The Probability of Chinese Mortgage Loan Default and Credit Scoring

A thesis

submitted in partial fulfilment

of the requirements for the Degree

of

Master of Commerce and Management

at

Lincoln University

By

Weizhuo Wang

Lincoln University
2010

ABSTRACT

Abstract of a thesis submitted in partial fulfilment of the
requirements for the Degree of M.C.M.

The Probability of Chinese Mortgage Loan Default and Credit Scoring

By Weizhuo Wang

Credit scoring is broadly applied in consumer lending especially in credit cards and mortgages. Credit scoring has not been widely used in business lending because business loan differ substantially across household borrowers, making it more difficult to build up an accurate scoring method. However, this has changed. The complexity and flexibility of statistical models and advanced computing technology have made such credit scoring possible in business lending. Thus, many banks are using credit scoring model to evaluate business loan applications, which is a cost effective credit management tool (Mester, 1997).

Preceding credit scoring valuation is important against the increasing financial risk, and adopting credit risk analysis method to value the consumer credit is a necessary step to resist serious credit loss (Li and Zhang, 2003). It is important for banks to understand that every borrower is a potential lemon and the probability to default is high. This is evidenced in today's U.S. subprime loan problems. In essence, a credit scoring model provides an objective estimate of a borrower's credit risk. Since a lender can not observe the borrower's probability to default, credit scoring models enable lending institutions to rank potential customers according to their default risk, and improve the allocation of loan resources.

Similarly, as China becomes more consumer credit culture and people start to borrow money for homes and other investments, an international risk standard management framework becomes more important for Chinese domestic banks. The efficiency advantage of credit scoring models certainly will help the banks to meet the growth in Chinese consumer loans and mitigate default risk.

The primary problem of any lender is to differentiate between "good" and "bad" debtors prior to granting credit. Such differentiation is possible by using a credit-scoring method. In this study, we examine how this inference problem impacts mortgage borrowers' characteristics on the probability of loan default. This includes volume of loan granted and loan price which is the interest charged on loans. We use a credit scoring model to investigate the bank mortgage lending policy in China during the period 2004 to 2009.

The results show Age group, Education level, Occupation type, and Region positively impact the amount of credit granted. Gender, Age group, Marital status, Annual income, Bank rating, Occupation type, Loan duration and Region have negative impacted on loan pricing. The findings also suggest that it is necessary to review the borrowers' creditworthiness periodically, as the changes in economic condition could affect the loan performance. The results are generally consistent with the existing literature.

Keywords: Credit scoring, Mortgage loans, Default, Logistic regression

ACKNOWLEDGEMENTS

The completion of this research is impossible without the support and encouragement of my supervisor, Dr. Christopher Gan under whose supervision I chose this topic and began the thesis. To Dr. Christopher Gan, who is always there to listen and to give advice, thanking him for all the patience, dedication, guidance, insightful comments, and invaluable suggestions throughout this research. Dr. Li Zhaohua, my associate supervisor had also been abundantly helpful, and has assisted me in numerous ways, including data collection, and analysis of the research findings. They are not only my academic supervisors, but also my minor advisors who had shown me different ways to approach a research problem and the need to the persistent to accomplish any goal. Sincere thanks also to staffs of Faculty of Commerce and Lincoln Library for their help and support.

I would also thank all my friends who had shared my joy and frustration, and support me while writing this thesis. I would like to thank Helen, Hill, and Nancy, who assisted me on the statistical methodology analysis, and Dong Guihua, who helped me in the data collection process.

I cannot end without thanking my family, on whose constant encouragement and love I have relied throughout my time at the university. Their unflinching courage and conviction will always inspire me. It is to them that I dedicate this work.

CONTENTS

Chapter 1

INTRODUCTION

1.1 Introduction

Credit scoring was one of the earliest financial risk management tool developed. It was used by the U.S. in the 1950s with early applications of portfolio analysis to manage and diversify the risk inherent in investment portfolios (Thomas et al., 2002). Credit scoring models are complex and often vary among creditors and for different types of credit. Credit scoring models are used as a tool for underwriting and administering all forms of retail credit, including credit cards, direct and indirect instalment loans, residential mortgages, home equity loans, and small business credit (OCC Bulletin, 1997). The models offer a cost-efficient way to make sound decisions based on bank or industry experience. Different types of credit scoring models are used for various activities. For example, credit scoring models can be used effectively to control risk selection, manage credit losses, evaluate new loan programs, improve loan approval processing time, and ensure that existing credit criteria are sound and consistently applied.

Credit scoring is a method combining the borrower's historical and statistical techniques to evaluate the credit risk of loan applications (Turvey and Brown, 1990). It is a statistical approach to predicting the probability that a credit applicant will default or become delinquent (Berger and Frame, 2007). A credit scoring system awards points for each factor that helps predict whether a borrower is likely to repay a loan. The total number of points--a credit score--helps to predict how creditworthy a borrower is, that is, how likely it is that a borrower will repay a loan and make the payments when due (National Consumer Law Center, 2005).

Credit scoring is broadly applied in consumer lending, especially in credit cards and mortgages. Credit scoring has not been widely used in business lending because business loan differ substantially across household borrowers, making it more difficult to build up an accurate scoring method. However, this has changed. The complexity and flexibility of statistical models and advanced computing technology have made such credit scoring possible in business lending. Thus, many banks are using credit scoring model to evaluate business loan applications, which is a cost effective credit management tool (Mester, 1997).

Most banks in China use judgemental techniques in assessing credit risk. The judgemental system depends on the experience and common sense of the credit evaluator. In addition, this method of assessing credit involves subjectivity of the credit assessing officer. This loan approval process can not ensure lenders are applying the same underwriting criteria to all borrowers regardless of race, gender, or other factors (Mester, 1997). Judgemental assessment has some serious limitations, for example, the difference in loan officers' experiences lead to different views regarding the relationships between risk and specific credit characteristics of loan applicators. Consequently, lending institutions can not ensure loan officers' approval of loan applications is consistent with the risk objectives of the institution (Avery et al., 1996).

As China becomes more consumer credit culture and people start to borrow money for homes and other investments, an international risk standard management framework becomes more important for Chinese domestic banks. The efficient credit scoring models certainly will help banks to meet the growth in Chinese consumer loans and mitigate default risk. For example, mortgage lending in China has been growing at a rate of 54.9% between 2000 and 2004. By November 2009, the total outstanding housing loans in China have reached RMB5.24 trillion (Anonymous, 2009). Preceding credit scoring valuation is important against the increasing financial risk, and adopting credit risk analysis method to value the consumer credit is a necessary step to resist serious credit loss (Li and Zhang, 2003). It is important for banks to understand that every borrower is a potential lemon and the probability to default is high. This is evidenced in today's U.S. subprime loan problems. In essence, a credit scoring model provides an objective estimate of a borrower's credit risk. Since a lender can not observe the borrower's probability to default, credit scoring models enable lending institutions to rank potential customers according to their default risk, and improve the allocation of loan resources.

For years, banks have been using credit scoring systems to determine whether a borrower is a good risk for credit cards and auto loans. Following this, credit scoring has been used to help banks evaluate a borrower's ability to repay home mortgage loans and whether to charge deposits for utility services. Many auto and home insurance companies use special credit scores to decide whether to issue a policy and for how much (National Consumer Law Centre, 2005).

1.2 Background of Chinese mortgage lending

Real estate industry is an important pillar industry in China. China is a developing country with a population of 1.3 billion, and housing is a basic need for people's livelihood as well as a public good. The development of the real estate industry closely relates to the overall economic growth. If real estate bubbles burst due to economic or political factors, housing prices will plummet before economic slowdown at a faster rate than overall economic growth, inflicting greater impacts on consumer confidence and economic growth than any other industry (Liu, 2007). It is important for banks to mitigate risk associated with real estate loans. In China, the Construction Bank of China (CBC) is the leader in mortgage lending (Anonymous, 2008). Table 1.1 shows CBS classifications of Chinese mortgages loans.

Table 1.1 Classified Real Estate Loans in China

Classification	Purpose of the loan
Personal housing loan	Used to support the individuals in cities and towns of mainland China to buy, build house and housing repair.
Second home loan	Used to support individuals purchase secondary market trading houses.
Housing provident fund loan	Use of bank credit and housing accumulation fund to buy various types of housing but who with full civil capacity and in full deposit housing provident fund loans.

Sources: Construction Bank of China

The Chinese housing loan reforms started in 1980. The main orientation of China's housing policy over the last twenty years has moved away from the traditional system of welfare allocation (*fuli fenpei*) to a system of monetised allocation (*huobi fenpei*) of housing benefits. This policy encourages people to buy apartments and become home owners (Burell, 2006). These changes in housing policy result in privatisation of public housing and the monetisation of housing benefits which placed new entrants on the housing market in a vulnerable position.

The Chinese people rely on their own earnings and try to save enough money to purchase houses (Burell, 2001). In addition, the people's salary has increased, but the prices of new houses rise much faster. For example, the price for urban housing in China was set between 500 to 2,000 RMB per square meter in 1992 (Wu, 1996). However, in late 1990s most large cities houses had reached 3,000 RMB per square meter. In large cities such as Shanghai and Beijing, the housing prices are higher and rise more rapidly (Burell, 2006). Majority of the urban households have their wages increased, but have not kept up with the rise in housing costs (see Table 1.2) (Burell, 2006). This leads to an increase demand for housing loans.

Table 1.2 Average annual per capita and household earnings and their housing purchasing power→ in two largest Chinese cities

Cities	Individual average annual earnings (RMB)	Couple average annual household earnings (RMB)	Time needed for a 60,000 RMB down payment.30 of HH earnings(years)	Time needed to purchase a 300,000 RMB apt. 30% of HH earnings(years)
Beijing	18,157	36,314	5.3	27.7
Shanghai	21,957	43,914	4.7	22.7
National	11,152	22,304	9.0	44.7

Source: China Labour Statistical Yearbook 2003,

Chinese housing loan grow rapidly over the past decade. According to the statistics from the People's Bank of China (2003), the balance of personal housing loans topped 1.18 trillion RMB ($142.51 billion in 1998), 26.64 times the figure in 1998 (Wang, 2004). At the end of October 2007, outstanding residential housing loans reached 4.69 trillion RMB, a growth of 30.75%, up 1.01 trillion RMB from the beginning of the year, accounting for 28.9% of total new RMB loans of commercial banks during the same period. Outstanding individual housing loan reached 2.6 trillion RMB, 619.2 billion RMB more than the beginning of the year 2007 (Liu, 2007). This leads to a large demand for Chinese mortgage loans.

After China's accession to the WTO foreign lenders are allowed to participate in Chinese banking markets. Both domestic and foreign banks recognise the opportunity in housing loans in China's housing market. According to the statistics from the People's Bank of China (2003), the balance of personal housing loans topped 1.18 trillion RMB ($142.51 billion), 26.64 times the figure in 1998 (Wang, 2004).

The competition between domestic and foreign banks can intensify the underlying risks in the housing market. For example, as banks try to capture a larger share of the housing market, some commercial banks lower their lending criteria, reduce loan application steps and relax the borrowers' background investigation through distorted practice (Liu, 2007). In 2006, local officers of some commercial banks lower lending criteria, reduce examination steps and relax credit background investigation. For example, the Chinese housing loans in 16 cities showed on average 22.31% of the borrowers did not meet the loan officer directly when getting loans, including in Beijing, Hangzhou and Guangzhou (Liu, 2007). At the end of August 2007, the outstanding refinance mortgages and additional loans in major cities reached 41 billion RMB, 3.5 times of that in the same period of 2006. Thus, it is necessary to strengthen credit information management in the real estate sector to fully leverage the credit information system that includes both enterprises and individuals (Liu, 2007).

Compare to foreign banks, the credit assessment and management tool of Chinese banks are outdated. Most Chinese banks currently depend on the experience and common sense of credit evaluators in assessing credit risk. These evaluators can not ensure lenders are applying the same underwriting criteria to all borrowers (Mester, 1997). For example, the evaluators in Chinese banks overemphasise collateral and largely ignore the borrowers' ability to repay the loans. Thus, these will lead to underestimating the risks confronting most Chinese banks (Rodman, 2006). According to the experience from the Western countries, the market would develop from cash transaction to credit transaction. Chinese banks still lag behind the Western modern management credit system. In China, cash transactions still play a big role in the market (Li and Zhang, 2003).

Therefore, understanding the mortgage loan default factors in Chinese mortgage loans are urged to help improving the loan default rate. The rationality of banks lending decisions includes reducing lending costs and increasing repayment rate and operating profits to the banks. As a consequence, a well-developed mortgage market would promote mitigate credit default risks. A well developed credit scoring system would provide the reasonable information to decision maker and identify the consumer credit risk in time is necessary (Li and Zhang, 2003).

Table 1.3 Sources & Uses of Credit Funds of Financial Institutions (by Sectors)

Unit: RMB100 Million Yuan

Total Loans 2007	Jan.	Feb.	Mar.	Apr.	May	Jun.	Jul.	Aug.	Sep.	Oct.	Nov.	Dec.
Loans to Resident Sector	40229.32	41097.77	42176.26	43412.63	44307.08	45596.42	46698.89	47981.52	49423.75	50105.83	50636.52	50652.26
Consumption Loans	24721.77	25152.50	25672.60	26260.76	26857.74	27842.87	28695.64	29711.69	30915.96	31624.79	32333.66	32729.00
Short-term Consumption Loans	2060.41	2118.10	2214.37	2330.11	2490.61	2669.92	2727.87	2833.70	2953.96	3038.09	3062.39	3104.11
Medium & Long-term Consumption Loans	22661.36	23034.40	23458.23	23930.65	24367.13	25172.95	25967.77	26877.99	27962.00	28586.70	29271.27	29624.89
Total Loans2008												
Loans to Resident Sector	51948.65	52057.46	52990.01	53718.38	54411.26	55198.94	55688.50	56192.54	56806.57	56846.35	56492.42	57057.92
Consumption Loans	33590.86	33587.80	33962.14	34268.33	34723.89	35309.33	35670.48	35996.43	36410.29	36535.82	36681.27	37210.29
Short-term Consumption Loans	3199.08	3117.76	3192.90	3255.64	3326.97	3465.55	3522.16	3612.36	3771.47	3864.52	3964.79	4136.86
Medium & Long-term Consumption Loans	30391.78	30470.04	30769.24	31012.69	31396.92	31843.78	32148.32	32384.07	32638.82	32671.30	32716.48	33073.43
Total Loans 2009												
Loans to Resident Sector	58415.34	58855.71	61423.56									
Consumption Loans	37935.36	38001.52	39392.92									
Short-term Consumption Loans	3996.58	3945.49	4366.85									
Medium & Long-term Consumption Loans	33938.78	34056.02	35026.07									

Source: People's Bank of China

1.3 Research Problem Statement

Understanding the mortgage loan default factors in Chinese mortgage loans help to improve the development of credit accessibility to households including mitigating default risk. The rationality of banks lending decisions includes reducing lending costs and increasing repayment rate and operating profits to banks. As a consequence, a well-developed mortgage market would promote economic growth.

When people cannot commit to pay back their loans and there is limited information about their characteristics, lending institutions must draw inferences about their likelihood of default. The primary problem of any lender is to differentiate between "good" and "bad" debtors prior to granting credit. Such differentiation is possible by using a credit-scoring method. In this study, we examine how this inference problem impacts mortgage borrowers' characteristics on the probability of loan default. This includes the volume of loan granted and the loan price which is the interest charged on loans. We use a credit scoring model to investigate the probability of Chinese mortgage loan default during the period 2004 to 2009.

The specific research objectives include:

- To identify the factors in mortgage lending performance of Chinese banks
- To investigate the factors affecting the volume of credit granted by Chinese banks
- To investigate the factors affecting the mortgage interest rate charged by Chinese banks

1.4 Significance of study

This research is expected to contribute to the development of Chinese mortgage market in additional to analysing the rational behaviour of mortgage loan default processes. Understanding the mortgage loan default factors in Chinese mortgage loans help to improve the development of credit accessibility to households including mitigating default risk. The rationality of banks lending decisions includes reducing lending costs and increasing repayment rate and operating profits to banks. As a consequence, a well-developed mortgage market would promote economic growth.

In addition, this research would benefit both lender and borrowers. Instead of using subjective evaluation decision-rules, which are bias and unreliable, lenders can apply an objective evaluation technique with a standard process and criteria to appraise their customer's credit risks and creditworthiness. A good credit risk management tool can effectively to control risk selection, manage credit losses, evaluate new loan programs, improve loan approval processing time, and ensure that existing credit criteria are sound and consistently applied.

Borrowers can predict whether they qualify for new loans or an extension of existing loans. They will be able to estimate the credit availability and the price of credit corresponding to their risk level. The information can enhance the borrowers' decision-making process when they acquire loans. Self-assessment by borrowers also benefits the credit suppliers as it can monitor the default risk and reduces administrative costs.

Our study uses the credit scoring model to examine the borrower's characteristics and banks loan granting decision and the borrower's default risk in China. The credit scoring model is applied to new loan applications and helps to determine after the loan is granted whether the borrower will default, and which borrower's characteristic affect the loan price.

1.5 Outline of thesis

The rest of this thesis is organised as follows. Chapter 2 presents an overview of the relevant literature and the theatrical background on credit scoring, credit availability model and loan pricing model. Chapter 3 describes the empirical models, the estimation techniques, the data, and the data collection method. Chapter 4 presents the results and discussion of the results generated by the analysis. Chapter 5 summarises the major findings followed by the limitation of the research and recommendations for the future study.

Chapter 2

LITERATURE REVIEW

This chapter reviews the relevant on mortgage loan default model (credit scoring model), demand for credit, and cost of credit. The chapter is structured as follows: Section 2.1 discusses the probability of mortgage loan default and credit evaluation techniques. Section 2.2 discusses the concepts of credit scoring and explains how credit scoring works. Section 2.3 discusses the benefits and limitations of credit scoring models. The variables and modelling techniques used in credit scoring models are summarised in Sections 2.4 and 2.5. Section 2.6 identifies the demand for credit and credit availability model, and Section 2.7 discusses the interest rate and loan pricing model.

2.1 Mortage loan default analysis: judgemental versus credit scoring

The probability of mortgage loan default depends upon the borrower credit risk that is the probability of the borrower not repaying the loan. Credit analysis includes the valuation of the financial history and financial statements of the applicant credit background. It is the primary method used in appraising credit risks. The objectives of credit analysis are to determine the financial strength of the borrower, to estimate the borrower's probability of repayment, and to reduce the risk of non-payment to an acceptable level. There are two major problems in credit analysis: the assessment of all important factors about an applicant simultaneously and the evaluation of all applicants objectively (Sinkey, 2002; Plata and Nartea, 1998). The factors used to assess loan applicators include monthly income, outstanding debt, financial assets, whether the applicant has defaulted or is over delinquent on a previous loan (Schreiner, 2000), and the loan applicant's subjective factors, such as age, gender, education, marital status, income, and other personal characteristics that will affect the applicant's loan performance (Boyes et al., 1989). The objectives of loan assessment are to predict the probability of loan default and mitigate the default risk (Jacobson and Roszbach, 2003).

There are two main techniques used to evaluate a borrower's creditworthiness (Crook, 1996): the loan officer subjective assessment (judgemental technique) and the credit scoring technique.

Creditworthiness is judged based on the characteristics of an individual that makes him or her qualify for a loan; and one who is not creditworthy will be unqualified for the loan (Lewis, 1992). The subjective assessment of a borrower's creditworthiness is based on 6 C's- Character, Capacity, Cash, Collateral, Conditions and Control (Rose, 1993) (see Table 2.1).

Glassman and Wilkins (1997), Crook (1996), and Lewis (1992) argue judgmental assessment of credit is inefficient, unexplainable, inconsistent and non-uniform. Traditional methods of deciding loan granting base on experience of previous decisions and use human judgement of the risk default. However, economic pressures resulting from increased demand for credit, allied with greater commercial competition and the emergence of new computer technology; have led to the sophistically models to aid the credit granting decision (Hand and Henley, 1997). Credit scoring methods produce more accurate classifications than subjective judgemental assessments by human experts. Rosenberg and Gleit (1994) and Chandler and Coffman (1979) discuss the credit scoring method advantages over the judgemental assessment. For example, credit scoring can increase efficiency by leaving loan officer to concentrate only on ambiguous cases, and it also allowed the lenders to review the creditworthiness of the borrowers periodically (Glassman and Wilkins, 1997). Thus, credit scoring models have become a preferable technique in credit risk appraisal.

Altman (1968) was the first to use a statistical model to predict probability of a borrower's to default aimed at identifying the borrower credit risk more objectively. Following this, many statistical credit scoring models have been developed, such as logistic regression, neural networks, smoothing nonparametric and expert systems and have been widely used in assessing credit risk (Hand and Hanley, 1997).

Table2.1 The Six basic C's in lending

Character	Capacity	Cash	Collateral	Conditions	Control
• Customer past payment record • Experience of other lender with current customer • Purpose of loan Customer track record in forecasting • Credit rating • Presence	• Identify of customer and guarantors • Copies of charters resolutions, agreements, and other documents bearing on the legal standing of the borrowing customer. • Description of history, legal structure, owners, natural of operations, productions, and principal customers and suppliers for a business borrower	• Past earnings, dividends, and sales record. • Adequacy of projected cash flow. • Availability of liquid reserves. • Turnover of payables, receivables, and inventory. • Capital structure and leverage. • Expense controls. • Coverage ratios. • Recent performance of borrower stock and P/E ratio. • Management quality. • Content of auditor report and statement footnotes. • Recent accounting changes.	• Ownership of assets • Age of assets • Vulnerability to obsolescence • Liquidation value • Degree of specialization in assets. • Liens, encumbrances and restrictions. • Leases and mortgages issued • Insurance covered • Guarantees and warranties issued • Bank relative position as creditor • Lawsuits and tax situation • Probable future financing needs	• Customer position in industry and expected market share • Customer performance vis-à-vis comparable firms in industry • Competitive climate for customer product • Sensitivity of customer and industry to business cycles and changes in technology • Labour market conditions • Impact of inflation on customer balance sheet and cash floe • Long-run industry outlook • Regulation; political and environmental factors	• Applicable banking laws and regulations regarding the character and quality of acceptable loans • Adequate documentation for examiners • Signed acknowledgements and correctly prepared loan documents • Consistency of loan request with bank written loan policy • Inputs from non-credit personnel (such as economists or political experts) on external factors affecting factors affecting loan repayments

2.2 What is credit scoring?

Credit scoring is used by U.S. retailers and mail-order firms in the 1950s with the early application of investment portfolios to manage and diversify borrowers default risks (Thomas et al., 2002). Nowadays the credit scoring models are becoming one of the most successful techniques of modelling in finance and banking (Abdou et al., 2007). Credit score is based on statistical analysis of the borrowers' credit files, using borrowers' historical data and statistical techniques; it tries to isolate the effect of various characteristics of applicators on delinquencies and defaults (Frame et al., 2001; Mester, 1997; Glassman and Wilkins, 1997; Turvey and Brown, 1990).

The techniques and practices of statistical credit scoring came to be applied by lenders, unevenly and in a relatively unplanned manner, to the problem of reducing losses due to the non-repayment of credit loans (Marron, 2007). Credit scoring models can assist banks to make lending decisions. Credit scoring can supplement or even replace the traditional subjective assessment of pertinent information in applicant's report, the statistically derived measure of the credit risk associated with a given credit history allow lenders to better and more quickly assess the strengths and weaknesses of applications (Avery et al., 1996).

Credit scoring is broadly used in consumer lending, especially in credit cards, and has became more commonly used in mortgage lending. The advanced computer technology increased data accessibility for business loans and this have made such scoring applicable in complex business loans. Thus, more and more banks are using credit scoring to evaluate loan applications, as credit scoring tends to standardise loans and make applicators' default risk more predictable (Mester, 1997). As banking markets in developing countries are maturing, banks need to face competitions from both domestic and foreign banks. The credit scoring models could give banks the substantial growth of retail credit and increased regulatory attention to risk management (Dihn and Kleimeier, 2007).

Credit scoring system is a computerised procedure generating a number of points (a score) according to a number of the borrowers' relevant characteristics, such as income, profession, age, wealth, previous loans, etc. The total score is obtained by summing the individual borrower's score. If the score is higher than a lender's "cut-off-level", credit will be granted, otherwise the credit will be refused (Steenackers and Goovaerts, 1989). The overall idea of credit scoring model is quite straightforward. Based on the statistical probabilities, the combinations of borrowers' characteristics differentiating "good" from "bad" generate a score (or probability) serve as an estimate of the risk level of each new loan when then lenders decide whether to make the loans or not (Crook, 1996). Crook (1996) argues that the aim of credit scoring is to predict risk, not to explain it. Thus, it is not necessary that the predictive model also explains why some borrowers default on the loan repayment and others do not.

Borrowers' credit history and other characteristics regarding repayment ability that are generally provided by borrower are electronically analysed. Credit models would predict the default risk of any loan granted based on previous experience with borrowers of similar loan profiles. A well-designed model should give high scores to borrowers whose loans would perform well and low scores to borrowers whose loans would not perform well. To develop a good credit scoring model, it is necessary to review the borrowers' credit worthiness periodically, as changes in economic condition could affect loan performances. In general, there is no best credit scoring models. It is possible some bad borrowers may get a high score and receive the loans, and vice versa (McAllister and Mingo, 1994). Jensen (1992) find that using credit scoring, approximately 8 percent of the applications would be approved when they were actually bad loans and 18 percent of the applications would be rejected when they were good loans. This is similar to the Type I and II errors in the hypotheses development.

In any credit market, borrowers have the option to default; defaulters are not exogenously excluded from future borrowing; there is free entry of lenders and borrowers; and lenders cannot collude to punish defaulters. Limited credit or credit at higher interest rates following default arises from the lender's response to limited information about the borrower's behaviour and earnings realisations. The lender learns from an individual's borrowing and repayment behaviour about his/her type and encapsulates his/her reputation for not defaulting in a credit score (Chatterjee, Corbae and Ríos-Rull, 2008).

2.3 Benefits and limitations of credit scoring

Credit scoring techniques have a number of benefits compared to judgemental techniques for both lenders and borrowers. The attributes in credit scoring that have led to its increasing use in loan evaluation include (Mester, 1997; Glassman and Wilkins, 1997; Crook, 1996; McAllister and Mingo, 1994; Lewis, 1992; Chandler and Coffman, 1979):

1. Increased efficiencies and reduced costs

   Credit scoring models allow immediate handling of the definite binary (Yes/No) decisions; this leaves more time for credit officers to focus on loans that least well-handled by the models. Credit scoring is also an efficient way to grant loan, the loan granting time is reduced to days or hours instead of weeks for consumers. Credit officer can handle more loan applications than in traditional loan assessment framework.

2. Reduced potential for bias

Credit scoring is a standard loan granting process. Lenders apply the criteria built into the model and applicants are measured against these criteria. The model considers the characteristics of both good and bad borrowers, but the judgemental methods are usually negatively biased towards bad borrowers. The model helps lenders ensure that the same underwriting criteria have been applied to all borrowers regardless of race, gender or other factors prohibited by commercial law used in credit appraisal.

3. Ability to control risk levels

Lenders use credit scoring to differentiate risk categories of loans. Instead of lender rejecting the high risk loans category, the credit analyst can estimate credit risk of the borrowers with reasonable degree of confidence and the cut-off score can be adjusted according the risk of the loan portfolio.

4. Lending system

Credit scoring models are base on the borrowers' previous history. The credit scoring system can re-estimate the model continuously base on the large amount of data. For lenders, the next applicant's behaviour can be better predicted base on the database and previous credit history recorded in the system. With the recent development of neural networks- "self-learning" computer programs, the level of the credit scoring process can develop in terms of complexity and accuracy is unknown. These models will be improved given time.

5. Risk monitoring

In cases which the bank uses a credit-scoring model, lenders are required to review the creditworthiness of the borrowers periodically. This credit review process is used to monitor the individual borrower's risk profile and to adjust aggregate loan loss reserves.

Although credit scoring can reduce costs and increase consistency to the loan granting process, the weakness of credit scoring should not be ignored. The limitations of credit scoring models include (Banasik, Crook, and Thomas, 2003; Crook, 1996; Glassman and Wilkins, 1997; Mester, 1997):

i. Data and accuracy of the model

Accuracy is an important consideration in using credit scoring. If the models are not accurate, these cost savings and other benefits of credit scoring could be negatively affected by poorly performing loans. Credit scoring models are complex; the models are only good if the data could feed them. Poor and inaccurate credit report information will invalidate results. The data used in credit scoring should include a sample of both well-performing and bad performing loans. The data also required to update regularly, as the model should be re-estimated frequently to ensure that changes in the relationship between potential factors and loan performance are captured.

ii. Knowing the customer

Borrowers' characteristics related to their likelihood of repayment and defaults. Although credit models attempt to predict a borrower's actual behaviour, there is no substitute for knowing the borrower. Most credit scoring models are embedded with human errors. Credit scoring will not replace the judgments of loan officers or loan groups based on informal qualitative knowledge. It is important in dealing with customer who has not had a pristine credit history or where the ability to repay the loan may be the primary factor on which a credit decision is based. In such cases, a working manual will be required to combine with credit scoring to make the lending decisions.

iii. Economic condition

An accurate model needs to make predictions in both expansions and recessions. Thus, the data used in the model should cover both good and bad economic periods. If the data on which the model is used does not include repayment behaviour in an economic downturn, lenders may face higher default risk than they have planned.

iv. Selection bias

Credit scoring models are used to predict the probability of default but the models are usually parameterised using a sample of accepted applicants only. This selection bias could lead to bias estimation by credit scoring model.

To avoid the bias in credit scoring model, Schreiner (2003) argues that banks should apply their credit scoring model to loans that are already conditionally approved by the credit officers. According to the First National Bank of Chicago, for small- business loans, about 25% of the applications rejected by the model were later approved by the credit officers. Thus, to fully extend the advantage of credit scoring and avoid bias, the bank's credit assessment can be a combination of the traditional lending via credit scoring models (Dinh and Kleimeier, 2007).

2.4 Variables commonly used in credit scoring models

The pragmatism and empiricism of credit scoring implies that any characteristics and environments of the borrower that has obvious connections with default risk should be used in the scoring system (Lewis, 1992). The variables should be sequentially added or deleted to maximise the model's predictive accuracy (Henley and Hand, 1997). There are two important standards for variable selection; first, the variables should have significant coefficients and contribute to explanation of the dependent variable's variance. Second, the variables should have close correlation with included variables (Dinh and Kleimeier, 2007). Lewis (1992) suggests that there is no need to justify the case for any variable. If it helps the predictions, it should be used.

However, the major factors commonly used in credit scoring models include the borrowers' income, age, gender, education, occupation, employer type, region, time at present address, residential status, marital status, home phone, collateral value, loan duration, time with bank, number of loans, and current account (Dinh and Kleimeier, 2007; Roszbach, 2004; Jacobson and Roszbach, 2003; Martinelli, 1997; Crook, Hamilton, and Thomas, 1992; Boyes, Hoffman, and Low, 1989; Capon, 1982;)

Income is a commonly used proxy of the borrower's financial wealthy and his/her ability to repay (Dinh and Kleimeier, 2007). There is a positive relationship between income and the borrowers' default rate; higher income is associated with lower default risk (Jacobson and Roszbach, 2003). Occupation is a common variable used in credit scoring model and is highly correlated with the borrowers' income level.

Education enhances the borrowers' ability to repay. The better educated borrowers are deemed to have more stable and higher income employment and thus a lower default rate. The borrowers' education level distinguished from post-graduate to non-high school graduate (Dinh and Kleimeier, 2007).

Employer refers to the type of company for which a borrower works such as stated-owned, foreign, joint-stock company, etc. The type of company a borrower works in could be a proxy for income level and stability. Missing values of this variable are also very informative as borrowers who do not answer this question show the highest probability of default.

Time with employer measures the number of years that the borrower has been working for the current employer. It reflects the satisfaction of the borrower with the current job. The higher the borrowers' job satisfactions, the more stable their employment will be and the higher their ability to repay their loans (Cook et al., 1992). The length of time with employer may discriminate against women, since women's length of employment reduces due to pregnancy and childbearing (Capon, 1982).

Age measures the borrower's age in years. Thomas (2000) and Boyle et al. (1992) confirm that older borrowers are more risk adverse, and therefore the less likely to default. Thus banks are more hesitant to lend to younger borrowers who are more risk averse.

Gender is a fair discriminatory - base on the statistical default rates of men versus women. There are ample evidences that women default less frequently on loans because women are more risk adverse (Coval et al., 2000).

Region means the area of the country that borrower lives. As people of similar wealth tend to live in the same location, the geographic criterion can indicate a borrower's level of financial wealth. Some suburb might attract richer residents and this could result increase in housing and property prices. This also affects the collateral value and probability of default.

The residential variable measures whether borrowers own their home, rent, or live with their parents. This could indicate the borrowers' financial wealth in the case of home ownership. Residential status also indicates financial pressure on borrowers' income, for example rental cost. Crook et al. (1992) find that borrowers living with their parents are less likely to default.

Time addresses the number of years that the borrowers have been living at their current address. According to Crook et al.'s (1992) study, the default risk drops with an increase in time at present address; it might be a proxy for the borrowers' maturity, stability, or risky aversion. Changing address might be a signal that a borrower's financial wealth is high or improving rapidly.

Marital status affects the borrower's level of responsibility, reliability, or maturity. The probability of default is higher for married than single borrowers. Dinh and Kleimeier (2007) discover that the marital status is typically related to number of dependants which in turn reflects financial pressure on the borrower and borrower's ability to repay a loan.

Collateral is a form of guarantee to support the loan. Borrowers' collateral can be a single of default risk, such as, if the loans that the house serves as collateral, the probability of default is very low. This is because the borrowers are risk adverse and fear of losing their house. Collateral reduces the bank's risk when it makes a loan (Gup and Kolari, 2005). The higher the collateral value the higher the incentive for the borrowers to repay the loan since they do not want to loose their collateral. The collateral value could also be a proxy for the borrowers' financial wealth since it is significantly positive correlated with the borrowers' income (Dinh and Kleimeier, 2007).

Loan duration indicates the maturity of loans in months. Loan duration reflects the borrowers' intention, risk aversion, or self-assessment of repayment ability.

Time with the bank indicates the borrowers' length banking relationship in years. It can be assumed that the longer a borrower stays with the bank, the more the bank knows about this borrower, and it could lower the probability of default. But this variable should be updated regularly due to adverse and unexpected changes in the borrowers' situation.

Number of loans measures the total number of loans a borrower has received from the bank during the whole relationship with the bank. Today, most borrowers have more than one loan from the same bank. This variable reflects the difficulty for a defaulted borrower to receive further loans from the same bank.

Current account indicates whether the borrower holds a current account with the bank. It partly represents the borrowers' financial wealth, and relationship between the borrower and the bank. The borrowers who hold current accounts with their banks have a lower default risk.

However, Boyes et al. (1989) recognised that if banks were minimising default risk, one should find the above variables with positive (negative) effect on the probability of granting a loan and a negative (positive) effect on default risk.

2.5 Credit scoring and modelling techniques

There are several statistical methods used to estimate credit scoring models in assessing borrowers' credits, such as discriminate analysis (Dunn and Frey, 1976), linear probability models (Turvey, 1991), probit models (Lufbuttow et al., 1984) and logit models (Mortensen et al., 1988). The last three methods estimate the default rate based on the historical data on loan performances and the borrowers' characteristics. The idea of linear probability is to look up for a linear combination of explanatory variables. It assumes there is a linear relationship between the default rate and the factors. The probit model assumes the probability of default follows the standard cumulative normal distribution function. The probability of default is logistically distributed in the logit model and discriminant analysis divides borrowers into high and low default-risk classes (Mester, 1997).

Discriminant analysis presents the critical assessment of the use of discriminant analysis in business. However, Hand et al. (1996b) show that the discriminant function obtained by segmenting a multivariate normal distribution into two classes' optimal discriminant function. Problems also arise in testing for the significance of individual variables when the assumption of normality does not hold and therefore we can not perform statistical inferences (Rosenberg and Gleit, 1994).

The linear probability model could present reasonable prediction results compared to discriminant analysis and logit models (Collins and Green, 1982). However, Pyndick and Rubinfeld (1998), Greene (1997), and Judge et al. (1985) indicate that the linear probability model could predict the default rate, but the predictive value might not necessary lie between zero and one. Moreover, because the variance of the models are generally heteroscedasticity, it leads to inconsistent estimation problem and invalid conventional measure of fit such as the R^2.

According to Hand and Henley (1997), the logistic approach is a more appropriate statistical tool than linear regression, when there are two discrete classes (good and bad risks) defined in the model. This gives the logistic approach superior classification rate. The probit model is very similar to the logit model. The logit model is generally preferred to the probit model because of its simplicity (Barney et al., 1999; Novak and LaDue, 1999; Lee and Jung, 1999) (see Table 1.3)

The logistic modelling approach is commonly used to model the bank's lending decision (Clarke, 2005). According to Collins and Green (1982), the logit model can increase the overall classification rate, and substantially reduce the error rate. The logistic approach also gives superior classification compare to dicriminant analysis (Wiginto, 1980). According to the literature, there is no best method for estimating credit scoring models and new methods continue to evolve. However, the logit models and neural networks have been applied frequently in previous research.

Table 1.3 Credit Scoring Techniques

	DA	LPM.	Logit	Probit	RPA.	ANN.
Dunn and Frey (1976)	✓					
Lufburrow et al. (1984)				✓		
Mortensen et al. (1988)			✓			
Miller and LaDue (1989)			✓			
Turvey and Brown (1990)			✓			
Jensen (1992)						✓
Altman et al. (1994)	✓					✓
Novak and LaDue (1999)			✓		✓	
Barney et al. (1999)		✓	✓			✓
Wu and Wang (2000)	✓					✓
Jacobson and Roszbach (2003)				✓		

Note: DA. = Discriminant Analysis, LPM. = Linear Probability Model, Logit. = Logistic Model, Probit. = Probit Model, RPA. = Recursive Partitionning Algorithms, ANN. = Artificial Neural Networks.

Clarke (2005), Buist et al. (1999), and Horne (1997) show the logistic model fitted well in their estimate of loan default risk in the mortgage industry. The logit model involves two choices represented by a binary, zero to one, dependent variable Y_i. The discussion begins with the limited case where loans are either defaulted or not (Pinder, 1996). Collins and Green (1982) find the logit model could increase the overall classification rate, and substantially reduce the error rate.

2.6 Credit availability model

Demand for credit is a derived demand. Households desire credit in order to make certain production and consumption expenditure as well as investments (Feder et al., 1993). The costs and returns of credit must be considered to determine the optimal level of credit utilization.

Bard et al. (2000) argue that there are several financial factors that may affect lender's decision on the amount of loan lend, such as, financial market structure and borrower, loan and lender characteristics. The borrower's characteristics indicates the credit risk, thus, it affect the loan amount. Bank attributes, such as lending policy, lending limits, reserve requirements, and available of funds are supply-side factors that could affect the availability of credit. Therefore, the credit availability model could be presented as follow:

$$A_i = f(B_i) \tag{2.1}$$

where A_j is the loan amount for loan ;

B_i is the vector of borrower characteristics believed to influence the loan amount.

2.7 Interest rate and loan pricing model

Banks' interest rate affects the potential borrowers' lending decision and the actions of borrowers. The interest rate charged by the bank could determine not only by the demand for capital but also the riskiness of the borrowers. The higher interest rate either presents riskier applicants or influences borrowers to choose other lenders. Lenders may optimally choose to ration the quantity of loans they have granted rather than raise the rate to clear the market (Petersen and Rajan, 1994).

In general, the interest rate charged on loan is comprised of four components (Ruthenberg and Landskroner, 2008):

Firstly, the financial funding cost.
Secondly, a premium reflecting market power exercised by the bank (example Inflation).
Thirdly, the sensitivity of the cost of capital raised to changes in loans extended.
Fourthly, a risk premium to compensate for the risk of default by borrower.

The default risk premiums explain the interest rates differences across the identical loan. Stiglitz and Weiss (1981) argues that the rate is charged to the borrower determines not only on the demand and capital but also the riskiness of the borrowers. A higher interest rate draws riskier applicants, and an increase in the interest rate increases the average riskiness of borrowers. Consequently, the differences in the interest rates are indicted the financial risk of the borrowers.

Lenders need to set the loan rate at the reasonable level consistent with the competition in the financial marketplace. Lenders want to charge a high enough rate to ensure that each loan will be profitable and compensate for the risks the lenders are exposed to. However, the rate of the loan should be low enough, in such a way that the borrowers can not be driven away to another lender. Therefore, in the loan market lenders are price takers, not price setters (Rose, 1993).

However, the competition between lenders might not be strong enough to eliminate the differences in interest rates across the lenders. It is more likely that lenders (for example, governmental banks and commercial banks) do not normally compete with each other for the same type of loans. Different lenders appear to have different lending practices and different cost structures; this can be another important factor explaining the difference in the interest rates charged (Limsombunchai et al., 2005).

Loans in general have the contract term extended over several years, and the expectation of future financial market conditions play an important role in shaping the demand and supply of loan funds. Interest rate can be viewed as the loan price from the interaction of the demand and supply for loanable funds. When the individual loan is considered at a point in time, the factors related to the expectations of the future economy and the future financial market conditions are expected to be homogenous across the competing lenders and borrowers. Therefore, the differences of the loan prices would be determined by the characteristics of the borrowers, the characteristics of the individual loans, and the likelihood that the loan will be repaid or defaulted.

In pricing a loan, the interest rate charged by a lender must cover the cost of loanable funds, the operating cost of the bank (such as, administration and service costs), the default risk premium and the profit for the lender (Lee et al., 1988). A loan pricing model can be expressed as follow (Rose, 1993):

Loan interest = rate	Marginal cost of rising loanable funds to lend to the borrower	+	Nonfunds bank operating costs	+	Estimated margin to protect the bank against default risk	+	Bank's desired profit margin

The loan pricing model can be formally expressed in general from as follows (Bard et al., 2000):

$$R_i = f(B_i) \tag{2.2}$$

Where R_i is the interest rate for loan i;

B_i is a vector of borrower and loan characteristics that may influence credit risk

Chapter 3

RESEARCH METHDOLOGY

This chapter discusses the research methodology and data. The research questions are presented in Section 3.1. The sample data and segmentation of variables are presented in Sections 3.2 and 3.2.1. The research models are described in Section 3.3. Sections 3.4 and 3.4.1 discuss the estimation techniques and the model sensitivity analysis. Section 3.5 concludes the chapter.

3.1 Research question

The objective of this research is to examine the effects of the borrower's characteristics in mortgage loan default, volume of loan granted and loan price. The credit scoring model is applied to investigate the probability of mortgage default in China during the period 2004 to 2009.

The whole sample mortgage loan are divided into different categories, such as borrowers' characteristics (age, gender, annual income, education level, occupation type, marital status, and region), loan amount, loan duration (shorter or longer than 5 years), relationship with the bank ("good" if the borrower does not delay loan repayment or "ok" if the borrower delays loan repayment), loan status (default loan or non-default loan), and loan price (the interest range charged by banks, less than or equal to 3 years, 3 years to 5 years, and greater than or equal to 5 years lending period).

Research Question One examines the Chinese mortgage loan default. According to Barney et al. (1999) and Wu and Wang (2000) studies, there are several statistical methods used to estimate credit scoring models. These include discriminant analysis, linear probability models, logit models, and probit models. The last three methods are standard statistical techniques for estimating the probability of loan default base on the borrowers' characteristics and historical data of loan performance.

The logistic approach is the most widely employed statistical procedure used to model the bank's lending decision (Clarke, 2005). Collins and Green's (1982) study find that the logit model can increase the overall classification rate, and substantially reduce the error rate. According to Mester (1997), Turvey (1991), Mortensen et al. (1988), Lufburrow et al. (1984), and Dunn and Frey's (1976) studies, logit model provides asymptotically consistent, efficient and unbiased estimates. Thus, logit model is used in this study to estimate bank lending decision via the maximum likelihood estimation technique.

The logit model assumes the borrowers' probability of default is logistically distributed. The model examines the parameters as dummy variables with multiple categories. The dummy regression model converts the variables of quantitative terms to numerical value of 0 and 1. In the bank lending decision model, the dependent and independent variables are dummy variables. For example, for dependent variable; the bad loan equals to 1 and good loan equals to 0. In this study, the result provides statistically significant estimates for all coefficients of parameters, so it can precede the interpretation of loan default and the borrowers' characteristics (Dimitrios Asteriou, 2006).

Research Question Two examines whether the borrowers' characteristics affect the volume of loan granted. Loan volume can be defined as the number and dollar amounts of new loans made during a period. The loan volume has been closely monitored by bank management, because it reflects both the market place and the bank's lending strategy. Bard et al. (2000) argue that the borrowers' characteristics could influence the lenders' decision on the loan amount. The borrower characteristics that signal credit risk may affect the loan amount.

Research Question Three examines how borrowers' characteristics influence loan pricing. Pervious studies investigate the interest rate charged by the bank could determine not only by the demand for capital but also the riskiness of the borrowers. Banks are required to set the loan rate at the reasonable level to be consistent with the competition in financial marketplace. The rate is charged high enough to ensure the lenders are profitable and compensate the risks lenders involved. However, the rate also needs to be low enough to ensure the borrowers will not be driven away and can repay the loan (Stiglitz and Weiss, 1981; Rose, 1993; Petersen and Rajan, 1994). Thus, it is important for lenders to understand how the borrowers' characteristics could influence the probability of mortgage loan default.

The ordinary least squares (OLS) method is utilised in both credit availability and loan pricing models. The regression analysis could direct the causation of which variable is causing/affecting the others. The result defines the correlation between dependent and independent variables whether the relationship is statistically significant under certain degree of freedom (Thomas, 2005).

3.2 Sample data

The research sample period is from 2004-2009. The data are obtained from one of the branches of the Construction Bank of China. The Construction Bank of China is one of the "big four" state owned banks in China. To date, it is ranked as the nation's second largest state owned bank in China. The Construction Bank of China is the leader in Chinese mortgage lending, which accounts for about 70 percent of all mortgages issued in China (Construction Bank of China). In this research, the name of the branch bank cannot be disclosed due to the agreement with the manager of the branch bank that provided the data for this research. Thus, only the parent's bank name can be announced.

The bank provided the personal number of each mortgage loan applicants, the date on which the application was submitted, the status of each loan (good or bad); the amount of the loan that was granted; the interest rate range of each loan is charged, and the date on which the loans were obtained from the data file. However, we are not allowed to identify the branches that are included in the sample and are not allowed to use the data set for any other research.

The data set included loans granted in 2004 to 2009. The total number of observations from the available data set was 7998 samples; the whole data set is from mortgage lending. The bank classifies loans with more than 90 days delay payment delay as default loans (bad loans). In the data set, there are 7681 good loans and 317 bad loans; the default rate is about 4.13%.

Banks' credit bureau recorded the personal information available on each applicant at the time the bank accesses the loan application, the unique personal number approved to each borrower. Before handing over the data for analysis, the personal numbers were removed, and the data for this research included borrowers' personal information, such as gender, age, education level, annual income, marital status, loan duration, occupation, region, and banking rating for each borrower ("good" versus "ok", "good"- if the borrower does not delay loan repayment, "ok"- if the borrower delays loan repayment).

3.2.1 Segmentation of variables

The variables used in this research overlaps with Crook (1992) as well as Dinh and Kleimeier (2007) list of commonly used variables for credit scoring model. Variables that are common to developing countries include gender, loan duration, marital status, and education. There are nine variables used in this research to assess the probability of mortgage loan default, loan availability and loan pricing. The variables are annual income, age, gender, marital status, education status, occupation, loan duration, region and bank rating. This is consistent with

Dinh and Kleimeier's (2007), Jacobson and Roszbach (2003), Coval et al. (2000), Thomas (2000), Crook, Hamilton, and Thomas (1992)'s studies.

According to Dinh and Leimeier's (2007) study, income can be used as a proxy of the borrower's financial wealth and ability to repay a loan. In this research, the annual income is categorized into two groups: annual income less than or equal to 36,000 RMB and annual income greater than 36,000 RMB.

Age measures the borrower's age in years. In this research, the age variable is categorized into three groups: 22-40 years old, 41-59 years old, and older than 59 year old.

Marital status affects the borrowers' responsibility, reliability, or maturity. According to Dinh and Kleimeier's (2007) study, the marital status is typically related to the number of dependants of the borrower which reflects financial pressure on the borrower and his/her ability to repay the loan. Marital status in this study is classified into a dummy variable, where, married equal to one, and zero otherwise.

Schreiner (2003) report there is ample evidence that woman default less frequently on loans. However, Schreiner (2004) discover gender effect disappears when other risk factors correlate with gender. However, we included gender in our model to test whether gender influence the probability of Chinese mortgage loan default.

Regarding to education level, the better educated people is expected to have more stable, higher-income employment and thus a lower default rate. There are three categories of education level; university or higher, college graduated, and high school graduated or education level lower than high school.

Occupation is highly correlated with income. Base on the eleven different occupations obtained from the sample, four categories are formed that pool occupations with similar default rates. The occupations are classified into group one as manager, group two as general staff, group three as professional employees and group four as others which include the military, farmer, teacher, consultant/broker, doctor, lawyer and journalist.

Loan duration measures the maturity of loans in months. In China, the loan duration is proposed by the lenders, this variable reflects the borrower's intention, risk aversion, or self-assessment of repayment ability. Based on the sample, there are two groups of borrower's loan duration: short-term loan (less than or equal to 5 years) and long-term loan (longer than 5 years).

Region represents the area where the borrower lives. As people of similar wealth tend to live in the same location, this geographic criterion can indicate a borrower's level of financial wealth. Region of the borrower is divided into two groups; within the district (in the same district where the bank is located) and out of the district (not in the same district where the bank is located).

The bank's rating relates to the borrower's previous loan performance as well the relationship with the bank. In this study, the bank rating of the individual borrowers are categorized as good and ok ("good" represents the borrower does not delay loan repayment, and "ok" represents the borrower delays loan repayment). These banking rates relate to the issues of credit availability and the interest rate charged to the borrowers, especially to small or individual borrowers. Previous studies documented the borrower with a close relationship with the banks is more likely to have a lower cost of capital, lower required collateral, and greater availability of funds compared to a borrower without such relationship (Petersen and Rajan, 1994; Berger and Udell , 2002).

Finally, because the sample contains only information of applicants who were accepted, the research can not perceive how these applicants who were rejected would have performed if they had been accepted. In the research, the variables have been categorized into different groups, one group of the variable may not be shown in the results, because it used as the benchmark for other groups. For example, there are four groups of occupation: manager, general staff, professional employee and others, these are dummy variables and one dummy variable is dropped to avoid the dummy trap problem in the model.

3.3 Research models

The objectives of this research is to use the credit scoring model to examine the co-integration of the borrowers' characteristics and banks loan granting decision, credit availability and loan pricing. According to the literature discussed in Chapter 2, mortgage loan default (credit scoring), credit availability, and loan pricing models are functions of the borrower's characteristics and a set of dummy variables (see Equation 3.1, 3.2, and 3.3).

$$\text{Mortgage loan default} = f\,(\text{Borrower characteristics, Dummy variables}) \qquad (3.1)$$

Credit availability = f (Borrower characteristics, Dummy variables) (3.2)

Loan Price = f (Borrower characteristics, Dummy variables) (3.3)

Where *Dependent variables* are:

- Mortgage loan default = 1 if loan is default (bad loan or not creditworthiness); 0 if loan is paid (good loan or creditworthiness),
- Credit availability = Volume of loan granted (in RMB),
- Loan price = Interest rate range charged, excluding application fee, borrowing fee, and other fees;

Borrower characteristics (Dummy variables) include:

AGE (+/-) = Dummy variables for age group

Age group 1; 1 if the applicant age is between 22 to 40 years old; 0 otherwise

Age group 2; 1 if the applicant age is between 41 to 59 years old; 0 otherwise

Age group 3; 1 if the applicant age is equal or above 60 years old; 0 otherwise

GENDER (+/-) = Dummy variable for gender; 1 if the applicant is female; 0 otherwise

MARITAL STATUS (+/-) = Dummy variable for marital status
1 if the applicant is married; 0 otherwise

EDUCATION (+/-) = Dummy variables for education levels

Educational level 1; 1 if the applicant completed a bachelor or higher degree; 0 otherwise

Educational level 2; 1 if the applicant completed three years of college; 0 otherwise

Educational level 3; 1 if the applicant completed primary school, middle school or high school; 0 otherwise

ANNUAL INCOME (+/-) = Dummy variable for annual income level

1 if the applicant annual income level is 36,000 RMB or higher; 0 otherwise

BANK RATING (+/-) = Dummy variable for bank rating

1 if the applicant bank rating is good (borrower did not delay loan repayment); 0 if applicant bank rating is ok (borrower delays loan repayment)

OCCUPATION (+/-) = Dummy variables for occupational status

Occupational status 1; 1 if the applicant is manager; 0 otherwise

Occupational status 2; 1 if the applicant is general staff; 0 otherwise

Occupational status 3; 1 if the applicant is a professional employee; 0 otherwise

Occupational status 4; 1 if the applicant is either a military, farmer, teacher, consultant/broker, doctor, lawyer or journalist; 0 otherwise

DURATION (+/-) = Dummy variable for duration status

1 if the applicant borrowed long term loan (greater than or equal to 5 years); 0 otherwise

REGION (+/-) = Dummy variable for region status

1 if the applicant lives within the bank district; 0 otherwise

The (+/-) indicate the hypothesised sign of the variables on mortgage loan default, credit availability, and loan pricing models. For example, annual income (+/-) is negatively related to the probability of bad loan, positively related to loan amount, but negatively related to the loan price.

3.4 Estimation techniques

The mortgage loan default (credit scoring) model will be analysed using logistic regression. According to Campbell and Dietrich (1983), the study estimate the loan default using data complied from the records of the bank. Horne (1997) employ the logistic regression to model bank's lending decision. The study finds the applicant's characteristics, such as age and marital status influence the decision of approval or denial the mortgage lending.

The mortgage loan default model is given as follows (Gujarati, 1995):

$$P_i = E(Y_i = 1 \mid X_{ij}) = \frac{1}{1+e^{-Z_i}} \frac{1}{1+e^{-(\alpha+\sum_j \beta_j X_{ij}+\varepsilon_i)}} \tag{3.4}$$

Where Y_i equals to 1 if loan is paid (good loan); 0 if loan is defaulted (bad loan);

P_i is the estimated probability of a good loan (high value of P_i implies low default risk);

$$Z_i = \alpha + \sum_j \beta_j X_{ij} + \varepsilon_i$$

α and β_j are an intercept term and parameters, respectively.

X_{ij} are Borrower characteristic, Credit risk proxies, Relationship indicators and Dummy variables;

ε_i is the error term,

Equation 3.4 represents the cumulative logistic distribution function. If P_i is the probability of a good loan, then, the probability of a bad loan or (1- P_i) given as follows:

$$(1-P_i) = \frac{1}{1+e^{Z_i}} \tag{3.5}$$

Therefore, the odds ratio in favour of a good loan or $\frac{P_i}{(1-P_i)}$ can be written as follows:

$$\frac{P_i}{(1-P_i)} = \frac{1+e^{Z_i}}{1+e^{-Z_i}} = e^{Z_i} \tag{3.6}$$

Taking the natural log on Equation 3.5 becomes:

$$Z_i = \ln(\frac{P_i}{1-P_i}) = \alpha + \sum_j \beta_j X_{ij} + \varepsilon_i \tag{3.7}$$

Where Z_i is the natural logarithm of the odds ratio in favour of a good loan.

The model is a binary choice model and the use of the ordinary least square estimation technique is inappropriate (Maddala, 1983). Thus, to obtain efficient parameter estimates, the maximum likelihood estimation is applied to the logistic regression. The likelihood function L for the model is given as follows (Maddala, 2001):

$$L = \prod_{Y_i=1} P_i \prod_{Y_i=0} (1 - P_i) \tag{3.8}$$

From Equation 3.7, the probability of a good loan can be obtained by the following equation (Greene, 1997):

$$P_i = \Pr ob(Y_i = 1 \mid X_{ij}) = \frac{e^{Z_i}}{1 + e^{Z_i}} \tag{3.9}$$

The credit availability and loan pricing models are given as follow (see Equation 3.10)

$$Y_i = \alpha + \sum_j \beta_j X_{ij} + \varepsilon_i \tag{3.10}$$

Where Y_i is interest charged;

α and β_j are an intercept term and the parameters, respectively;

X_{ij} are Borrower characteristics and Dummy variables;

ε_i is the error term.

To estimate the credit availability and loan pricing models, the ordinary least squares (OLS) method is utilised.

3.4.1 Sensitive analysis

The Maximum Likelihood Estimates (MLE) is the empirical estimation of logit model; it assumes large sample properties of efficiency, consistency, normality of parameter estimates and validity of the t-test significance (Studenmund, 2001; Greene, 1993). Beside these properties, the logit model could help to avoid the major problem associated with Ordinary Least Square (OLS) which estimate the standard linear probability model (Judge et al., 1982; Hair et al., 1998). The MLE coefficient estimates from the logit analysis have no direct interpretation with respect to the probability of the dependent variable (Loan default =1) other than indicating a direction of influence of probability. Liao (1994) and Maddala (1991) recommend calculating the changes in probabilities to indicate the magnitude of the marginal effect. It refers to the partial derivatives of the non-linear probability function at each variable's sample mean (Liao, 1994; Pindyck and Rubinfeld, 1991; Hosmer and Lemeshow, 1989). In order to identify the most and the least important variables influencing the mortgage loan default (good loan versus bad loan), the marginal effect for each of the estimated coefficients in the empirical model are calculated. The marginal effect reveals the marginal change in the dependent variable given a unit change in a selected independent variable, holding other variables constant (Liao, 1994). The marginal effect indicates the level of important for the estimated coefficients in the empirical model.

3.5 Conclusion

This chapter discusses the research questions, methodology, and data. Section two described the data source used in this study. Section three presents the research models. The empirical models and statistical test methods for the research questions are discussed in Chapter four. This includes a discussion of the logit, multiple regression test methods and the sensitive analysis of marginal effect test.

Chapter 4

EMPIRICAL FINDINGS AND RESULTS DISCUSIION

This chapter reports the results on the mortgage loan default, credit availability and loan pricing models, and the factors affecting the lending decision models, amount of credits and loan prices. The estimated results reflect the borrower’s risk level, the amount of credit availability and the lending rate ranges.

The chapter is organised as follows: Section 4.1 presents the results of the sample selection. Section 4.2 describes the characteristics of the sample data including the descriptive statistics. Following this, Section 4.3 discusses the results of mortgage loan default models, including the marginal effects. Sections 4.4 and 4.5 discuss the estimated results on credit availability and loan pricing models, respectively. Section 4.6 summarises the research findings.

4.1 Sample selection results

The total sample from the mortgage loans of Construction Bank of China was 7,998. The loan data set includes 7,681 good loans and 317 bad loans (see Table 4.1). The average default rate for the total loans is 4.127%. The data set contained no missing observations. The samples include the borrowers’ demographic characteristics (such as age, gender, marital status, education, occupation, region), bank relationship indicators (such as loan amount, loan duration, bank rating, interest rate range), borrowers’ financial detail (such as annual income), and debt repayment history (such as the number of days that the loan repayment has been delayed). In this research, all loans are mortgage loans for individual borrowers.

Table 4.1 Total Number of Mortgage Loans (Good and Bad Loans)

Year	Good loans	Bad loans	Default rate (in %)
2004	706	37	5.24
2005	2604	92	3.53
2006	2061	92	4.46
2007	1505	55	3.65
2008	749	38	5.07
2009(First quarter)	56	3	5.35
Total	7681	317	4.13

Table 4.1 presents the distribution of good and bad loans by lenders' identity from 2004 to the first quarter of 2009. Table 4.1 also shows the total amount of good loans issued is significantly larger than the total amount of bad loans issued. The number of mortgage loan increased significantly since 2005 to 2007. For example, the number of good loans increased from 706 to 2,604 from the year 2004 to 2005. This could be explained by China's membership to the WTO, which has a serious effect on the Chinese mortgage loan market. Following China's accession to the WTO, foreign lenders were allowed to participate in Chinese banking markets. Both domestic and foreign banks compete for the housing loans in China's housing market. For example, at the end of October 2007, the outstanding residential housing loans reached 4.69 trillion RMB, a growth of 30.75%, up by 1.01 trillion RMB from the beginning of the year. This amounted 28.9% of the total new RMB loans of the commercial banks during the same period. The outstanding individual housing loan reached 2.6 trillion RMB, 619.2 billion RMB more than the beginning of 2007 (Liu, 2007). Furthermore, the Chinese mortgage lending was heavily affected by the government monetary policies. In 2005, the personal mortgage loan researched 1.9 trillion RMB, an increased of 12.8% compared to 2004. At the end 2006, the personal mortgage loan increased 19%, up by 3.3% compared to the same time period in 2005. But in 2007, the amount of personal mortgage loan decreased compared to 2006. This is because of the government regulation and the tight monetary policy on mortgage lending. In January-December 2008, the personal mortgage loans amounted to RMB 357.3 billion, down from 29.7% compared to the pervious year (People's Bank of China, 2009). Table 4.1 also shows the default rate has increased from 3.53 percent to 5.07 percent from 2005 to 2008. The results show during this time period the lending amount has increased dramatically, but because of the lack of proper techniques in assessing credit risk, the default rate also increased. The results confirmed that a well-developed credit scoring system is necessary for Chinese banks. This is consistent with Li and Zhang's (2003) findings.

4.2 Descriptive Statistics

Table 4.2 Panel A: Descriptive statistics (sample from 2004 to 2006)

	Age	*Annual income (in RMB)*	*Loan duration*	*Loan amount*	*Days of overdue*
Mean	41.095	25619.132	9.558	84995.787	5.978
Standard Error	0.151	429.671	0.085	1030.827	0.507
Standard Deviation	8.846	25204.523	4.990	60477.1167	29.768
Minimum	21.919	8000	1	10000	0
Maximum	65.988	240000	20	490000	937

Panel B: Descriptive statistics (sample from 2007 to the first quarter of 2009)

	Age	*Annual income (in RMB)*	*Loan duration*	*Loan amount*	*Days of overdue*
Mean	38.282	28589.341	10.234	87675.393	6.337
Standard Error	0.138	622.678	0.068	970.741	0.372
Standard Deviation	9.315	42029.60	4.553	65523.196	25.111
Minimum	22.958	7900	1	14000	0
Maximum	65.988	483600	20	437061	445

* 1 US Dollar ≈6.82 RMB

Table 4.2 shows the descriptive statistics of the borrowers' age, annual income, loan duration, loan amount, and the loan performance. Panel A in Table 4.2 presents the descriptive statistics for the sample from 2004 to 2006; and Panel B presents the descriptive statistics for the sample from 2007 to the first quarter of year 2009.

Table 4.2 also reports the borrowers' mean annual income, loan duration, loan amount and days of overdue increase between Panel A and Panel B. In Panel B, the mean of borrowers' age is slightly lower than those in Panel A, which indicates there are more young aged people applying for mortgage loans. The borrowers' annual income increased from *25,619 RMB* to *28,589RMB*; the loan duration increased from 9.6 years to 10.2 years; the average loan amount also increased from *84,995.79 RMB* to *87,675.39 RMB*; and the days of over due increased from 5.97 to 6.34 between the two sub samples. According to Thomas (2000) and Boyle et al.'s (1992) study, older borrowers are more risk adverse, and less likely to default. Loan duration measures the maturity of loans in months, which reflects the borrowers' intention of the borrowing period, risk aversion, or self-assessment of repayment ability, and the longer the loan duration, the higher the borrowers' default risk. According to Jacobson and Roszbach (2003), the decrease in borrowers' age or increase in loan duration both increases the probability of loan default. There is a positive relationship between the borrowers' income and probability of default; higher income is associated with lower default rate.

In Panel A, the standard deviation for the borrowers' annual income (*25204.5)* and loan amount (*84995.7)* are extremely high. This indicates that the borrowers' annual income and loan amount varies widely among the sample mean. The standard deviations for the borrowers' age, the loan duration, and the days of overdue are *8.85, 4.99* and *29.77* respectively. In Panel B, the standard deviation for the borrowers' annual income and loan amount are similar to those in Panel A. The average borrowers' age, annual income, loan duration, loan amount and the days of overdue are higher in the Panel B than in the Panel A. The results show the borrowers' age decreased while the annual income, loan duration, loan amount and the days of overdue increased between the time period 2004 to 2006 and 2007 to the first quarter of 2009. This can be explained by the economic growth in China in the late 1990s where peoples' annual income increase, but the prices of new houses in China increase much faster. The wage increase for the majority of urban households has not kept up with the rise in housing costs, which leads to an increase demand for housing loans (Burell, 2006). As banks try to capture a larger share of the housing market, some commercial banks lower their lending criteria, reduce loan application procedures and relax the borrowers' background investigation through distorted practices (Liu, 2007). This could explain the increase of the loan default rate during this time period.

Table 4.3 Characteristics of borrowers and loans (2004-2009)

Variable	Good loans	Bad loans	Total
Loan repayment	7681	317	7998
Volume of total loan granted	688.9669 million (RMB)	28.8625 million (RMB)	717.8294 million （RMB）
Borrower characteristics			
Gender			
_Female	2904	106	3010
_Male	4777	211	4988
Age			
22-40	3949	216	4165
≥ 40	3732	101	3833
Education			
< College	1432	102	1534
≥ College	6249	215	6464
Annual income			
＜36000 (RMB)	7027	296	7323
≥ 36000 (RMB)	654	21	675
Marital status			
Married	4788	221	5009
Un-married	2893	96	2989
Loan duration			
Short term loan (≤ 5yrs)	2121	91	2212
Long term loan （> 5 yrs)	5560	226	5786
Occupation			
Manager	1863	76	1939
Stuff	4448	186	4634
Professional	947	31	978
Other	423	24	447
Region			
Local (borrower living within the district)	7669	315	7984
Un-local（borrower living outside the district）	12	2	14
Bank rating			
Good (borrower who did not delay loan repayment)	4225	4	4229
Ok (borrower who delayed loan repayment)	552	313	865

Table 4.3 shows the characteristics of the borrowers classified by loan performance as good or bad loans. A total of 7,998 new loans were issued from 2004 to 2009, where 7,681 (96.04 percent) are considered as good loans and 317 (3.96 percent) are bad loans. According to Dinh and Kleimeier (2007), bad loans are classify as loans with more than 90 days of delay payment and can be treated as loan default. Following Dinh and Kleimeier's classification, a total of 317 bad loans are considered as default loans. The data in Table 4.3 shows the default rate on average is higher among male borrowers than female borrowers; among age group between 22-40 years old and among low education level groups.

The correlation matrix in Appendix Ⅰ shows the loan amount granted has a positive correlation with *Interest, Age (1), Age (2), Education (1), Annual income, Region, Occupation type (2), and Bank rating*. However, it is negatively correlated with *Age (3), Education level (3), Occupation type (1) and Occupation type (4)*. The low and insignificant correlation coefficients between the loan amount granted and *Gender* (r = -0.01), *Marital status* (r = 0.00), *Education level* (2) (r = - 0.01), *Occupation type (3)* (r = 0.00) and between the loan amount granted and the *Loan duration* (r =0.00) imply that these variables do not impact the amount of credit granted.

The correlation coefficients show *Age group (2), Education level (1), Annual income, Occupation type (3)* and *Bank rating* are negatively correlated with the loan default. On the other hand, *Marital status, Education level (2), Region, Occupation type (1)* and *Duration* are positively correlated with the loan default (see Appendix I). The correlation coefficient does not control for the other factors' influence, and further investigation should be conducted to examine these relationships.

4.3 Mortgage loan default model

The empirical estimation of the logit model used is Maximum Likelihood Estimates (MLE), which assumes large sample properties of consistency, efficiency, normality of parameter estimates and validity of the t-test significance (Studenmund, 2001; Greene, 1993). Given these properties, the logit model avoids the major problem associated with Ordinary Least Square (OLS) estimation of the standard linear probability model (Judge, Hill, Griffiths, Lutkepohl, and Lee, 1982). The MLE coefficient estimates from the logit analysis have no direct interpretation with respect to the probability of the dependent variable (Y=1) other than indicating a direction of influence of probability. Liao (1994) and Maddala (1991) recommend calculating changes in probabilities to indicate the magnitude of the marginal effect. This refers to the partial derivatives of the non-linear probability function evaluated at each variable’s sample mean (Liao, 1994; Pindyck and Rubinfeld, 1991). As a result, in order to identify the most and the least important variables influencing the banks’ lending decision, the marginal effects for each of the estimated coefficients in the empirical model were calculated. The marginal effect reveals the marginal change in the dependent variable given a unit change in a selected independent variable, holding other variables constant (Liao, 1994). The marginal effect indicates the level of importance for the estimated coefficients in the empirical model.

Table 4.4 Mortgage loan default model (2004 to 2009)

Number of Observations:	7997
Log likelihood function:	-1106.248
Restricted log likelihood:	-1349.780
Chi-Squared Statistics:	487.0636
Degrees of Freedom:	18
Prob[ChiSqd > value]:	0.000000
McFadden R^2:	0.18042

	Coefficients	Std Error	t-statistics	Marginal Effects
Loan Amount	3.105012473	0.63260217	4.908**	0.4751713197E-01
Interest rate(1)	0.5909588811	0.22974721	2.572**	0.4712655758E-02
Interest rate(2)	0.5067970889	0.18492724	2.741**	0.3755480046E-02
Gender	-0.7999524478E-01	0.11901032	-0.672	0.5037217488E-03
Age (1)	-1.650097968	0.26398226	-6.251**	-0.1261567876E-01
Age (2)	-1.876146791	0.24730046	-7.587**	-0.1172089233E-01
Marital status	0.1044613454	0.12258393	0.852	0.6432009566E-03
Education (1)	-0.3059537648	0.21417965	-1.428	-0.1730050790E-02
Education (2)	-0.3565166173E-03	0.16682475	-0.002	0.2223101219E-05
Annual income	-0.4028412951	0.17497164	-2.302**	-0.2955789748E-02
Bank rating	-0.5972558099	0.20872121	-2.862**	-0.4928165859E-02
Occupation (1)	-0.4950888665	0.21926146	-2.258**	-0.2750534552E-02
Occupation (2)	-0.6053185713	0.20566173	-2.943**	-0.4015860309E-02
Occupation (3)	-0.8036251380	0.27012376	-2.975**	-0.3809004189E-02
Loan duration	0.1380944467	0.13092320	1.055	0.8358586063E-03
Region	-5.227391706	0.55996579	-9.335**	-0.9053778240E-01

Note: * *denote statistically significant at 0.05 level of significance

The mortgage loan default models (include the full sample model and two sub-sample models) are estimated using the logistic regression via the maximum likelihood estimation techniques. The estimated results of the full sample mortgage loan default model are presented in the Table 4.4. The table shows eleven out of sixteen predicted influencing factors are statistically significant (Chi-Square= 487.0636, P value= 0.000, degree of freedom = 18). The estimated coefficients in Table 4.4 are significantly different from zero at the 5 percent level of significance.

The negative (or positive) coefficient values in Table 4.4 imply decreases (or increases) in the probability of a loan default relative to the borrower's characteristics. For example, *Loan amount, Interest range (1)* and *Interest range (2)* positively impact the probability of loan default. An increase in the loan amount or interest rate level increases the probability of loan default (Petersen and Rajan, 1994; Stiglitz and Weiss, 1981).

Similarly, the coefficient *Age group (1)* (22 to 40 years old), *Age group (2)* (41 to 59 years old), *Annual income, Bank rating, Occupation type (1)* (manager), *Occupation type (2)* (general staff), *Occupation type (3)* (professional employee) and *Region* negatively impact the probability of loan default. This implies the demographic variables are significant in explaining the borrowers' probability of default. For example, the Age group between 22 to 59 years is less likely to default the loan, as the majority of borrowers in this age group are employed, have a steady income, and have the ability to repay their loans. As expected, the higher the respondents' income the lower the probability of loan defaults. Income is commonly use as a proxy of the borrower's financial wealth. For example, the probability of loan defaults of a borrower who earns an annual income higher than *36,000 RMB* is lower than borrowers whose annual incomes are less than *36,000 RMB.* The significant and negative coefficients for Occupation type (1) (manager), (2) (general staff), and (3) (professional employee) suggest that borrowers that are employed are less likely to default their loans. The results are consistent with Jacobson and Roszbach's (2003) findings who argue that there is a negative relationship between income and the borrowers' default rate; higher income would appear to be associated with lower default risk. Thomas (2000) and Boyle et al. (1992) confirm that older borrowers are more risk adverse, and therefore are less likely to default. Furthermore, banks are more hesitant to lend to younger borrowers who are more risk averse. Dinh and Kleimer (2007) discover that occupation is commonly incorporated into a credit scoring model since it is highly correlated with income. However, Gender, Marital status, Education level (1) and (2) ((1) = bachelor or higher degree, (2) = college graduate) and Loan duration are statistically insignificant in explaining the banks lending decision in this study.

Table 4.5 Mortgage loan default model (sub samples model with two different time periods)
Panel A: Mortgage loan default model (2004 to 2006)

Number of Observations:	5595
Log likelihood function:	-722.4542
Restricted log likelihood:	-940.2830
Chi-Squared Statistics:	435.6575
Degrees of Freedom:	15
Prob[ChiSqd > value]:	0.000000
McFadden R^2:	0. 23166

	Coefficients	Std Error	t-statistics	Marginal Effects
Loan Amount	0.3247914589E-02	0.11618745E-02	2.795**	0.1758102664E-04
Interest rate(1)	1.811986168	0.29353797	6.173**	0.2277236848E-01
Interest rate(2)	1.505638373	0.24583051	6.125**	0.1441588596E-01
Gender	-0.4376850643E-01	0.14546516	-0.301	0.2382276411E-03
Age (1)	-1.923722108	0.31674529	-6.073**	-0.1422129515E-01
Age (2)	-2.195525276	0.29351574	-7.480**	-0.1149939075E-01
Marital status	-0.6450714823E-02	0.15009772	-0.043	-0.3494832272E-04
Education (1)	-0.3527133039	0.26906577	-1.311	-0.1696130596E-02
Education (2)	-0.1008168646	0.20698375	-0.487	-0.5541667429E-03
Annual income	-1.825351673	0.19669382	-9.280**	0.1532525561E-01
Bank rating	-0.7473880542	0.23962219	-3.119**	-0.5755212538E-02
Occupation(1)	-1.507709939	0.30570540	-4.932**	-0.5319890456E-02
Occupation(2)	-1.227101956	0.24960676	-4.916**	-0.8705693840E-02
Occupation(3)	-1.458532266	0.31758507	-4.593**	-0.5061000020E-02
Loan duration	0.1868483820	0.16521067	1.131	0.9608605911E-03
Region	-4.932877165	0.59085852	-8.349**	-0.6956058859E-01

Note: * *denote statistically significant at 0.05 level of significance

Panel B: Mortgage loan default model (2007 to 2009)

Number of Observations:	2403			
Log likelihood function:	-323.1416			
Restricted log likelihood:	-409.5255			
Chi-Squared Statistics:	172.7677			
Degrees of Freedom:	15			
Prob[ChiSqd > value]:	0.000000			
McFadden R^2:	0. 21094			
	Coefficients	Std Error	t-statistics	Marginal Effects
Loan Amount	0.1579846766E-02	0.23619264E--02	0.669	0.9667455472E-05
Interest rate(1)	0.48995532310	0.45353809	1.080	0.3650561278E-02
Interest rate(2)	0.8159748171	0.30624971	2.664**	0.6654401297E-02
Gender	-0.4828111667E-01	0.22249812	-0.217	-0.2939009868E-03
Age (1)	-1.819914563	0.48756930	-3.733**	-0.1253355758E-01
Age (2)	-1.843899819	0.45563371	-4.047**	-0.1271038646E-01
Marital status	-0.1253452502	0.22814849	-0.549	0.7578151029E-03
Education (1)	-0.5348950401	0.38105172	-1.404	-0.2833061904E-02
Education (2)	-0.6259355024E-01	0.30575976	-0.205	-0.3867139286E-03
Annual income .	-1.240288089	0.25159530	-4.930**	0.1159873951E-01
Bank rating	-1.048138946	0.37698950	-2.780**	-0.1077622366E-01
Occupation(1)	-0.2014353287E-01	0.33844851	-0.060	-0.1231574193E-03
Occupation(2)	-0.2903971631	0.36317753	-0.800	-0.1712211786E-02
Occupation(3)	-0.7895297292	0.55802590	-1.415	-0.3613772208E-02
Loan duration	0.5521074946E-01	0.22392601	0.247	0.3364161128E-03
Region	-4.971146481	1.0082568	-4.930**	-0.7600265674E-01
Note: * *denote statistically significant at 0.05 level of significance				

Table 4.1 shows that the lending amount dropped significantly from 2006 to 2007; and the descriptive statistics shows the mean and standard deviations of *Age*, *Annual income*, *Loan duration*, *Loan amount*, *Days of overdue* are different in the different time period (see Table 4.2). According to Mays (2001) and Thomas et al. (2001) , if the sample's internal differences is large and the sample size is big enough, in order to improve the model prediction accuracy, the sample should be divided into sub-sample to test the credit scoring model. Table 4.5 shows the sub samples of the mortgage loan default model in the two different time periods; the first model includes time period from 2004 to 2006 and the second model from 2007 to the first quarter of 2009. The estimated coefficients in Table 4.5 are statistically significant at the 5 percent level. Panel A shows eleven out of sixteen predicted influencing factors are statistically significant (Chi-Square =435.6575, P value= 0.000, degree of freedom= 18). Panel B shows six out of sixteen predicted influencing factors are statistically significant (Chi-Square = 172.7677, P value= 0.000, degree of freedom= 15)

The data in Panel A in Table 4.5 shows the coefficients *Loan amount*, *Interest rate (1)*, *Interest rate (2)* are positively correlated with loan default; *Age (1)* (22 to 40 years old), *Age (2)* (41 to 59 years old), *Annual income*, *Bank rating*, *Occupation (1)* (manager), *Occupation (2)* (general staff), *Occupation (3)* (professional employee), and *Region* are negatively correlated with the probability of loan default. The results are similar to the full sample model. For example, annual income is commonly used as a proxy of the borrower's financial wealth, and the result shows the probability of loan default of the borrower who earns an annual income higher than *36,000 RMB* is lower than the borrower whose annual income is less than *36,000 RMB*.

Panel B shows the time period 2007 to the first quarter of year 2009. The results show *Interest rate (2)* is positively correlated with loan default; and *Age(1)*, *Age(2) Annual income*, *Bank rating* and *Region* are negatively correlated with the probability of loan default.

The results in the Table 4.5 reflect the changing economic conditions during the two time periods. For example, China became a member of WTO in 2006 and foreign lenders are allowed to participate in the Chinese banking markets. The competition between domestic and foreign banks could potentially intensify the underlying risks in the Chinese housing markets. As banks try to capture a larger share of the housing loan markets, some commercial banks lowered their lending criteria; relaxed the borrower's background investigation through distorted practice (Liu, 2007). Therefore, the loan amount increased dramatically from 2006 to 2007 and this could lead to an increase of the probability to default by the borrowers.

4.3.1 Interpretation of the marginal effects

Table 4.6 shows the marginal effects of the full sample mortgage loan default model. The marginal effects show changes in the conditional probability results from the changes in the independent variables. For example, a unit change in the increment in the borrower's total loan amount value would increase the probability of a bad loan by 4.75%. Furthermore, for borrowers in the 22 to 40 years old age group, the probability of loan default decreases by 1.17%, and for borrowers in the 41 to 59 years old age group, the probability of loan default decreases by 1.26%.

Table 4.6 Marginal effects of the full size mortgage loan default model

Variables	**Marginal effects**	**Importance Ranking**
Loan Amount	0.0475	2
Interest rate (1)	0.471	6
Interest rate (2)	0.376	7
Age (1)	-0.126	3
Age(2)	-0.117	4
Annual income	-0.296	10
Bank rating	-0.493	5
Occupation (1)	-0.275	11
Occupation (2)	-0.402	8
Occupation (3)	-0.381	9
Region	-0.905	1

Types of occupation also impact the probability of loan default. For example, if the borrower is a manager (*Occupation (1)*), the probability of loan default decreases by 0.275%; if the borrower is a general staff (*Occupation (2)*) the probability of loan default decreases by 0.402%; and if the borrower is a professional employee (Occupation 3), the probability of loan default decreases by 0.381%. If bank rating for the borrower is good, the probability of the loan default decreases by 0.493%; and if the region of the borrower is local, the probability of the loan default decrease by 9.05%.

4.3.2 Important factors to banks in lending decisions

According to the marginal effect variables (see Table 4.6), when a bank makes a lending decision, the bank should rank the risk factor from the most important to the least important factor in making the lending decision. The marginal effects show how strongly these independent variables influence the mortgage loan default.

From Table 4.6, we may conclude that the lenders rank the borrower's region as the most important factor when making their lending decision. This implies that the borrowers' region may affect the mortgage loan default. *Loan amount* and the age group between 22 to 40 years old are the second and third important variables impacting mortgage loan default, respectively. The age group between 41 to 59 years old is the fourth most important variable impacting the lenders' decision. The borrower's bank rating is the fifth most important factor impacting the lending decision. Interest range (1) and (2) are the sixth and seventh factors influencing the mortgage loan default. Professional employee followed by general staff is the ninth important factors influencing the mortgage loan default. The borrower's annual income greater than *36,000* RMB is the tenth most important factor affecting mortgage loan default. However, *Occupation (1)* (managers) is the least important variable influencing the mortgage loan default.

In summary, the results from the logistic regression models reveal the factors that determine mortgage loan default include demographic characteristics (*Age, Gender, Marital status, Education, Occupation and Region*), bank relationship indicators (*Loan amount, Loan duration, Bank rating* and *Interest rate range*), and the borrower's financial detail (*Annual income*). The results suggest that an increase in the loan amount or interest rate increases the probability of loan default. In addition the increase in *Age group (1), (2)* ((1) =22 to 40 years old, (2) =41 to 59 year old), *Annual income*, *Bank rating*, *Occupation type (1) (manager), (2) (general staff), and (3)* (professional employee) decreases the loan default rate. The results are consistent with Dinh and Kleimeier (2007) and Jacobson and Roszbach (2003) findings.

4.4 Credit availability model

The estimated results of the ordinary least square (OLS) regressions for the credit availability model is shown in Table 4.7. The results show that all explanatory variables can explain 55 percent of the total variation in the credit availability. The results indicate that the model fits the data quite well.

Table 4.7 Effects of Borrowers' Characteristics on Bank Loan Availability

Dependent variable:	Loan amount			
Method:	Ordinary Least Squared			
Number of Observations:	7997			
Log likelihood function:	-5134.1599			
Restricted log likelihood:	-5235.7876			
R-Squared Statistics:	0.55096			
	Coefficients	Std Error	t-statistics	P-value statistics
Gender	0.8743946899E-02	0.10564658E-01	0.828	0.4079
Age(1)	0.1626502149	0.26187620E-01	6.211**	0.0000**
Age(2)	0.1584651285	0.25181129E-01	6.293**	0.0000**
Marital status	0.8417023603E-03	0.10686562E-01	0.079	0.9372
Education (1)	0.3989511067E-01	0.18074910E-01	2.207**	0.0273**
Education (2)	0.7859162553E-02	0.14559287E-01	0.540	0.5893
Annual income	0.733735306E-02	0.17761595E-01	0.413	0.6795
Bank rating	0.2211610804E-01	0.20098137E-01	1.100	0.2712
Occupation(1)	0.7850838277E-01	0.23012222E-01	3.412**	0.0006**
Occupation(2)	0.6613715583E-01	0.21841237E-01	3.028**	0.0025**
Occupation(3)	0.7672546138E-01	0.25572066E-01	3.000**	0.0027**
Loan duration	-0.7063415883E-02	0.11784720E-01	-0.599	0.5489
Region	0.1527864145	0.10266333E-01	14.882**	0.0000**
Note: * *denote statistically significant at 0.05 level of significance				

Table 4.7 shows the estimated coefficients *Age group (2)* (41 to 59 years old), *Education level (1)* (bachelor or higher degree), *Occupation type (1)* (manager), *Occupation type (2)* (general staff), *Occupation type (3)* (professional employee), and *Region* are highly significant and positively correlated with banks credit availability. For example, the coefficient Age group between 41 to 59 years old showed that senior borrowers are more likely to grant a larger

amount of loan. Senior borrowers are more risk adverse and will therefore be less likely to default. Thomas (2000) and Boyle et al. (1992) confirm this empirically, where banks are more hesitant to lend to younger borrowers who are more risky and tend to default their loans repayment. Borrowers with a university or post graduate degree are more likely to obtain bigger loans, because these borrowers have a better capability to repay the loan. Dinh and Kleimeier (2007) argue that better educated borrowers have a more stable employment and higher income, thus have a lower default rate. Banks prefer to retain these customers with a large amount loan. The positive coefficient of *Occupation types (1) (manager), (2) (general staff), and (3) (professional employee)* show that borrowers who are employed, and worked as a manager, a general staff, or a professional employee should have a low earning volatility.

The above results are consistent with Bard et al.'s (2000) findings who argue that the borrowers' characteristics affect the loan amount. Similarly, Limsombunchai et al.'s (2005) study show that the borrower with a low earning volatility, a high asset value, high average earnings, high education, high work experience, and high profitability would be granted a larger loan.

4.5 Loan pricing model

Multiple linear regression is used to estimate the loan pricing model. The results of the regression model are presented in Table 4.8. The volume of credit is not included in the loan pricing model, since it is assumed to be an endogenous variable.

Table 4.8 Effects of Borrowers' Characteristics on Loan Pricing model

Dependent variable: Interest rate rang
Method: Ordinary Least Squared
Number of Observations: 7997
Log likelihood function: -8466.6207
Restricted log likelihood: -8048.4137
R-Squared Statistics: 0.110256

	Coefficients	Std Error	t-statistics	P-value statistics
Gender	-0.4928939889E-01	0.16047701E-01	-3.071**	0.0021**
Age(1)	-1.620136738	0.46059903E-01	-35.175**	0.0000**
Age(2)	-1.639682776	0.43873026E-01	-37.373**	0.0000**
Marital status	-0.5512506384E-01	0.16271656E-01	-3.388**	0.0007**
Education (1)	-0.3386618553E-01	0.27553649E-01	-1.229	0.2190
Education (2)	0.2112611130E-01	0.22265456E-01	0.949	0.3427
Annual income	-0.1495664916	0.26978619E-01	-5.544**	0.0000**
Bank rating	-0.4511061610	0.34091991E-01	-13.232**	0.0000**
Occupation(1)	-0.3294403796	0.35365670E-01	-9.315**	0.0000**
Occupation(2)	-0.3522479831	0.33640920E-01	-10.471**	0.0000**
Occupation(3)	-0.3734204863	0.39189367E-01	-9.529**	0.0000**
Loan duration	0.9633629149E-01	0.17896327E-01	5.383**	0.0000**
Region	-0.1645126270	0.15583306E-01	-10.557**	0.0000**

Note: * *denote statistically significant at 0.05 level of significance

The estimated coefficient *Loan duration* is positively correlated with the loan price and statistically significant at the 5% level of significance (see Table 4.8). This shows that the borrower who maintains a longer relationship with the bank would have a lower borrowing cost. The result is consistent with the findings of Berger and Udell (1990), Berger and Udell (1995), and Bodenhorn (2003).

The estimated coefficients *Gender, Age (1) (22 to 40 years old), Age (2) 41 to 59 years old), Marital status, Annual income, Bank rating, Occupation (1) (manager), Occupation (2) (general staff), Occupation (3) (professional employee)* and *Region* are significant and negatively influence the loan pricing model. The mortgage loan default model showed that the factors except *Gender* are negatively correlated to the probability to default (see Table 4.4). Therefore, there is a positive relationship between the loan price and credit risk, which indicates the higher interest rate attract riskier applicants. For example, applicants with low annual income have a higher probability to default, thus, the loan price is relatively high for the low income applicants. This is consistent with Stiglitz and Weiss (1981) and Petersen and Rajan's (1994) findings.

Education has no significant impact on the loan price. It could be due to when duration is introduced into the model, the influence of some of the borrowers' characteristics could be eliminated by the length of the bank-borrower relationship. This result is consistent with Limsonbunchai et al.'s (2005) study.

The results in Table 4.8 confirm that the borrowers' demographic characteristics (such as gender, age, marital status, annual income and region) negatively impact the bank's mortgage loan price. Furthermore, the longer the relationship the borrower has with the bank, the lower the borrowing cost they will have to pay. In addition, a higher bank rating leads to a lower loan price; the borrower who does not delay his or her loan repayment will be rated as "good" borrower. Bodenhorn (2003), Athavale and Edmister (1999), Blackwell and Winters (1997), and Berger and Udell (1995) find evidence of an inverse relationship between the bank-borrower relationship and loan price, which shows that the good relationship between the bank and borrower could result in a lower loan rate.

4.6 Summary of findings

In this study, we developed and analysed three models (Mortgage loan default model, Credit availability model, Loan pricing model) of the Chinese residential mortgage markets. The mortgage loan default model was divided into two sub-sample; the first sub-sample model is from year 2004 to 2007, the second sub-sample model is from year 2007 (see Table 4.5). From our results, we can conclude that the borrower's demographic characteristics (age, gender, marital status, education, occupation and region), bank's relationship indicators (loan amount, loan duration, bank rating and interest charged) and borrower's financial detail (annual income) significantly affect banks' lending decision, loan amount granted and loan price charged.

For the mortgage loan default model, we divide the full sample into two sub- samples. Our result shows the factors affect the mortgage loan default model varies in different time period since the economic condition changes during the two time periods. The empirical result confirms that banks in screening and monitoring borrowers require reviewing the borrower's credit worthiness periodically.

For credit availability, the borrower's with the lower default risk tends to get a bigger amount of bank loan. This confirms the finding of Limsombunchai et al. (2005), which suggest borrowers with lower earning volatility, high asset value, high average earnings, higher education, and longer work experience could be granted a larger loan.

Moreover, our results suggest that the loan price is positively correlated with the credit risk, where a higher interest rate draws riskier applicants, and an increase in the interest rate increases the average riskiness of the borrowers. This result is consistent with Stiglitz and Weiss's (1981) study.

Chapter 5

SUMMARY AND CONCLUSIONS

This chapter summarises the research. Section 5.1 presents a summary of the research objectives, data and methodology, and major findings. The implications of the research findings are discussed in Section 5.2. Section 5.3 describes the research limitations and Section 5.4 provides recommendations for further research.

5.1 Summary and major findings

Credit scoring is broadly applied in consumer lending, especially in credit cards, and has been used in mortgage lending recently. Mester's (1997) study discovered that credit scoring is a cost effective credit management tool for banks to evaluate loan applications. Based on Berger and Frame's (2007) and Turvey and Brown (1990) studies, credit scoring uses the borrower's financial history and financial statement to evaluate the credit risk of loan applications. It is a statistical approach to predict the probability whether a credit applicant will default or become delinquent.

In China, the rigorous risk management systems have been implemented and practised by most banks, but financial fraud happens often, and the bad loans of state owned commercial banks' portfolio are huge. Since China becomes a member of the WTO, the degree of Chinese economy opening would be higher, consumer credit activity will further extend, and consumer credit problem will increase. For example, some local bank officers of Chinese commercial banks lower the lending criteria, reduce the examination steps and relax investigation through distorted practice to increase market share of the credit market. A few branches and sub-branches of Chinese commercial banks even collaborate with real estate developers to provide housing mortgage loan contracts and these loans are used to help developers to speculate on difficult-to-sell buildings (Liu, 2007). Therefore, it is important for Chinese banks to practise a robust consumer credit scoring system, providing reasonable information to decision makers to guarantee the financial order and stability in the Chinese financial market.

This research is expected to contribute to the development of Chinese mortgage market in additional to analysing the rational behaviour of mortgage loan default processes. The purpose of this research is to identify critical factors in the mortgage loan default process and to investigate factors affecting the amount of credit granted and interest rate charged for Chinese mortgage lending.

The data used in this research are obtained from a branch of the Construction bank of China, a major lender in Chinese mortgage sector. During the period of 2004 to 2009, a total of 7998 credit files under the mortgage loan scheme were obtained. The credit scoring model is analysed using the logistic regression with the maximum likelihood estimation.

5.2 Results and implications

5.2.1 Results for research objective one and implications

Objective one of this research is to identify the factors in mortgage lending decision process of Chinese banks. Our results of mortgage loan default models show that the borrower's demographic characteristics impact banks' lending decision process. The empirical results suggest the *Loan amount* and *Interest range* are positively correlated with the probability of loan default; which means an increase in the loan amount or interest rate level increases the probability of loan default. There are seven variables negatively impact the probability of loan default, such as age group between 22 to 59 years old, annual income greater than 36,000 RMB, good bank rating (no delayed loan repayment history), manager (occupation type (1)), general staff (occupation type (2)), professional employee (occupation type (3)), borrower resides within the same district with the bank. For example, if the borrower's annual income level is higher than 36,000 RMB, the probability of loan default for the borrower will be lower. The results are consistent with Dinh and Kleimer (2007) Jacobson and Roszbach (2003), Thomas (2000), Petersen and Rajan (1994), Boyle et al. (1992), and Stiglitz and Weiss (1981).

The full sample period of this research is from the year 2004 to the first quarter of 2009, and the banks lending amount has dropped dramatically from 2006 to 2007. Based on Thomas et al. (2001) and Mays's (2001) studies, further tests should be performed on the sample's internal differences problem. Therefore, the full sample of the mortgage loan default model is divided into the different time periods (2004 to 2006, and 2007 to the first quarter of 2009).

Our results indicate during the different lending period and as the economic condition changes, there are different numbers of factors which impact the mortgage loan default process. The sub sample (2004 to 2006) results are similar to the full size sample where loan amount and interest range are positively correlated with the probability of loan default. Demographic characteristics of age group between 22 to 59 years old, annual income greater than 36,000 RMB, good bank rating, occupation such as manager, general staff, and professional employee and local borrower negatively impact the probability of loan default. But in the sub sample from 2007 to the first quarter of 2009, there are only six factors which impact the mortgage loan default model, higher range of interest charged is positively correlated to the loan default; age group between 22 to 40 and 41 to 59 years old, annual income, good bank rating and local borrower are negatively correlated to the probability of loan default.

The results in this research show that the borrower's demographic characteristics could affect the mortgage loan default process. A good credit scoring model has the ability to detect bad loans; this could help the bank to reduce the loan losses from loan default. Consequently, it can improve the profitability and the financial stability of the bank. Therefore, the credit scoring model should be developed and used to support credit officers in monitoring the Chinese mortgage loan applications. The results also show it is necessary to review the borrowers' creditworthiness periodically, as the changes in economic condition could affect the loan performance; this is consistent with McAllister and Mingo's (1994) finding.

5.2.2 Results for research objective two and implications

Research objective two is to investigate the factors affecting the volume of credit granted by Chinese banks. Our results suggest that age group between 41 to 59 years old, university or post graduated, occupation types such as manager, general staff, and professional employee, and local borrower are highly significant and positively correlated to banks' credit availability. The results show that the borrower with a low earning volatility, a high asset value, high average earnings, high education, high work experience, and high profitability would be granted a larger loan, which is consistent with Boyle et al. (1992), Bard et al. (2000), Thomas (2000), Limsombunchai et al. (2005), and Dinh and Kleimeier's (2007) findings.

The results in this research show that the borrower's demographic characteristics that signal credit risk may affect the loan amount. A good model could help the bank to determine the optimal level of credit extended to the borrower. As a result, this could limit the borrowers to borrow from more expensive or non-financial institution lenders (such as loan sharks).

5.2.3 Results for research objective three and implications

Multiple linear regressions are used to estimate the loan pricing model. Our results show that *Loan duration* is positively correlated with the loan price and statistically significant at the 5% level. There are nine variables significantly and negatively correlated with the loan pricing model, such as female, age group between 22 to 59 years old, married, annual income is greater than 36,000 RMB, good bank rating (no delayed loan repayment history), manager (occupation (1)), general stuff (occupation (2)), professional employee (occupation (3)), borrower resides within the district with the bank. For example, according to Dinh and Kleimeier's (2007) study, the default rate for female is less than male. Thus, the loan charge for female will be lower than male.

The results in this research show that the borrower's demographic characteristics could affect the loan price. The results suggest that borrowers could consider swapping banks less frequently to keep a good repayment history and maintain a better relationship with the bank; this could help to lower the borrowing cost. For example, the longer and better the relationship the borrower has with the bank, the lower the borrowing cost, which is consistent with Berger an Udell (1995), Blackwell and Winters (1997), Athaval and Edmister (1999), and Bodenhorn's (2003) findings.

5.3 Research limitations

There are a number of limitations related to the data set, the estimation techniques, and the variables used in this research. These include:

Firstly, the data set used in this research is from one branch of Construction bank of China. Therefore, the results in this research have limited implications and robustness. The results can not represent the lending behaviour of all commercial banks in the mortgage lending in China.

Secondly, only the data and information of the applicants who have been granted a credit in the past are used in the study. There is no data and information on the applicants who were rejected. Thus, the models are parameterized using a sample of accepted applicants only. This may lead to biased estimates of the parameters. For example, equation (3.1) is used to examine the effect of the various borrowers' characteristics on whether loans are defaulted or not, but not what criteria the bank used to make their loans decision. This is due to the limitation of the data. There is no data and information on the applicants whose loans were rejected. Thus the model is parameterized using a sample of accepted applicants only. Therefore, equation (3.1) examines only the effect of the various borrowers' characteristics on whether loans are defaulted or not after the lending decision has been made.

Thirdly, the model ignores the potential exposure to the future credit risk. The credit scoring, credit availability and loan pricing models are typically static models in nature and they ignore the potential exposure to future credit risk In addition, the information about financial distress of the borrowers' in the past is not included in the models. Thus, the credit scoring model and judgmental technique can be combined to improve the probability of mortgage loan default, which can help the bank to make better lending decisions in terms of assessing the credit quality ex-ante based on the borrowers' characteristics and loan evaluators' personal experience.

Table 5.1- Impact of borrowers' characteristics on the mortgage loan default, loan availability and loan pricing models

Mortgage loan default model(Full Sample)	Bank loan availability model	Loan pricing model
Loan amount (+)		
Interest rate(1) (+)		
Interest rate(2) (+)		
Age(1) (-)	Age(1) (+)	Age(1) (-)
Age(2) (-)	Age(2) (+)	Age(2) (-)
Annual income (-)		Annual income (-)
Bank rating (-)		Bank rating (-)
Occupation(1) (-)	Occupation(1) (+)	Occupation(1) (-)
Occupation(2) (-)	Occupation(2) (+)	Occupation(2) (-)
Occupation(3) (-)	Occupation(3) (+)	Occupation(3) (-)
Region (-)	Region (+)	Region (-)
	Education(1) (+)	Loan duration (+)

Fourthly, the research findings identify some important borrowers' characteristics which impact both the mortgage loan default and the availability and pricing of loans. Table 5.1 shows the influence of the borrowers' characteristics with the three models are different. For example, annual income and bank rating impact the mortgage loan default and loan pricing, but not bank loan availability. However, the parameters which affect the borrowers' probability to default should also affect the loan amount and loan pricing. Table 5.1 shows there is a positive relationship between *Loan amount* and *Mortgage loan default.* Petersen and Rajan (1994) and Stiglitz and Weiss (1981) revealed that an increase in the loan amount or interest rate level increases the probability of loan default. This inconsistency in our results could be due to the credit scoring model which is not wildly used in Chinese banks and the credit assessment and management tool of Chinese banks are outdated. Most Chinese banks depend on the experience and common knowledge of credit evaluators in assessing credit risk. These evaluators can not ensure lenders are applying the same underwriting criteria to all borrowers (Mester, 1997). Therefore, the findings of this research suggest the credit scoring model should be developed to assist credit officers in processing Chinese mortgage loan applications.

5.4 Recommendations for future research

To improve the research results and to increase the generalisability of the research findings, the credit file from other commercial banks can be collected and included in the data set. Furthermore, the number of the sample of borrowers should be increased and the information of the applicants who did not qualify for the credit should be taken into account.

In addition, there are a number of variables that can be added to the models to enhance the performance of the models. These include geographic distance between the lending bank and the borrower (see Degryse and Ongena, 2005), the number of loans with the current bank (Dinh and Kleimeier, 2007), the borrower's credit bureau score (see Wu and Wang, 2000; Athavale and Edmister, 1999; Elsas and Krahnen, 1998), and the management ability of the borrower (see Angelini et al., 1998; Ellinger et al., 1992).

This study only provides a single estimation of the credit scoring model. The combination models could be used in the future study. For example, the study could combine the logit and artificial neural net works (ANN) together, where the calculation of the neural network weights is known as training process. The process starts by randomly initialising connection weights and introduces a set of data inputs and actual outputs to the network. The network then calculates the network output and compares the actual output, and then calculates the error. In an attempt to improve the overall predictive accuracy and to minimise the network total mean squared error, previous study discovered that neural network model offers a more flexible and efficient framework to analyse economic and financial relationships than can be potentially highly nonlinear. The neural network examines all information at once, thus facilitating a highly automated input interface (Wu and Wang, 2000).

In practice, banks need to update their credit scoring model on a regular basis. For developing countries where economic changes should more pronounced than the relatively stable developed countries. A regular update of the credit scoring model is even more important for banks in developing countries. Furthermore, it might be interesting to compare credit scoring models of banking markets that are in various stages of development. This way could explore any linkages that exist between the features that determine loan default, for example, the variables included in the credit scoring model and their relative importance, and financial and economic development (Dinh and Kleimeier, 2007).

Collateral requirement is considered as another key issue in the lending decision process. However, it has not been addressed in this research. Therefore, further research can be extended to include the determinants of collateral (what factors determine the extent of total debt collateralization) and the influence of the bank-borrower relationship on the collateral requirement. These could be achieved by using the multiple regression analysis and the artificial neural networks techniques.

REFRENCES

Abdou, H., A. E. Masry, and J. Pointon (2007), "On the applicability of credit scoring models in Egyptian banks", Banks and bank system, 2(1), Pg 4.

Allen N. Berger, Iftekhar Hasan and Mingming Zhou (2009), "Bank ownership and efficiency in China: What will happen in the world's largest nation?", *Journal of Banking and Finance*, 33(1), January, Pg 113-130.

Altman, E. I. (1968) "Financial ratios, discriminant analysis and the prediction of corporate bankruptcy", The Journal of Finance, 23(4), Pg 589-609.

Altman, E. I., M. Glancario, and F. Varetto (1994), "Corporate distress diagnosis: comparisons using linear discriminant analysis and neural networks (the Italian experience)", *Journal of Banking and Finance*, 18, Pg 505-529.

Angelini, P., R. Di Salvo, and G. Ferri (1998), "Availability and cosy if credit for small business: Customer relationships and credit cooperatives", *Journal of Banking and Finance*, 22(6-8), Pg 925-954.

Anonymous (2008), "China construction bank corporation-corporate and strategic assessment report", M2 Presswire, Oct 7.

Anonymous (2009), "China's personal mortgage hit CNY 295trn by Nov. end", Sinocast China Business Daily News, Jan. 7, London (UK).

Athavale, M., and R. O. Edmister (1999), "Borrowing relationships, monitoring, and the influence on loan rates", Journal of Financial research, 22(3), Pg 341-352.

Asteriou, D. (2006), Applied Econometrics, Palgrave Macmillan.

Avery, R. B., R. W. Bostic, P. S. Calem, and G. B. Canner (1996), "Credit risk, Credit scoring, and the Performance of Home Mortgages", *Federal Reserve Bulletin 82* (July 1996), Pg 621-48

Banasik, J., J. Crook, and L. Thomas (2003), “ Sample selection bias in credit scoring models”, *Journal of the Operational Research Society*, 54, Pg 822-832.

Bard, S. K., P. J. Barry, and P. N. Ellinger (2000), “Effects of commercial bank structure and other characteristics on agricultural lending”, *Agricultural Finance Review*, 60, Pg 17-31

Barney, D. K., O. F. Graves, and J. D. Johnson (1999), “The farmers home administration and farm debt failure prediction”, *Journal of Accounting and Public Policy*, 18, Pg 99-139.

Beck, R. E., and S. M. Siegel (2001), Consumer lending (4th ed.), American Bankers Association.

Berger, A. N. and G. F. Udell (1990). “Collateral, loan quality, and bank risk”, *Journal of Monetary Economics,* 25(1), Pg 21-42.

Berger, A. N. and G. F. Udell (1995). “Relationship lending and lines of credit in small firm finance”, *Journal of Business*, 68, Pg 351-381.

Berger, A. N., and G. F. Udell (2002), “Small business credit availability and relationship lending: the importance of bank organizational structure”, *Economic Journal*, 112, Pg 32-53

Berger, A. N., and W. S. Frame (2007), “Small business credit scoring and credit availability”, *Journal of Small Business Management*, Jan., 45, Pg 1.

Blackwell, D. W., and D. B. Winters (1997), “Banking relationships and the effect of monitoring on loan prices”, *Journal of Financial Research*, 20, Pg 275-289.

Bodenhorn, H. (2003), “Short-term loans and long-term relationships: Relationship lending in early America”, Journal of Money, Credit, and Banking, 35(4), Pg 485-505.

Boyes, W. J., D. L. Hoffman, and S. A. Low (1989), “ An econometric analysis of the bank credit scoring problem”, *Journal of Econometrics*, 40, Pg 3-14

Boyle, M., J. N. Crook, R. Hamilton, and L. C. Thomas (1992), Methods for credit scoring applied to slow payers. In; Thomas, L. C., Crook, J. N., Edelman, D. B. (Eds), "Credit scoring and credit control", *Oxford University Press*, Oxford, Pg 75-90.

Buist, H., P. D. Linneman, and I. F. Megbolugbe (1999), "Residential-Mortgage Lending Discrimination and Lender-Risk-Compensating Polices", *Real Estate Economics*, 27(4), Pg 695

Burell, M. (2001), "The rule-governed state: China's labor market policy, 1978-98. Doctoral dissertation, Department of Government, Uppsala University.

Burell, M. (2006), "China's housing provident fund: its success and limitations", *Hosing Finance International*, 20(3), Pg38

Capon, N. (1982), "Credit Scoring Systems: A Critical Analysis", *Journal of Marketing* (pre-1986); Spring 1982; 46, 00002; ABI/INFPRM Global, Pg 82

Campbell, T.S. and J. K. Dietrich (1983), "The Determinants of default on insured conventional residential mortgage loans", *Journal of Finance*, 38, Pg 1569-1581.

Chandler, G.C., and J.Y. Coffman (1979), "A comparative analysis of empirical vs. judgmental credit evaluation", *Journal of Retail Banking*, 1, Pg 15-26.

Chatterjee, S., D. Corbae, and J. V. Ríos-Rull (2008), "A finite-life private-information theory of unsecured consumer debt", *Journal of Economic Theory*, 142(1), Pg 149-177

Chen (2005), X. Chen, M. Skully, and K. Brown. "Banking efficiency in China: Application of DEA to pre-and post-deregulation eras: 1993-2000". *China Economic Review*, 16, Pg 229-245.

Clarke, J. A. (2005), "Implications of stratified sampling for fair lending binary logit models", *Journal of Real Estate Finance and Economics*, 30(1), Pg 5.

Collins, R. A., and Richard D. Green (1982), "Statistical Methods for Bankruptcy Forecasting", *Journal of Economics and Business,* 34, Pg 349-354.

Coval, J., T. Shumway (2000). “Do behavioural biases affect prices?” University of Michigan, Working Paper.

Crook, J. N. (1996), “Credit scoring: an overview”, Working paper no. 96/13, Department of business studies, The University of Edinburgh.

Crook J.N., R. Hamilton, and L.C. Thomas (1992), “A comparison of a credit scoring model with a credit performance model”, *The Service Industries Journal*, 12, 4

Degryse, H., and S. Ongena (2005), “Distance, lending relationships, and competition”, Journal of Finance, 60(1), Pg 231-266

Dinh, T. H. T., and S. Kleimeier (2007), “A credit scoring model for Vietnam’s retail banking market”, *International Review of Financial Analysis*, 16(5), Pg 471-495

Dunn, D. J., and T. L. Frey (1976), “Discriminant analysis of loans for cash grain farms”, *Agricultural Finance Review,* 36, Pg 60-66.

Ellinger, P. N., N. S. Splett, and P. J. Barry (1992), “Consistency of credit evaluations at agricultural banks”, *Agribusiness*, 8(6), Pg 517-536.

Elsas, R., and J. P. Krahnen (1998), “Is relationship lending special? Evidence from credit-file data in Germany”, Journal of Banking and Finance, 22(10-11), Pg 1283-1316.

Feder, G., L. J. Lau, J. Y. Lin, and X. Luo (1993), “The nascent rural credit market in China”, in Hoff, K. A. Braverman, and J. E. Stiglitz (Eds.), *The economics of rural organization: theory, practice, and policy*, Oxford University Press.

Frame, W. S., A. Srinivasan, and L. Woosley (2001), “The effect of credit scoring on small-business lending”, *Journal of Money*, Credit and Banking, 33(3), Pg 813-825.

Gup B., and J. Kolari (2005), Commercial Banking: The management of Risk, (3rd Eds), John Wiley & Sons.

Glassman, C. A., and H.M. Wilkins (1997), “Credit scoring: probabilities and pitfalls”, *Journal of Retail Banking Service*, 19(2), Pg 53-56

Greene, A. H. (1983). Using econometrics: a practical guide, (4^{th} Eds.), Addison Wesley Longman Inc.

Greene, A. H. (1993). Using econometrics: a practical guide, (4^{th} Eds.), Addison Wesley Longman Inc.

Greene, W. H. (1997), Econometric Analysis, (3^{rd} Eds.), Prentice Hall.

Gujarati, D. N. (1995), Basic Econometrics, (3^{rd} Eds.), McGraw-Hill.

Hair, J. F., Anderson, R. E. Anderson , R. L. Tatham, and W. C. Black (1998). Multivariate data analysis, (5^{th} Eds.), Upper Saddle River, N. J.: Prentice-Hall International Inc.

Hand, D. J., and W. E. Henley (1997), "Statistical classification methods in consumer credit scoring: a review", *Journal of the Royal Statistical Society: series A* (Statistic in Society) 160, Pg 525-541.

Hand, D. J., J. J. Oliver, and A. D. Lunn (1998), "Discriminant analysis when the classes arise from a continuum" Pattern Recognition, 31(5), Pg641-650.

Horne, D.K. (1997), "Mortgage lending, race and model specification", *Journal of Financial Services Research*, 11, Pg 43-68

Hosmer, D. W., and S. Lemeshow (1989). Applied logistic regression. New York: John Wiley and Sons.

Huang, Z. (2004), "Is WTO accessions an opportunity or challenge for Chinese state commercial banks?" MSc dissertation, University of Middlesex.

Jacobson, T., and K. Roszbatch (Apr., 2003), "Bank lending policy, credit scoring and value-at-risk". *Journal of Banking and Finance*, 27(4), Pg 615

Jesen, H. L. (1992), "Using neural networks for credit scoring", *Journal of Managerial Finance,* 18(6), Pg 15-16.

Jude, G. G., R. C. Hill, W. E. Griffith, H. Lutkepohl, and T. Lee (1982). Introduction to the theory and practice of econometrics. New York: Wiley.

Judge, G. G., W. E. Griffiths, RC. Hill, H. Luktepohl, and T. Lee (1985), *The Theory and Practice of Econometrics*, (2nd Eds.), John Wiley and Sons.

Roszbatch, K. (2004), "Bank lending policy, Credit Scoring and the Survival of Loans", *The review of economics and statistics*, 86(4), Pg 946

Lee, T. H., and S. C. Jung (1999), "Forecasting creditworthiness: logistic vs. artificial neural net", Journal of Business Forecasting, winter, Pg 28-30.

Lewis, E. M. (1992), An Introduction to Credit Scoring, (2nd Eds.), Fair, Isaac and Co.

Li, S. G., and M. Y. Zhang (2003), "A study of present condition and development foreground personal credit scoring in China", *USA-China Business Review*, 3(3) (serial No. 16)

Liao, T. F. (1994). Interpreting probability models: logit, probit and other generalized linear models. Sage university papers series. Quantitative applications in the social sciences; no. 07-101. Thousand Oaks, Calif: Sage Publications.

Limsombunchai, V., C. Gan, and M. S. Lee (2005), "An analysis of credit scoring for agricultural loans in Thailand", *American Journal of Applied Sciences,* 2(8), Pg 1198-1205.

Lin Xiaochi and Yi Zhang (2009), "Bank ownership reform and bank performance in China", *Journal of Banking and Finance*, Jan, 33(1), pp 20-29.

Liu, S. Y. (2007), "Earnestly strengthening commercial real estate management to adjust real estate credit structure and prevent credit risks", Paper presented at the Conference on Strengthening Commercial Real Estate Credit Management on December 11th.

Lufburrow, J., P. J. Barry, and B. L. Dixon (1984) "Credit scoring for farm loan pricing", *Agricultural Finance Review*, 44, Pg 8-14.

Ma, J. (1997), China's Economic Reform in the 1990s. *World Bank*, mimeo

Maddala, G. S. (1983), Limited-dependent and qualitative variables in econometric, Cambridge University Press.

Maddala, G. S. (2001), Introduction to econometrics. Macmillan Publishing Company.

Marron, D. (2007), "Lending by numbers'; credit scoring and the constitution of risk within American consumer credit", *Economy and Society*, 36(1), Pg 103-133.

Martinelli C. (1997), "Small firms, borrowing constraints, and reputation", *Journal of Economic Behaviour and Organization*, 33(1), Pg 91-105

Mays, E. (2001) Handbook of Credit Scoring. Fitzroy Dearborn.

Mester, L. J. (1997), "What's the point of credit scoring?" Business Review, Federal Reserve Bank of Philadelphia, Sep/Oct., Pg 3-16.

McAllister, P. H., and J. J. Mingo (May, 1994), "Commercial loan risk management, credit scoring and pricing: the need for a new shared database", *Journal of Commercial Lending*, Pg 6-22.

Miller, L. H., and E. L. LaDue (1989), "Credit scoring assessment models for farm borrowers: a logit analysis", *Agricultural Finance Review,* 49, Pg 22-36.

Mortensen, T., D. L. Watt, and F. L. Leistriz (1988), "Predicting probability of loan default", Agricultural Finance Review, 48, Pg 60-67.

Nicholas Lardy (1998), "China's Unfinished Economic Revolution" *Brookings Institution Press*, Washington, D.C.

Novak, M. P., and E. LaDue (1999), "Application of recursive partitioning to agricultural credit scoring", Journal of Agricultural and Applied Economics, 31(1): Pg 109-122.

OCC Bulletin 97-24, "Credit scoring models", May 20, 1997

Petersen, M. A., and R. G. Rajan (1994), "The benefit of lending relationships: evidence from small business data", *Journal of Finance*, 49, Pg 3-38

People's Daily (2004), "One hundred foreign banks are allow to do Renminbi (Chinese currency) business", 19 July, P1.

Pinder J.P. (1996), "Decision analysis using multinomial logit models: mortgage portfolio valuation", *Journal of Economics and Business*, 48: 67-77

Pindyck, R. S. and D. L. Rubinfeld (1991). Econometric models and economic forecasts, (3rd Eds.) New York: McGraw-Hill.

Plata, V., and G. N. Nartea (1998), "Credit analysis procedures of rural lenders in Canterbury", Paper presented at the Fifth Annual Conference of the New Zealand Agricultural and Resource Economics Society (Inc.)

Pyndick, R. S., and D. L. Rubinfeld (1998), Econometric Models and Economic Forecasts, (4th Eds.), McGraw Hill.

Rodman, J. (2006), "Viewpoint: Will China crack down on realty tricks?", *American Banker*, New York, Jan. 6, 171(4), Pg 10 .

Rose, P. S. (1993), Commercial Bank Management, (2nd Eds.), IRWIN.

Rosenberg, E., and A. Gleit (1994). "Quantitative methods in credit management: A survey", *Operations Research*, 42, Pg 589-613.

Ruthenberg D. and Y. Landskroner (2008), "Loan pricing under Basel II in an imperfectly competitive banking market", Journal of Banking & Finance, 32(12), Pg 2725-2733.

Sayuri Shirari (2002), "Bank's lending behaviour and Firm's corporate financing pattern in the People's Republic of China. *ADB Institute research paper*

Schreiner, M. (2000), "Credit scoring for microfinance: can it work?" Microfinance Risk Management, Washington University.

Schreiner, M. (2003). The next breakthrough in micro credit? CGAP Occasional Paper No. 7, http://www.cgap.org

Schreiner, M. (2004). Scoring arrears at a micro lender in Bolivia. *Journal of Microfinance,* 6, 65-88.

Shinkey, J. F. (2002), *Commercial Bank Financial Management*, (6th Eds.), Prentice Hall.

Stiglitz , J. E., and A. Weiss (1981). "Credit rationing in markets with imperfect competition", American Economic Review, 71, Pg 393-410.

Steenackers, A., and M. J. Goovaerts (1989). "A credit scoring model for personal loans", Insurance: Mathematics and Economics, 8, Pg 31-34.

Studenmund, W. H. (2001). Econometric analysis, (2nd Eds.) Macmillan Publishing Company.

Thomas, L. C. (2000). "A survey of credit and behavioural scoring: Forecasting financial risk of lending to consumers". *International Journal of Forecasting* 16, Pg 149-172

Thomas, R. L. (2005). Using Statistics in Economics, McGraw-Hill

Thomas, L. C., D. B. Edelman and L. N. Crook (2002), "Credit scoring and its applications", Philadelphia: Society for industrial and Applied Mathematics.

Thomas L. C., J. Banasik, and J. Crook (2001), "Recalibrating Scorecards. Journal of the Operational Research Society, 52, Pg 981-988.

Turvey, C. G. (1991), "Credit scoring for agricultural loans: a review with application", *Agricultural Finance Review*, 51, Pg 43-54

Turvey, C. G., and R. Brown (1990), "Credit scoring for a federal lending institution: the case of Canada's farm credit corporation", *Agricultural Finance Review*, 50, Pg 47-57.

Wang, J. (2004). "Sweet mortgage loan deals", *Beijing Review*, 47(34), pg 42.

Weitseng Chen (2006), "WTO: Time's Up for Chinese Banks- China's Banking Reform and Non-Performing Loan Disposal". *Chicago Journal of International Law*, Chicago: Summer, 7(1), Pg 239

Wiginton, J. C. (1980), "A note on the comparison of logit and discriminant models of consumer credit behaviour, *The Journal of Financial and Quantitative Analysis*, 15(3), Pg 757-770.

Wong, Y. C. Richard and M. L. Sonia Wong (2001), "Competition in China's Domestic Banking Industry", *Cato Journal*, 21(1), Spring/Summer

Wu, F. L. (1996), "Changes in the structure of public housing provision in urban China", *Urban Studies*, 33(9), Pg 1601-1627.

Wu, C., and X. M. Wang (2000), "A neural network approach for analyzing small business lending decisions", *Review of Quantitative Finance and Accounting*, 15(3), Pg 259-276.

Yi, G. (1994), "Money, banking, and financial markets in China". *West view Press*, San Francisco.

APPENDICIES

Appendix: Correlation Matrix

t-Statistic	LA	I	D	G	A1	A2	A3	M	E1	E2	E3	AI	R	O1	O2	O3	O4	LD	BR
LA	1.00																		
I	0.41**	1.00																	
D	0.23	0.00	1.00																
G	-0.01	-0.00	0.01	1.00															
A1	0.03**	-0.00	-0.01	-0.02	1.00														
A2	0.03**	-0.03**	-0.02**	0.02**	-0.97**	1.00													
A3	-0.02**	-0.02**	0.00	0.01	-0.16**	-0.04**	1.00												
M	0.00	-0.02**	0.03**	-0.02**	0.08**	-0.08**	-0.01	1.00											
E1	0.02**	0.00	-0.01**	0.01	0.03**	-0.03**	-0.01**	0.04**	1.00										
E2	0.01	0.01	0.02**	-0.03**	0.25**	-0.23**	-0.13**	0.04**	-0.60**	1.00									
E3	-0.01**	-0.01**	-0.01	0.02**	-0.34**	0.31**	0.18**	-0.08**	-0.21**	-0.65**	1.00								
AI	0.03**	-0.03**	-0.01**	0.00	0.04**	-0.04**	0.00	0.02**	0.04**	0.03**	-0.07**	1.00							
R	0.02**	0.01	0.01**	0.01	0.16**	-0.16**	0.00	0.05**	0.03**	0.11**	-0.16**	0.03**	1.00						
O1	-0.02**	0.01	0.01**	0.00	0.05**	-0.04**	-0.04**	0.03**	0.02**	0.04**	-0.07**	0.29**	0.04**	1.00					
O2	0.02**	-0.02**	0.00	-0.01	-0.10**	0.09	0.03**	-0.02**	0.00	-0.04**	0.05**	-0.26**	-0.03**	-0.67**	1.00				
O3	0.00	-0.02**	-0.02**	0.00	0.06**	-0.05**	-0.03**	-0.04**	-0.03**	0.03**	-0.01	-0.07**	0.00	-0.21**	-0.44**	1.00			
O4	-0.02**	0.00	0.00	0.02**	0.03**	-0.03**	0.04**	0.03**	0.00	-0.02**	0.02**	0.14**	0.00	-0.14**	-0.28**	-0.09**	1.00		
LD	0.00	0.26**	0.01**	-0.01**	-0.09**	0.08**	0.05**	-0.15	0.04**	-0.10**	0.09**	-0.20	-0.06**	-0.11**	0.12**	0.03**	-0.12	1.00	
BR	0.11**	-0.13**	-0.21**	0.01	0.02**	-0.01**	0.00	-0.02**	0.00	0.00	0.01	-0.02**	-0.01	0.00	-0.01	0.00	0.01	0.02	1.00