent variables and original explanatory variables with the denotation in STATA below. The explanatory variables in our estimations include all variable data which are available, in case that these data may have potential influence on the loan performance.

	Variables Specification	Denotation in STATA
Dependent variables	Current loan delinquency status 1/0 for delinquent or not for more than 30/60/90 days	d30/ d60/ d90
Explanatory variables	Credit score (FICO) ranged from 301-850	credit_score
	Mortgage insurance percentage (MI %) ranged from 1%-55%	mi_p
	Number of Units denotes the unit of mortgage property ranged from 1-4	no_unit
	Number of borrowers ranged from 1-2	no_borrower
	Original debt-to-income (DTI) ratio ranged from 0%-65%	dti
	Original unpaid principal balance (UPB)	upb
	Original loan-to-value (LTV) ranged from 6%-105%	ltv
	Original interest rate (%)	oir
	Loan purpose (C=cash-out refinance, N=no cash-out refinance, P=purchase)	p1=C, p2=N

Table 1 The description of the dataset variables
(Source: https://freddiemac.embs.com/FLoan/Data/download.php)

4.2 Model Construction

We estimate our models using logistic regression. Logistic regression is a useful way of describing the relationship between independent variables and a binary dependent variable, that has only two possible values—the variable equals to 1 if the observation is the applicant who 'delinquent', and the variable equals to 0 if the applicant who ‘does not delinquent’. Here is the model construction.

$p_i = E(y_i \mid x_i) = 1/(1+\exp(-\beta' x_i))$ for each individual i, where $y_i \in \{0,1\}$ is an indicator variable for non-default (non-delinquent) or default (delinquent), x_i is a vector of covariates, and β is the vector of associated coefficients. The estimated $\hat{p}_i$ of the probability of default are derived from the estimated model $\hat{p}_i = 1/(1+\exp(-b' x_i))$, where b is the maximum likelihood estimator of β.

Let $Z = b' x_i$, then $Z = ln(\hat{p}/(1-\hat{p}))$ represents the estimated log-odds.

$Z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k$ where β_0 is the intercept and $\beta_1, \beta_2, ..., \beta_k$ are the regression coefficients of $x_1, x_2, ..., x_k$, respectively.

Since D90 within one year is a common performance proxy for the default rate (more precisely, if a loan is delinquent for 270 days, it goes into default), and because of the small ‘delinquent’ sample proportion with only approved loans included in the dataset, our estimations predict the probability of going 30 days past due (DPD)/ 60 DPD/ 90 DPD instead of actual default rate during the single calendar year of 2012 after origination. Here is the formation of the logit equation:

$$\text{Probability } (\text{D90} \setminus \text{D60} \setminus \text{D30 in 2012}) = \text{F } (\text{origination variables, error terms})$$

4.3 Model Estimation

The descriptive statistics of the dependent variables are presented in **Table 2**. Only 0.11% of the applicants are delinquent for more than 90 days. Even when we extend the delinquency range to more than 30 days’ delinquency, we only get 1.30% of the ‘delinquent’ sample, which is a small proportion from the whole sample.

Dependent variables	1= 'delinquent' / percentage		0= 'not delinquent'/ percentage	
d90	41	0.11%	37459	99.89%
d60	76	0.20%	37424	99.80%
d30	488	1.30%	37012	98.70%

Table 2 The descriptive statistics of dependent variables

One of the explanatory variables we are most interested in is the FICO credit score. The summarize statistics is shown in the **Table 3** below. From the table we can see that all applicants get a high score above 603 points, which is consistent with the fact that only a few applicants who delinquent for more than 90 days, and that most of the declined bad loans are already excluded from the dataset.

Variable	Observations	Mean	Standard Deviation	Min value	Max value
credit_score (301~850)	37500	765.1565	37.7770	603	832

Table 3 The descriptive statistics for FICO score

We use the available nine variables (credit score, number of borrowers, number of units, unpaid principal balance, debt to income ratio, loan to value ratio, original interest rate, mortgage insurance percentage and the loan purpose) to run the logistic regression and the results are reported in **Table 4**. Since the number of unpaid principal balance (UPB) is large which makes the influence on delinquency rate tiny, we divided it by 10,000 for convenience and get the variable 'upb'.

	(1) d90	(2) d90	(3) d60	(4) d60	(5) d30	(6) d30
main						
credit_score	-0.0434***		-0.0269***		-0.0146***	
	(-9.54)		(-9.60)		(-13.76)	
no_unit	0.619	0.292	0.0715	-0.0918	0.348**	0.279*
	(1.36)	(0.70)	(0.16)	(-0.21)	(2.61)	(2.12)
upb	-0.142	-0.288	-0.130	-0.191	0.0818*	0.0540
	(-0.93)	(-1.95)	(-1.23)	(-1.84)	(2.31)	(1.54)
no_borrower	-0.689*	-0.481	-0.795**	-0.675**	-0.561***	-0.518***
	(-2.08)	(-1.49)	(-3.25)	(-2.78)	(-5.98)	(-5.53)
dti	4.937*	5.769**	6.104***	6.653***	0.785	1.559**
	(2.32)	(3.02)	(3.89)	(4.56)	(1.57)	(3.20)
ltv	5.072**	2.889	0.886	0.456	-0.253	-0.203
	(2.58)	(1.76)	(0.78)	(0.44)	(-0.65)	(-0.54)
oir	34.72	238.1***	119.0***	240.6***	69.32***	130.6***
	(0.74)	(6.01)	(3.54)	(8.07)	(4.51)	(8.89)
mi_p	-0.707	-1.011	1.864	1.701	-0.121	0.0140
	(-0.33)	(-0.51)	(1.36)	(1.29)	(-0.22)	(0.03)
p1	1.247*	0.951*	0.459	0.460	-0.0202	0.0337
	(2.51)	(2.03)	(1.34)	(1.39)	(-0.15)	(0.25)
p2	1.561***	1.110**	0.569	0.313	-0.0305	-0.123
	(3.73)	(2.76)	(1.92)	(1.08)	(-0.27)	(-1.09)
_cons	17.35***	-20.24***	6.906*	-17.44***	4.128***	-9.491***
	(4.15)	(-9.13)	(2.39)	(-10.96)	(3.49)	(-14.37)
N	37500	37500	37500	37500	37500	37500

t statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001

Table 4 Explaining the effect of loan variables on the three delinquency status

As can be seen from the results in **Table 4**, FICO score has a statistically significant influence at the 0.1% level on predicting delinquent under the three delinquency status. Take the first logit equation on 90-day delinquency rate as an example. A high credit score will lead to a low probability of delinquency. Every additional point's increase in an applicant's credit score will decrease the predicted log odd of delinquency for 0.0434. That is, the predicted delinquency rate of an applicant is $e^{-0.0434} = 0.9575$ times than that of an applicant who get a one-point-lower credit score. Since we divided the UPB by 10,000 for convenience, we can see from the table that every ten thousand's increase in the unpaid principal balance will decrease the delinquency probability by 0.142. Additionally, the more applicants for a loan, the smaller an applicant's debt-to-loan ratio, the smaller an applicant's loan-to-value ratio will lead to a higher probability of delinquency. Also when estimating on 90-day delinquency rate, the loan purpose dummy variables p1 and p2 both have significant effects at 5% level. The results reveal that a no cash-out refinance mortgage (p2) borrower is more probable to delinquent than a cash-out refinance mortgage (p1) borrower or a purchase one on 60-day and 90-day. But a purchase mortgage borrower is more likely to delinquent on 30 days' period.

As the proportion of delinquency sample size become larger, the significance of effects from explanatory variables varies as the dependent variables change from 90-day to 30-day. Most of the variables such as number of borrowers, number of units, unpaid principal balance, and original interest rate become more significant from 90-day to 30-day. Concerning about the endogeneity of the variable, especially the explanatory variable FICO score may have simultaneity problem with the independent variable D90/D60/D30 since the delinquency conditions may be considered in the process of establishing the score. Thus we drop the FICO score and estimate the remaining variables on the three delinquency status. It also can be seen from **Table 4** that some of the variables become significant after taking out the score. For example, the DTI is significant at 1% level under every status, and the original interest rate becomes significant then on 90-day delinquency rate. Variables like LTV on 90-day and UPB on 30-day respectively become insignificance then. That means an applicant's FICO score has correlation with his/her original debt-to-income ratio, the original interest rate when applying for a loan, the loan-to-value ratio and the original unpaid principal balance. As the coefficients of these variables become larger after the drop, we consider FICO scores covered a lot of the effects of these variables.

We also evaluate the marginal effects of variables before and after dropping the FICO score from 90-day to 30-day. After the drop, the negative influence of UPB on delinquency rate increases, the positive influence of DTI and LTV also increase a lot. Specially, the positive marginal effect of original interest rate increases for approximately 40 times; and the effect from DTI and LTV increase for about 6 and 3 times respectively. The reason is also related to the process of establishing the FICO credit score, which may have considered the loan information when scoring a loan and the applicant, and the score has covered the effect of all the other explanatory variables on one's delinquency rate.

4.4 Variable Analysis

Due to the lack of personal information data such as the income, occupation, marriage status, age and other loan related factors that may have significant influence on predicting one's delinquency probability and may have included in the process of creating FICO scores, we are unable to establish an instrumental variable to further study the endogeneity problem of the equations. However, it is still meaningful for enterprises to manage the credit risk by explaining the influence of some of the variables (LTV, DTI, number of borrowers) on predicting the probability of delinquency/default.

Original unpaid principal balance (UPB) of the mortgage is the remaining portion of the monthly payment applied to interest and principal. It is calculated by using original loan amount minus the total of loan payments to data plus the total interest paid to data. Since the next period's interest is derived from the unpaid principal balance at the end of the preceding period, it may have positive and significant influence on short periods as shown in 30-day's regression. But this influence is not significant during longer periods like 60-day and 90-day, and we can see that the effect of UPB on the probability of delinquent become negative then.

Original loan-to-value (LTV) is the ratio obtained by dividing the original mortgage loan amount on the note date by the lesser of the mortgaged property's appraised value on the note date or its purchase price in the case of a purchase mortgage loan; in the case of a refinance mortgage loan, the ratio obtained by dividing the original mortgage loan amount on the note date by the mortgaged property's appraised value on the note date; in the case of a seasoned mortgage loan, if the seller cannot warrant that the value of the mortgaged property has not declined since the note date, the seller will provide a new appraisal value to calculate the LTV. The results show that, to 90-day and 60-day, the larger the ratio, the more probable an applicant will delinquent. But in short period such as 30-day, this trend is not obvious, sometimes revealing an adverse effect.

Original debt-to-income (DTI) is based on the sum of the borrower's monthly debt payments, including monthly housing expenses that incorporate the mortgage payment the borrower is making at the time of the delivery of the mortgage loan to Freddie Mac, divided by the total monthly income used to underwrite the borrower. It makes sense that DTI always positively influences the probability of delinquent. A larger amount of debt with little income will give pressure to borrowers, thus makes them feel difficult to repay the loan.

Original interest rate determines the total amount of money to repay, thus it has a positive effect on delinquency in both short and long period. The number of borrower denotes only whether there is one borrower or more than one borrower associated with the mortgage note. More borrowers to one loan will share the responsibility to pay the loan as well as supervising each other in repaying, which will make the probability of delinquency lower. The loan purpose indicates whether the mortgage loan is a cash-out refinance mortgage[2] no cash-out refinance mortgage[3] or a purchase mortgage.

To conclude, the estimation of our model has several obstructions which make the result biased. First is the less detailed credit information. Since the bad loans have already been dropped from the dataset, our estimation only relies on the data of accepted applicants which makes the sample proportion of delinquent status restricted and sample selection bias problem. If we estimated with the bad loans sample, the results should be more accurate and there should be less variables that should have influence on delinquency rate but omitted in the estimation. Second is that the FICO scores are known at origination. If we have access to the original data which is used for estimating FICO scores, we will be able to reduce the simultaneity bias.

---

2 A cash-out refinance mortgage loan is a mortgage loan in which the use of the loan amount is not limited to specific purposes. A mortgage loan placed on a property previously owned free and clear by the Borrower is always considered a cash-out refinance mortgage loan.

3 A no cash-out refinance mortgage loan is a mortgage loan in which the loan amount is limited to the uses: pay off the first mortgage, regardless of its age; pay off any junior liens secured by the mortgaged property, that were used in their entirety to acquire the subject property; pay related closing costs, financing costs and prepaid items; and disburse cash out to the borrower not to exceed 2% of the new refinance mortgage loan or $2,000, whichever is less.

Chapter 5 Situation in China

As the Chinese credit data are limited, we are not able to analyze the credit scoring model for the situation. However, the importance of credit scoring cannot be disregarded. By analyzing the current situation, we will be able to promote the benefits of credit scoring to the Chinese financial market in a better way.

With the globalization and the intensified volatility in financial markets in recent years, challenges from credit risk management become more and more important in China's financial markets. Financial institutions need to be more rigorous and careful in assessing the personal credit history, credit behavior and other factors to determine whether to do business with a particular client. Due to the low degree of data sharing and a lack of scientific and reasonable assessment systems, most of the domestic financial institutions are still using relatively backward ways to evaluate customers, facing high credit costs, low operating efficiency and great risk of fraud.

As the personal credit scoring system in joint-equity commercial banks developed faster than the state-owned commercial banks in China, we take one of the joint-equity commercial banks, Min Sheng Bank's current credit scoring system as an example. They use five levels (AAA, AA, A+, A and B) to evaluate the loan applicants. The analyzing content includes the applicants' natural condition, financial solvency, ability of indemnification and the relationship with the bank. Analysts make the decision based on these characters of the applicants. Specific criteria are as follows: personal credit score consists of two components. The first component includes natural conditions, solvency and ability of indemnification. Specially, natural conditions include the applicant’s age, residence, family structure, education, occupation, working experience, housing conditions; solvency includes the applicant’s income, financial assets, other assets; ability of indemnification includes applicant’s medical insurance, endowment insurance and accumulation fund; credit conditions include credit history, the loan-to-deposit ratio, and the bonus in the particular bank. The second component is bad financial records of the applicant. The personal credit score is the sum of the two components. The higher the credit score, the lower the risk. The specific criterion of scoring is presented in **Table 5** at the end of this chapter.

Although some of the banks in China are still using the rough and simple traditional methods, credit scoring has already started and achieved initial success in some of the domestic commercial banks. China Merchants Bank Company introduced the Fair Isaac credit systems for credit approval process; Guangdong Development Bank introduced SAS's risk rating system. The use of the credit risk management tools helps grasp the customer's credit information and strengthen risk management in China’s banking sectors.

In 2002, People's Bank of China and Chinese government jointly launched the business credit registration and consulting system and achieved a national network query. The system is queried for 320 000 times a month in 2002, and the average monthly inquiries are 1,100,000 times in 2003. In 2004, it reached 200 million times a month. China gradually built up the data collecting system for enterprises and individuals. In 2006, the People's Bank of China built a national credit information database on enterprise and individuals. By the end of 2007, the database established files for more than 13,700 companies and nearly 600 million households. Commercial banks will check the applicant's credit history when they apply for loans for buying a car, a house or applying for credit cards. Personal credit data has been accumulated from financial institutions which provided more than 100 million people's personal credit reports.

In February 2010, the Virtual Economy and Data Research Center of Chinese Academy of Science reported its 2007 to 2009 results that the national personal credit rating system developed by Professor Shi Yong with the People’s Bank of China, has been on trail in a large number of commercial banks in China, and will cover 1.3 billion people’s credit activities. This scoring model considered indicators from individual’s performance in the economic and social activities as well as their occupations, wages, and other hundreds of variables to give the final credit scores. Different from the US standard whose score ranged from 300 to 850 points, China initially identified 350-1000 points as the standard score. Besides, the database include individual credit information for about 0.65 billion people, in which 0.1 billion is complete and useable.

Such a situation of slow development in credit scoring in China has its historical reasons. In the early China, banks are owned by the government. The banking sector is a monopolized market. The government had the power to order banks to lend money to the national mainstay industry no matter whether these enterprises will default and have no money to repay. The interest rate of loans is also decided by the government rather than by the market. As a result, credit scoring is unnecessary in the early China in some sense. The obstacles and improvements to establish an effective credit scoring system are as follows. Most importantly, the availability to relevant individual credit information is limited. Just like the industrial revolution opened up the material transactions, opening the data is the trend of the times. And the benefits from the era of big data greatly exceed the issues it brings to us. Currently the individual credit information is mainly provided by some of the commercial banks, information from other public institutions cannot be easily obtained. The reviewing methods from different departments are different, which makes it hard to ensure the authenticity and completeness of customers' information. The opaque of personal income tax system makes it difficult for financial institutions to determine the borrowers' property by only analyzing their history information. Since credit history does not include taxation, insurance and other non-banking credit information, the integrity of credit data cannot be fully guaranteed. The era of big data is to get valuable products, services and profound insight through massive data analysis in an unprecedented way (Victor, 2013). The obstacles to develop the 'big data' in China are the data 'liquidity' and 'accessibility'. The US government created Data.gov website opens the door for big data; Britain and India also have 'disclose data' activities. To catch up with such a large revolution, China should try to make the data publicized for enterprises and individuals.

Moreover, the professional individual credit rating agencies should be set up. As China's financial market is a bank-based system, the central bank plays a main role in credit evaluation. The credit database's development and management are all led by the central bank, which restricts the development of professional non-bank credit assessing agencies. As the marketization of interest rate in recent years caused fierce competition between banks, banks become more cautious on setting the interest rates of deposits and loans. Thus different people with different credit conditions will obtain loans from banks in different interest rates. As a result, it is hard for central bank to make the credit evaluation for various financial institutions; more professional agencies that focus on different groups of institutions and applicants should be built for credit analysis.

Furthermore, an effective credit scoring system is needed for evaluation. In China, individual credit scoring is still new, and is currently limited to the level of aggregated personal credit information. There are no established scientific evaluation criteria and assessment methods which are built according to China's situation. China is different from western countries in economic, political, historical and cultural aspects which should be considered in choosing evaluation methods.

Finally, due to individual credit factors are complicated and the privacy of data and other concerns, many credit researchers only focus on corporate credit, ignoring the in-depth study of individual credit scoring. More research should be encouraged to conduct on individual credit scoring.

About the applicant's housing conditions (15 points)		About the applicant's economic support (34 points)		About the applicant's personal stability (27 points)		About the applicant's background (24 points)	
Own the house	8	≥6,000 CNY (monthly income）	26	Public official	16	Local	5
				Institutional employee	14	Nonlocal	2
No house	0	3000-6000 CNY	22	State-owned business	13	Junior high school and below	1
				Joint-stock enterprise	10		
Rent the house	4	2000-3000 CNY	18	Other works	4	High school	2
Welfare house	4	1000-2000 CNY	13	Retired	16	Secondary School	4
Mortgage Secured	7	300-1000 CNY	7	Unemployment with social assistance	10	University and above	5
Unsecured	0			Unemployment without social assistance	8		
		No debt	8	Stayed at the present house for ≥6 years	7	Female aged ≥30	5
						Male aged ≥ 30	4.5
		10-100 CNY's monthly debt	6	Stayed for 2-6 years	5	Female aged ≤30	3
				Stayed for ≤2 years	2	Male aged ≤30	2.5
		100-500 CNY's monthly debt	4	Unmarried	2	No investigation on credit	0
				Married without children	3	No Record on credit	0
		≥500 CNY's monthly debt	2	Married with children	4	Default for once	0
						Default for more than twice	-9
						No default	9

Table 5 The traditional criteria of credit scoring in China's Min Sheng Bank (Source: http://creditcard.cmbc.com.cn/)

5.1 Implications to China's financial markets

The credit scoring is originated in the US. The phenomenon that one third of the bonds default in the end except for those high-rated ones during the Great Depression in 1930s contributes to the government and investors' emphasis in credit scoring. However, China's domestic credit rating industry sprouted in the late 1980s. As the lag in the institutional infrastructure and the limitation in the auditing securities and trading equity, China's credit development is rather slow. The comprehension of credit scoring in China remains in the traditional moral level. People think credit as a moral standard. Borrowing money for consumption has not yet been widely accepted and the saving rate has been maintained between 37% and 42%, thus the contribution of consumption to GDP is lower than US by about 20%. People are more accustomed to make transactions through cash than credit in order to guard against risks. As a result, it is applicable for China's commercial banks or mortgage companies to use the proposed model as a part of the credit evaluation process. Based on the model, the default probability can be estimated, banks can set a cut-off point and reject those with a high probability of default and make loans with a better rate. By adopting the model, more financial services will be able to reduce the high-risk loan and manage the risk of default properly.

To adapt our empirical results to the credit conditions in China, the applicants' credit information for evaluation should be relatively different from that in the US. In China, most of banks focus more on personal information than loan information because of the limitation in credit information. As the maturing of credit system, both of the information could be considered in the process of evaluation. A study from the Institute of Economics Chinese Academy of Social Sciences reveals that several unique variables with Chinese characteristics are important in credit scoring. Apart from loan information variables (payment history, amounts owed, length of credit history, new credit and types of credit used) that are included in the US FICO credit score evaluation process, the ignored variables whose data can be access to could also be included in China's credit scoring system. The ignored variables in the FICO evaluation include sex, marital status, age, education background, salary, occupation, title, employer, employment history, residence, transportation, securities and insurance.

Salary is the most powerful and persuasive factor in predicting default. We believe that people with higher salary have higher and more stable ability to pay the loan.

Age is related with applicant's ability of earning and amount of disposable wealth. Generally in China, applicants under 20-year-old represent a low credit and a high risk of default limited by the work experience and education level. Applicants who are older than 50 years will also face a high probability of default because of the decrease in income and the health problems after the retirement.

People believe that women default less frequently on loans than men (Schreiner, 2004) because of their attitude of risk aversion. In China there's a specific relationship between gender and income which in turn might be predictive of default. In China, women will get retired earlier than men for 5 years, that is, women's mid-career is around 30-40, while men will reach it around 35-45, thus their income level varies with both gender and age.

Marital status will matter if it has effects on the maturity, the reliability of borrowers as well as the number of dependants the borrowers has to support in a family. More children in a family indicate more school fees and other living expenses which will bring pressure to borrowers in repayment. On the other hand, in China, marriage means stable lives which will make the couple work harder and more responsible.

Education background is usually directs proportional to one's income. With higher education level, the ability to earn is relatively higher, thus the higher ability of repayment.

Occupation and title are more or less Chinese-characteristic factors that will affect one's default rate. In China, salaries from a state-owned enterprise and that from private enterprise might be in large difference. Besides, the rank of title in China is also linked tightly with the income, thus influences the default probability.

Residence and region might reflect applicants' stability of living status. Whether applicants own their home, or rent, or subsidized by the government, or live with their parents will reflect their income in some extent. This phenomenon may vary from countries. For example, Cook et al. (1992) find that in the US borrowers living with their parents are least likely to default. In China, as the prizes of house increase, owning one house might reveal heavy burden of loans to pay in the following years or reveal high level of income. Besides, the living expenses in different regions in China might also represent ones' income and pressure of payment. For example, the consumption cost is much higher in Beijing than in remote areas like northwest of

China.

By incorporating these variables into the credit scoring model, future research could make more persuasive prediction in consumer's credit performance.

Chapter 6 Credit scoring and economic development

Resource scarcity is the fundamental economic problem in a world of limited resources. As stated in Paisi (2010), credit constrain is one of the significant problems for enterprises as well. From the view of economic development, credit scoring system not only plays a positive role in banks' decision-making process (as explained in Mester (1997)'s paper, credit scoring is consistent and cost-effective) but also makes resource allocation more efficient thus expanding the credit constraint and promote economic growth, which fully reflected in the booklet of TransUnion (2007).

On the one hand, credit scoring is beneficial to the banks in several aspects such as reducing the cost and time of evaluation process, increasing sales and customer service, and enable regulatory compliance. Banks need to make profit by providing loans, so it is important for them to choose profitable targets to lend money to. Credit scoring helps banks to do the screening. Besides, bad loans will blow customers' confidence in banks, making it hard for banks to get more deposit from customers. Credit scoring makes the result more accurate thus reducing bad debt.

On the other hand, it also enhances the overall efficiency of the society. First, credit scoring relaxes the credit constraint and expands the access to credit thus increasing the wealth of the society. From the experience of the US, the occurrence of credit default is the main cause of rising bad loans. A resource from the US in January 2009 indicates that the default rate reaches 7.1% which is unprecedentedly high. Due to the fact that some of the credit card companies cannot cover the loss of about $56.8 billion of debt and the threat of short of money or even bankruptcy, they reduce the credit limit and close the account to protect the company. By using a judgmental system, banks are cautious to lend money to borrowers concerning that the borrowers may default and lead to losses of the banks. Therefore, many banks set a conservative credit constraint for safety. However, those enterprises or individuals who will make profits may not be able to be approved as the limited of amount of loans. The credit scoring system improves and rationalizes the credit decision process of banks, relaxes the credit constraint and increases all the enterprises' profits and individuals' welfare. The system enables banks to make accurate, fast and competitive decisions to approve more applications and expand access to credit. In this way, people who are poor or unemployed will be able to get the financial resources in need; since little money may make a big increase in their welfare, the consumption demands in the whole economy will increase, thus promoting the economic growth. The small or medium enterprises will also be able to access consumer credit which provides financial resources for their activities when business loan are unavailable.

Secondly, the credit scoring system allocates existing resources effectively. Without credit scoring, it is hard for banks to price the products according to borrowers' risk level. Banks use the traditional judgmental system may face the problem of subjective and imprecise decisions, then the costs for both low-risk and high-risk customers to financial resources are high. By increasing access to credit and reducing credit costs, credit scoring spreads risk more fairly, prevents the misallocation of resources, thus fairly pricing the cost of financial resources, thus increasing the efficiency of the society.

Thirdly, for the national economy, credit scoring helps families break the generation cycle of low economic status by increasing their access to homeownership. This is one of the most important steps in the accumulation of wealth. Besides, credit scoring plays a role in promoting economic expansion and preventing recession by reducing liquidity constraints. It helps consumption smoothing during cyclical periods of unemployment as well as reduces the swings of the business cycle. Credit scoring also makes it possible to bundle credit products with loans according to the measured risks, thus encouraging customers' consumption in the secondary markets and increasing the availability of capital to be invested in the economy.

Chapter 7 Conclusion

In this thesis, we did a tentative work in applying a logit model to one of the representative mortgage companies in US. By using the delinquency status—30, 60 and 90 days past due, we estimate several loan performance and origination variables. The results reveal that apart from the loan information contained in the FICO credit score, several other variables such as the original debt-to-income, original interest rate also have significant influence on predicting the loan delinquency/default rate. We also analyzed Chinese-characteristic credit variables and the economic benefits from credit scoring. It is tempting to conclude that we could do better by using better 'data': if we have access to effective credit data from the Chinese market, we will also be able to predict the probability of delinquency/default and managing the credit risk in China.

Technology progress facilitates the economic growth in many aspects. As a kind of new technique, credit scoring plays a significant role in increasing the welfare of the society. The economic benefits credit scoring can bring are considerable, which can be applied in various institutions apart from banks and real estate market. It will help the financial departments enhance the efficiency and control the risk, thus avoiding a financial crisis. China is encouraged to adopt the new innovation and improve the credit policy, making the access to data easier and more systematically, thus helping the improvement of credit scoring condition.

In the future, more advanced techniques or integrated statistical models, such as neural networks, are recommended for use in the credit market, especially in China.

Besides, more data and relevant variables can be added to the model for accurate prediction results. China is suggested to make the credit information data available and standardize the analytical variables reporting platform for convenience. At the time when researchers have access to the real dataset, further research could be done to examine our result in Chinese financial market by using the real data and check the similarity with the performance in US. New models could also be designed for China's credit market.

Bibliography

Abdou, Hussein A., and John Pointon. "Credit scoring, statistical techniques and evaluation criteria: A review of the literature." *Intelligent Systems in Accounting, Finance and Management* 18.2-3 (2011): 59-88.

Abdou, Hussein, Ahmed El-Masry, and John Pointon. "On the applicability of credit scoring models in Egyptian banks." *Banks and Bank Systems* 2.1 (2007): 4-20.

Abdou, Hussein, John Pointon, and Ahmed El-Masry. "Neural nets versus conventional techniques in credit scoring in Egyptian banking." *Expert Systems with Applications* 35.3 (2008): 1275-1292.

Aumeboonsuke, Vesarach, and Arthur Lance Dryver. "Developing Credit Scoring Models When Small Sample Sizes Are Available." *Journal of Business Review, Cambridge* 20.1 (2012): 138-143.

Avery Robert B., Bostic Raphael W., Calem Paul S., and Canner Glenn B. "Credit risk, credit scoring, and the performance of home mortgages." *Fed. Res. Bull.* 82 (1996): 621.

Barakova, Irina, Dennis Glennon, and Ajay A. Palvia. "Sample Selection Bias in Acquisition Credit Scoring Models: An Evaluation of the Supplemental-Data Approach." *Available at SSRN 1722382* (2013).

Berger, Allen N., and W. Scott Frame. "Small business credit scoring and credit availability." *Journal of Small Business Management* 45.1 (2007): 5-22.

Bielecki, Tomasz R., and Marek Rutkowski. *Credit risk: modeling, valuation and hedging*. Springer, 2002.

Bierman Jr, Harold, and Warren H. Hausman. "The credit granting decision."*Management Science* 16.8 (1970): B-519.

Blöchlinger, Andreas, and Markus Leippold. "Economic benefit of powerful credit scoring." *Journal of Banking & Finance* 30.3 (2006): 851-873.

Bluhm, Christian, Ludger Overbeck, and Christoph Wagner. *An introduction to credit risk modeling*. CRC Press, 2002.

Bolton, Christine. "Logistic regression and its application in credit scoring." (2010).

Boyes, William J., Dennis L. Hoffman, and Stuart A. Low. "An econometric analysis of the bank credit scoring problem." *Journal of Econometrics* 40.1 (1989): 3-14.

Brandenburger, Thomas. *A Markov multinomial regression model for predicting consumer credit risk*. South Dakota State University, 2010.

Crook, J. N., Hamilton, R., and Thomas, L. C. A comparison of discriminations under alternative definitions of credit default. In L. C. Thomas, J. N. Crook, & D. B. Edelman (Eds.), Credit scoring and credit control. Oxford: Oxford University Press. (1992): 217-245.

Crook, Jonathan N., David B. Edelman, and Lyn C. Thomas. "Recent developments in consumer credit risk assessment." *European Journal of Operational Research* 183.3 (2007): 1447-1465.

Desai, Vijay S., Conway D.G., Jonathan N. Crook, and Overstreet G. "Credit-scoring models in the credit-union environment using neural networks and genetic algorithms." *IMA Journal of Management Mathematics* 8.4 (1997): 323-346.

Desai, Vijay S., Jonathan N. Crook, George A., and Overstreet G. "A comparison of neural networks and linear scoring models in the credit union environment." *European Journal of Operational Research* 95.1 (1996): 24-37.

Dinh, Thi Huyen Thanh, and Stefanie Kleimeier. "A credit scoring model for Vietnam's retail banking market." *International Review of Financial Analysis*16.5 (2007): 471-495.

Emel, Ahmet Burak, Oral Muhittin, Reisman Arnold, and Yolalan Reha. "A credit scoring approach for the commercial banking sector." *Socio-Economic Planning Sciences* 37.2 (2003): 103-123.

Hand, David J., and William E. Henley. "Statistical classification methods in consumer credit scoring: a review." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 160.3 (1997): 523-541.

Jacobson, Tor, and Kasper Roszbach. "Bank lending policy, credit scoring and value-at-risk." *Journal of banking & finance* 27.4 (2003): 615-633.

Lando, David. *Credit risk modeling: theory and applications*. Princeton University Press, 2009.

Lawrence, Edward C., and Nasser Arshadi. "A multinomial logit analysis of problem loan resolution choices in banking." *Journal of Money, Credit and Banking* (1995): 202-216.

Lee, Tian-Shyug, and I-Fei Chen. "A two-stage hybrid credit scoring model using artificial neural networks and multivariate adaptive regression splines."*Expert Systems with Applications* 28.4 (2005): 743-752.

Lee, Tian-Shyug, Chiu Chih-Chou, Lu Chi-Jie, and Chen I-Fei. "Credit scoring using the hybrid neural discriminant technique." *Expert Systems with applications* 23.3 (2002): 245-254.

Longenecker, Justin G., Carlos W. Moore, and J. William Petty. "Credit scoring and the small business: A review and the need for research." *A paper presented in USASBE 1997 National Conference, San Francisco, California*. 1997.

Malhotra, Rashmi, and D.K. Malhotra. "Evaluating consumer loans using neural networks." *Omega* 31.2 (2003): 83-96.

Min, Jae H., and Young-Chan Lee. "A practical approach to credit scoring." *Expert Systems with Applications* 35.4 (2008): 1762-1770.

Orgler, Yair E. "A Credit Scoring Model for Commercial Loans." *Journal of Money, Credit & Banking (Ohio State University Press)* 2.4 (1970).

Parisi, Jeff. "Commercial Credit & Collection Scoring – Part I & II." *Business Credit*, 107.4 (2005):18-19.

Pohar, Maja, Mateja Blas, and Sandra Turk. "Comparison of logistic regression and linear discriminant analysis: a simulation study." *Metodolski Zvezki* 1.1 (2004): 143-161.

Press, S. James, and Sandra Wilson. "Choosing between logistic regression and discriminant analysis." *Journal of the American Statistical Association*73.364 (1978): 699-705.

Pulina, Manuela. "Consumer behaviour in the credit card market: a banking case study." *International journal of consumer studies* 35.1 (2011): 86-94.

Rosenberg, Eric, and Alan Gleit. "Quantitative methods in credit management: a survey." *Operations research* 42.4 (1994): 589-613.

Rud, Olivia Parr. *Data mining cookbook: modeling data for marketing, risk, and customer relationship management*. John Wiley & Sons, 2001.

Steenackers, A., and M. J. Goovaerts. "A credit scoring model for personal loans." *Insurance: Mathematics and Economics* 8.1 (1989): 31-34.

Šušteršič, Maja, Dušan Mramor, and Jure Zupan. "Consumer credit scoring models with limited data." *Expert Systems with Applications* 36.3 (2009): 4736-4744.

Thomas, Lyn C. "A survey of credit and behavioural scoring: forecasting financial risk of lending to consumers." *International Journal of Forecasting*16.2 (2000): 149-172.

Thomas, Lyn C., David B. Edelman, and Jonathan N. Crook. *Credit scoring and its applications*. Siam, 2002.

TransUnion. “The importance of credit scoring for economic growth.” (2007)

Viganó, L. A credit scoring model for development banks: An African case study. Savings and Development, 4 (1993): 441−482.

Vojtek, Martin, and E. Kocenda. "Credit scoring methods." *Czech Journal of Economics and Finance* 56.3-4 (2006): 152-167.

Wiginton, John C. "A note on the comparison of logit and discriminant models of consumer credit behavior." *Journal of Financial and Quantitative Analysis* 15.03 (1980): 757-770.

Yobas, Mumine B., Jonathan N. Crook, and Peter Ross. "Credit scoring using neural and evolutionary techniques." *IMA Journal of Management Mathematics* 11.2 (2000): 111-125.