INTERNATIONAL HELLENIC UNIVERSITY

Credit Scoring and Bank Lending Policy in Consumer Loans

A dissertation

submitted in partial fulfillment

of the requirements for the Degree

of

Master in Banking and Finance

at

International Hellenic University

By Ganopoulou Maria and Giapoutzi Fotini

2010

ABSTRACT

Abstract of a dissertation submitted in partial fulfilment of the
requirements for the Degree of MSc in Banking and Finance.

Credit Scoring and Bank Lending Policy in Consumer Loans

By Maria Ganopoulou and Fotini Giapoutzi

Consumer loan performance determines the profitability and stability of banks and other lending institutions and screening the loan applications is a key process in minimizing credit risk. Therefore, the primary problem of any lender is to differentiate between "low risk" and "high risk" debtors prior to granting credit. The main method used in assessing credit risks is the credit scoring analysis. The incentive of our research is the recent global financial crisis that erupted in September, 2008 which caused a dramatic reduction of loans granted by financial institutions. In this study a sample set of applications from a large Greek financial institution was focused on in order to estimate a credit scoring model for the consumer loans in Greece during the period 2007-2009. Taking into consideration that during a period of financial distress the banks' lending criteria change rapidly, we separated the full data into tree periods and estimated a probit model per year in order to examine the probability of granting a loan changes through years. Moreover, by constructing a bivariate probit model in order to avoid the sample selection effect, we analyzed how the borrower's characteristics influence the decision of granting consumer loans and their performance. Concisely, according to our empirical results, we verified that the probability as well as the criteria of granting a loan changes through years especially in periods of financial distress. Last but not least, we demonstrated that the financial institutions' lending policies are compatible with default risk minimization and thus it is important for banks in order to minimize default risk, to review the borrowers' creditworthiness periodically.

***Keywords*:** Credit Scoring, Greek Banks' Lending Policy, Financial Distress, Bivariate Probit

ACKNOWLEDGEMENTS

This work would not have been possible without the support of the managers in Retail Banking Credit Risk Division. We would like to specially thank them for providing us the data and giving us the chance to actualize the topic we had chosen. We would like to gratefully acknowledge the supervision of Dr. Kosmidou Kyriaki who has been abundantly helpful.

Moreover, we would like to express our sincere gratitude to Dr. Moysiadis Theodoros for his support and insightful comments throughout this research and his assistance on the statistical methodology analysis. We also would like to say a big thank to PhD candidate Ganopoulos Giannis for his support and assistance in the writing process and to Assistant Vice President of Barclays Capital Ioannidis Aristeidis for his help and continuous encouragement. Likewise, we would like to thank the assistant branch manager Mr. Triantafyllou Alexandros for his guidance, support, and invaluable suggestions through his experience in the banking sector.

Our final words go to our families. We want to thank our families; Ganopoulos Basilios, Ganopoulou Theodora, Giapoutzis Nikolaos, Kalaitzoglou Despina and Giapoutzis Anastasios, whose love and constant encouragement are always with us in whatever we pursue. Their undaunted courage and passion will always inspire us; therefore, we dedicate this work to them.

Contents

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Introduction

Consumer credit is granted by banks, retailers and a variety of other lending institutions and is a sector of the economy that has seen rapid growth over the last 50 years. Furthermore, consumer credit constitutes a significant instrument in the financial planning of households. When current income falls below a household's permanent level and assets are not available, credit constitutes a means to increase their welfare and maintain consumption at a level which is consistent with permanent income.

Despite the wide variety of banking services, lending to the public constitutes the core of the income of banks and other lending institutions. Traditional methods of credit risk evaluation incorporated the use of human judgment, based on experience of previous decisions to determine whether to grant credit to a borrower. The emergence of advanced computer technology and the economic pressures, resulting from the increased demand for credit, have led to expansion of sophisticated statistical models to support the credit granting decision.

Credit scoring models may be used as a tool for underwriting and administering all kind of retail credit, including credit cards, direct and indirect installment loans, mortgages, and small business credit (OCC Bulletin, 1997). A variety of types of credit scoring models are used for various activities such as to control risk selection, to manage credit losses, to improve loan approval processing time, to assess new loan programs, and to ensure that existing credit criteria are reliable and consistent.

From a technical perspective, the lending process is a relatively straightforward series of actions involving two principal parties. These actions go from the initial loan application to the successful repayment of the loan or its default. Although retail lending is among the most profitable investments in lenders' asset portfolios, at least in developed countries, increases in the amount of loans also causes increases in the number of defaulted loans. Therefore, the primary problem of any lender is to differentiate between "low risk" and "high risk" debtors prior to granting credit. Credit scoring is a statistical approach that predicts the probability that a credit applicant will default or become delinquent. Credit scoring is broadly applied in consumer lending. Primarily, the amounts lent are much smaller in the case of retail lending, and therefore from the risk management point of view, retail loans are dealt with using a portfolio approach.

In the following section 1.2 it is represented how the recent global financial crisis in 2008 motivated us to proceed with this particular research.

1.2 Financial Distress and Banking Statistics

The onset of the recent financial crisis that erupted in September, 2008 has thrown economies around the world into recession. The starting point of this crisis was sown in the credit boom that peaked in mid-2007, followed by the meltdown of business and consumer loans, mortgages and securitized products. The consequential concerns about the health of financial institutions became a complete banking panic following the collapse of Lehman Brothers and Washington Mutual, and government takeovers of Fannie Mae, Freddie Mac, and AIG. Although the panic in U.S.A. subsided in the first half of October of 2008, the financial results in the Greek Banking and economic environment became observable in the balance sheets and financial statements of banks and companies at the end of 2009.

During the financial distress in Greece a large number of banks has shown to be insufficiently attentive to risks within their portfolios. After a variety of government actions to promote the liquidity and solvency of the financial sector, prices across most asset classes and commodities fell drastically and the cost of corporate and bank borrowing rose substantially. The instructions of Bank of Greece, concerning the commercial and retail banking system, were to diminish their assets originating from loans and advances to customers. Almost every bank in Greece in order to purge their finances reacted simultaneously the same way by decreasing dramatically the amount of loans granted. Evidently, this extraordinary policy of all banks involved that henceforth the criteria to grant a loan would be much more strict and consistent.

Since the period under examination of our research includes the credit crunch, it would be of interest to see the progress of consumer loans in the Greek Banking sector through some evidential statistics. As we can notice, Table 1.1 presents a detailed picture of credit extended to the Greek economy by domestic Monetary Financial Institutions excluding the Bank of Greece. At the end of 2007 the amount of loans granted was €31.915 millions whereas by the end of 2008 it reached the amount of €36.412 millions, which equals to a total increase of 14%. On the other hand, the amount of loans granted between the periods 2008-2009 decreased by 1% (from €36.412 million in 2008 to €36.023 in 2009).

To the extent that X Bank* is concerned, we observe from Table 1.2 that the percentage of change of the amounts granted reduced from 25,3% (period 2007-2008) to 2,5% (period 2008-2009). This is also an indicator that enhances the consequences of the financial crisis.

Before proceeding with the objective of our research, it would be essential to explain in detail the procedure of submitting a consumer loan application in Greece and the characteristics of the loan. This procedure is described in the following section 1.3.

* We denote as X Bank the bank of our sample

Table1.2.1 Credit to domestic non-MFI residents by domestic MFIs excluding the Bank of Greece
(outstanding amounts at end of period in EUR millions)

		Dec-2007	Dec-2008	Dec-2009
1	**GROSS DOMESTIC PRODUCT (GDP) at market prices**	**226.437**	**239.141**	**237.494**
2	**TOTAL CRETIT**	**246.540**	**280.998**	**294.787**
3	**CONSUMER LOANS**	**31.915**	**36.412**	**36.023**
3.1	in Euros	31.908	36.402	36.014
3.2	In non-euro currencies	7	11	8
3.a	Short-term Loans up to 1 year	15.164	16.549	15.558
3.b	Middle-term Loans over 1 and up to 5 years	6.049	6.537	6.164
3.c	Long-term Loans over 5 years	10.702	13.326	14.301
3.i	Credit Card Debt	9.275	10.044	9.538
3.ii	Other Consumer Loans	22.640	26.368	26.485
	Total Debt of Households in Domestic banks (%GDP)	**45,6**	**48,2**	**49,9**

Source: Bank of Greece

Table 1.2.2 Consumer Loans of X Bank for the period 2007-2009 in Euro millions

	31.12.2007	31.12.2008	31.12.2009	Δ% 2007-2008	Δ %2008-2009
Consumer, Credit Cards and other Loans	4.846.256	6.073.941	6.225.000	25,3%	2,5%

Source: Annual Reports of X Bank at the end of 2007, 2008 and 2009
* Mortgages are excluded from the table

1.3 Procedure of consumer loan applications and the characteristics of the loan

The starting point of every loan is the application. When lending institutions receive an application for a loan, the process by which it is evaluated and its degree of complexity may vary greatly. The purpose of this paper is to examine extensively the applications of a consumer loan of a Greek Bank during the period 2007-2009. The main characteristics of this particular loan are its predetermined duration as well as its stable monthly installments. Moreover this product appeals both to already existent customers of the bank but also to new ones.

To begin with, this personal consumer loan is an amortizing loan with exact expiration. The aim of this product is to service the needs of households under the assumption that the borrower is a natural entity and a permanent resident of Greece. The amount of the granted loan can range from €1.500 to €30.000 and its duration fluctuates between 6 to 96 months. The interest rate which yields monthly remains stable for the whole loan duration. According to the transactional behavior of the customer, the bank may decide to diminish the interest rate. This case applies only if the borrower is performing well and has not defaulted. Furthermore, there is a fixed charge for editing the applications of each customer which the bank receives when the loan is granted.

The customer submits an application that is edited electronically by the bank's officer. Moreover, the applicant must submit to the bank all the available documents such as a copy of his/her identity card and a statement by the Internal Revenue Service referring to the last tax year. Occasionally, the bank requires collateral for the loan. In this case the same documents are required for the guarantee as well. After the fulfillment of the application and the submission of all documents, the bank officer forwards the application to the Retail Banking Credit Management Department. After the collection of the applications, the Credit Management Department evaluates the applicants via credit scoring model and, at the same date, informs the branch about the approval or rejection of each application. There is also the possibility according to which the Credit Management Department communicates with the branch due to deficient or even wrong documents of the applicant. In this case the bank officer has to adjust or to correct the documents and re-forward the application to the qualified department. The last stepping-stone is the notification of the applicant about the approval or rejection of his/her loan by the bank officer.

As far as the payments of the installments are concerned, the bank offers various alternative ways to the borrowers in order to facilitate the repayment of the loan. Such ways include cash in hand in any branch of the bank, via a standing order linked to the customer's deposit account, via charge of credit card or via web banking.

Finally, the bank gives the opportunity to the borrower to skip up to two installments per year as well as the option to rearrange the amount of payment by increasing or decreasing the loan duration. The last two benefits are offered by the bank under the assumption that the loan is performing well.

1.4 Objective

In this paper we focus on an analysis of the determinants of obtaining a consumer loan as well as the probability of a loan to default in Greece for the period 2007-2009. It is also very important to refer that the period of our research includes the latest global financial distress that affected Greece too. Therefore it is worthwhile and interesting to see what is happening in the case of the Greek banking system.

The intention of this research is firstly to examine if the criteria, which influence the decision of granting consumer loans, change through time. According to McAllister et al. (1994), it is necessary to review the borrowers' creditworthiness periodically as the changes in economic condition could affect the criteria on lending policy. We separate the full data into tree periods and estimate a probit model year to year. Taking for granted that, in periods of financial distress the financial decisions change rapidly, it is important to examine the financial institutions' reaction to their lending policy.

Secondly, we construct and estimate a bivariate probit model in order to examine how the variables affect the probability of obtaining a loan and the probability of a good loan. We choose this model in order to avoid the sample selection effect that might arise. More specifically, a predictor of default risk in a given population of applicants might be systematically biased because this given population is not made by a random sample. In our research this population is constructed only by the applications that were being accepted and therefore it is important to take it into consideration (Greene, 1998).

Last of all, we explore if our results are in accordance with the strength of the finding that banks' lending policies are not consistent with default risk minimization (Boyes et al. 1989). In other words, we examine if the factors that affect positively (negatively) the probability of obtaining a loan, also affect positively (negatively) the probability of a good loan or not and we compare our results with the inferences of other studies with the same researching view.

1.5 Outline of thesis

The rest of this thesis is organized as follows. Chapter 2 presents an overview of the relevant literature and the background on credit scoring. The benefits and limitations of credit scoring are described in detail as well as the commonly variables used in these models. Moreover, we illustrate the modeling techniques used in other studies with similar researching view. Chapter 3 describes the methodology, the sample data collection method and the estimation of the econometric model. Chapter 4 presents the empirical results and discussion of the results generated by the analysis. Chapter 5 summarizes the major findings followed by the limitation of the research and recommendations for future study.

CHAPTER 2

LITERATURE REVIEW

In this section of the thesis we examine the role of credit scoring and its contribution in the bank's lending decision and we exemplify the advantages and disadvantages of credit modelling. Following this, we proceed with a description of the most commonly variables used in credit scoring and finally we end up with the modelling techniques used in other studies with the same researching view.

2.1 Banks' lending decision

Consumer loan performance determines the profitability and stability of the financial institutions and the key process in minimizing credit risk is to screen the loan applications. Before making any credit decisions, credit analysis should be completed as part of the screening process. Credit analysis, which consists of the valuation of the financial history and financial statements of the applicant credit background, aims to evaluate the borrower's probability of repayment, to determine the financial strength of the borrower, and to minimize the risk of non-payment to an acceptable level. Good borrowers with low credit risk would be granted a loan, while a high risk borrower would be denied. Credit analysis incorporates two major problems: the appraisal of all important factors about an applicant simultaneously and the evaluation of all applicants impartially. Quantitative and qualitative variables, which will be analyzed comprehensively in section 2.4, are used to assess loan applicators. Two main methods are used to estimate a borrower's creditworthiness (Crook, 1996): the loan officer subjective assessment known as judgmental technique and the credit scoring technique. Creditworthiness of an applicant is judged based on the characteristics that meet the requirements for a loan. In case of someone who is not creditworthy, he/she will be unqualified for the loan (Lewis, 1992). The judgmental technique of an applicant's creditworthiness is based on 6 C's namely Character, Capacity, Cash, Collateral, Conditions and Control (Rose, 1993) (see Table 2.1). Glassman et al. (1997) claim that judgmental technique of credit is inefficient, unexplainable, incompatible and nonstandardized. Traditional methods of deciding whether to grant a loan are based on experience of previous decisions and use human judgment of the risk default.

However, the augmented demand for credit, associated with greater commercial competition and the emergence of new computer technology have led to the development of credit scoring technique. Credit scoring methods produce more precise classifications in comparison with subjective judgmental assessments by loan officers. Rosenberg et al. (1994) argue about the credit scoring method advantages over the judgmental technique. For instance, credit scoring is more efficient since it lets loan officer to focus only on ambiguous cases, and it also assists the lenders to review the borrowers' creditworthiness frequently. Therefore, credit scoring models is a preferable technique in credit risk assessment.

In section 2.2 it is stated a comprehensive analysis of credit scoring and its use. The first one who used a statistical model to predict a borrower's probability to default was Altman (1968). His intention was to identify 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 et al., 1997). Following this, many statistical credit scoring models have been developed, such as logistic regression, discriminant analysis, linear probability method, probit model, and neural networks and have been widely used in assessing credit risk (Hand et al., 1997). The main characteristics of each of these techniques are going to be studied in section 2.5.

Table 2.1.1 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

Source: Wang W. (2010)

2.2 What is credit scoring?

Credit scoring was firstly used by U.S. retailers and mail-order companies in the 1950's with the early application of investment portfolios in order to manage and diversify borrowers default risks (Thomas et al., 2002). Currently, the credit scoring models turn out to be one of the most successful techniques of modelling in banking and finance. Based on statistical analysis, credit scoring uses the borrowers' historical data and credit characteristics and the point of this method is to detach the effect of several characteristics of applicants on delinquencies and defaults. Credit scoring models can assist banks to make lending decisions. Credit scoring enhances and sometimes replaces the traditional subjective assessment, since it measures the credit risk of applicants much more accurately and quickly than the latter one. This method of measuring credit risk is mostly used in consumer loans, especially in credit cards but, nowadays, has become commonly used in mortgage lending as well. Furthermore, this scoring method has become applicable in complex business loans, because of the advanced computer technology which increases data accessibility for companies. Therefore, many banks are using credit scoring models to evaluate loan applicants, with the intention to make applicants' default risk more predictable. Credit scoring system is a computerized process producing a score according to various relevant characteristics of the borrower, such as income, profession, age, wealth, previous loans, etc. The final score is obtained by summing the individual borrower's score. Credit will be granted if the score is higher than a predetermined bank's "cut-off-level", otherwise the credit will be refused. The overall idea of credit scoring is based on the statistical probabilities, or in other words the combinations of the borrowers' characteristics differentiating good from bad. In such way a score is generated in order to act as an estimation of the risk level of each new. Crook (1996) argues that the goal 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. Credit scoring analyzes electronically the borrowers' credit history and other characteristics regarding repayment ability that are, in general, provided by borrowers. Based on previous experience with borrowers of similar loan profiles, credit models could predict the default risk of any loan granted. A successful credit model should give high scores to borrowers whose loans would perform well and low scores to borrowers whose loans would not perform well. A fundamental procedure to develop a superior credit scoring model, is to review the borrowers' credit worthiness periodically, since changes in economic environment could affect loan performances. By and large, there is no best credit scoring models. It is probable that some bad borrowers may get a high score and receive the loans, and vice versa. Jesen (1992) argues that using credit scoring, almost 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.

Borrowers have the option to default, in any credit market. Those ones who default are not excluded from future borrowing, which means that there is free entry of lenders and borrowers and also lenders are not allowed to conspire to punish defaulters. Instead, the lender learns from an individual's borrowing and repayment behavior about his/her type and summarizes his/her reputation for not defaulting in a credit score. Limited information about the borrower's behavior and earnings realizations might have as a consequence lenders to grant limited credit or credit at higher interest rates following.

2.3 Pros and cons of credit scoring

Credit scoring has some obvious benefits compared to judgmental techniques for both lenders and borrowers which led to its increasing use in loan evaluation. Firstly, credit scoring models allow immediate handling of the explicit binary decisions. This leaves more time for credit officers to concentrate on the less clear-cut cases that are least well-handled by the models. Credit scoring is also an efficient way to save time, since the loan granting time is reduced to days or hours instead of weeks for consumers. Credit officer can cope with more loan applications than in traditional loan assessment method. This time savings means cost savings to the bank and benefits to the customer as well. Specifically, customers have to provide only the information used in the scoring system, so applications can be less time consuming. Moreover, another significant advantage of credit scoring is that it reduces the probability for bias. This can be justified by the fact that credit scoring is a standard loan granting procedure. Bank officers apply the criteria assembled by the model and applicants are evaluated against these criteria. In comparing to the judgmental methods which are usually negatively biased towards bad borrowers, the scoring models take into consideration the characteristics of both good and bad borrowers. The model assists lenders to ensure that the same underwriting criteria have been applied to all borrowers regardless of gender, nationality or other factors prohibited by commercial law used in credit appraisal.

As far as the classification of loans is concerned relatively to credit risk, the credit analysts can evaluate the riskiness of each borrower with reasonable degree of confidence and the cut-off level can be adjusted according to the risk of each loan portfolio. Thus credit scoring offers the ability to lenders to control risk levels.

Apart from that, this computerized method allows bank officers to review the creditworthiness of the borrowers at regular intervals. As a result, risk monitoring of the individual borrower becomes simple.

Last but not least, credit scoring has the advantage that it is based on historical data. This indicates that lenders can better predict the next applicant's probabilities of performing well or defaulting, due to the database and credit history recorded in the system.

Even though credit scoring reduces costs and increases the efficiency to the loan granting process, the weaknesses of credit scoring should not be disregarded.

To begin with, credit history scoring models offer the benefit of low-cost and reliable quick screening, and since these models are based only on data enclosed in credit reporting agency files, they can be used to monitor more or less any potential customer. However, because they are based on less information than that commonly used in consumer credit screening, they entail the probable disadvantage of being less precise than models based on a fuller set of data. Avery R.B. et al. (2004) examines the potential costs of failing to incorporate situational data into consumer credit evaluations and also discusses practical difficulties related with the development of credit scoring models that incorporate situational data. These complexities arise because of intrinsic limitations of the credit reporting agency databases used to build many scoring models.

Accuracy is also a vital consideration in using credit scoring. The cost savings and other advantages of credit scoring could be negatively affected by loans which are not performing well if the models are not accurate. Credit scoring models are complicated; that means that the models are good only if the data are sufficient and accurate enough. Otherwise the model will generate imprecise results. A sample of both well-performing and bad performing loans should be included in the data used in credit scoring. In order to ensure that changes in the relationship between potential factors and loan performance are captured, the data need to be reviewed regularly and the model should be re-estimated frequently.

Another important characteristic of scoring models is that the borrowers' features are much correlated with their likelihood of repayment and defaults. Although credit models endeavor to forecast the probability of a borrower's default, there is no alternative for knowing the borrower. Often, credit scoring models incorporate human errors because credit scoring cannot replace the decisions of loan officers which are based on informal qualitative knowledge. Therefore, it is important for banks history in order to make a credit decision to deal with customers who have not had an immaculate credit and to combine this knowledge with credit scoring.

An essential factor for a model to be accurate is to make predictions when the economy is either in recovery or in recession. Thus, the data used in the model should cover both good and bad economic periods.

As it is already mentioned credit scoring models are used to predict the probability of default. However, the models, usually, use a sample of accepted applicants only. This selection bias could lead to bias estimation by credit scoring model. According to Schreiner (2003) to avoid the bias in credit scoring model, banks should apply the credit scoring model to loans that are already conditionally accepted by the credit officers. For instance, the First National Bank of Chicago, for small- business loans, rejected about 25% of the applications by using credit modelling while the same applications were later approved by the credit officers. To conclude with, in order to avoid bias in credit scoring, the bank's credit evaluation can be a mixture of the traditional lending through credit scoring models.

2.4 Variables commonly used in credit scoring

Credit scoring entails that all characteristics and available information of the borrower that have obvious connections with default risk should be used in the model. The model's predictive accuracy is maximized if the variables are in sequence added or deleted (Henley et al., 1997). There are two important criteria for variable selection. Firstly, the variables should have significant coefficients and contribute to explanation of the dependent variable's variance. Secondly, the variables should have close correlation with variables included (Dinh et al., 2007). Lewis (1992) claims that there is no need for justification of each variable. If it helps the predictions, it should be used. However, the main factors generally used in credit scoring models include the borrowers' personal characteristics such as income, age, gender, education, occupation, region, time at present address, residential status, marital status, and followed by the borrowers' banking relationship such as collateral value, loan duration, time with bank, number of loans, and current account (Dinh et al., 2007; Roszbach, 2004; Jacobson et al., 2003).

Income denotes the borrower's annual income and is a commonly used proxy of the borrower's financial wealthy and his/her ability to repay (Dinh et al., 2007). Income and the borrowers' default rate are positively correlated since higher income is related with lower default risk (Jacobson et al., 2003).

As far as occupation is concerned, this is a variable used in credit scoring which is greatly associated with income.

Regarding education, it increases the borrowers' ability to repay. We can discriminate borrowers by their educational level as the better educated people are considered to have more stable and higher income employment and thus a lower probability of default.

Employer signifies the type of company for which a borrower works such as stated-owned, foreign, joint-stock company, etc. This variable is important for the reason that the type of the borrower's company could be a proxy for income level and stability. Missing values of this variable are also very enlightening since borrowers who do not answer this question show the highest default rates.

The variable "time with employer" measures the number of years that the borrower has been working for the current employer. It displays 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). It should be emphasized that the length of time with employer may discriminate against women, since women's length of employment decreases due to pregnancy and childbearing.

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

Although gender is considered to be biased in many industrialized countries due to statistical rates of men versus women, there is plenty of evidence that women default less frequently on loans possibly because women are more risk adverse (Coval et al., 2000).

Region depicts the area of the country that borrower lives. A common proxy for this variable could be the postal code. As people of similar wealth are likely to live in the same location, the geographic criterion can imply a borrower's level of financial wealth. Some suburb might attract richer residents and this could lead to increase in real estate property prices. Moreover, this also influences the collateral value and probability of default.

The residential status points out whether borrowers own their home, rent, or live with their parents. In the case of home ownership this variable could indicate the borrowers' financial wealth. In addition, residential status also signifies financial pressure on borrowers' income through insurance fees, taxes, or electricity costs. Crook et al. (1992), notice that borrowers living with their parents are less likely to default.

'Time at present address" refers to the number of years the borrowers have been living at their current address. As stated by Crook et al. (1992), the default risk and the time at present address are negatively correlated, illustrating that it might be a proxy for the borrowers' maturity, stability, or risk aversion. Changing address might be a signal that a borrower's financial wealth is high or improving rapidly.

Marital status has an impact on the borrower's level of responsibility, reliability, or maturity. Statistics show that default rates are higher for married than single borrowers. Dinh et al. (2007) demonstrated that the marital status is associated with the number of dependants which in turn replicates financial pressure on the borrower and borrower's ability to repay a loan.

Collateral is a type of guarantee in order to minimize the borrowers' probabilities of default. In the retail loan sector, requiring collateral may be a signal of risk. For example, if the loans that the house serves as collateral, the probability of default is very low. This is due to the fact that the borrowers are risk adverse and fear of losing their house. The higher the collateral value the higher the incentive for the borrowers to repay the loan since they would avoid loosing 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 et al., 2007).

Loan duration indicates the maturity of loans in months. This variable is an outcome of the negotiation between the bank and the borrower. There is a probability that a borrower might be rejected for a shorter loan while he/she might be accepted for a longer one of the same size as the pressure on his/her income is decreased. Loan duration represents the borrowers' intention, risk aversion, or self-assessment of repayment ability.

"Time with the bank" measures the length of the 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 the lower the default risk becomes. However, this variable should be revised regularly due to unexpected changes in the borrowers' situation.

Number of loans counts the number of loans a borrower has received from the bank during the whole relationship with it. Many borrowers have a sequence of historical loans and quite often more than one loan from the same bank. This proxy is informative about the borrower's default risk because a borrower who has not met his/her pay off obligations to an already granted loan, will also have difficulties in receiving a new loan. This variable, therefore, reflects the difficulty for a defaulted borrower to receive further loans from the same bank.

Current account is a binary variable which reflects whether the borrower holds a current account with the bank. This variable is relevant and indicates up to some degree the borrowers' financial wealth, and relationship between the borrower and the bank. However, the borrowers who hold current accounts with their banks have a lower default risk.

2.5 Modelling techniques of credit scoring

The most common statistical methods used to estimate credit scoring models in assessing borrowers' credits, are discriminant analysis linear probability models probit models and logit models. The last three methods estimate the default rate based on the historical data on loan performances and the borrowers' characteristics. Traditional credit scoring methodology has focused on using techniques such as discriminant analysis and linear regression to distinguish between applicants who are assumed to belong to one of two classes, namely good and bad credit risks. Previous papers which include the application of these methods to credit scoring include Myers and Forgy

(1963), Eisenbeis (1978), and Reichert et al. (1983). However, according to Eisenbeis (1978) and Reichert et al. (1983) both these techniques are subject to the conceptual problems.

In particular, the main purpose of discriminant analysis is to predict group membership based on a linear combination of the interval variables. The procedure begins with a set of observations where both group membership and the values of the interval variable are known. The end result of the procedure is a model that allows prediction of group membership when only the interval variables are known. Furthermore, a second purpose of the discriminant analysis is an understanding of the data set, as a careful examination of the prediction model that results from the procedure can give insight into the relationship between group membership and the variables used to predict group membership.

According to Mester (1997) discriminant analysis divides borrowers into high and low default-risk classes. On the other hand, Hand et al. (1996) show that the discriminant function obtained by segmenting a multivariate normal distribution into two classes' optimal discriminant function. Altman et al. (1981) based on discriminant analysis, generates indices depending on whether or not the applicant belongs to the population of those who would be defaulters. However, as stated by Boyes et al. (1989), since the overall incentive of a lender is profit maximization, this view of Altman et al. (1981) may be misleading.

In order to overcome this problem Boyes et al (1989) suggests the probit model which assumes that the probability of default follows the standard cumulative normal distribution function. The goal of his research is to display how expected earnings on revolving credit loans rely both on probability of default and maintained balances. Boyes et al. (1989) and Jacobson et al. (2003), applying a bivariate probit model, provide evidence that by classifying applicants according to predicted default probabilities, banks can minimize the expected default rate but this does not solve the problem of profit or utility maximization. Tor Jacobson et al.(1998) investigate how banks provide loans in a way that is not consistent with default risk minimization and they concluded that the size of the loans does not affect the default risk. Banks are not faced with a trade-off between risk and return even if they are risk averse. It is also shown how estimating Value at Risk can enable banks to develop alternative lending strategies on the basis of their implied credit risks and loss rates

Regarding the linear probability model, it could present reasonable prediction results compared to discriminant analysis and logit models (Collins et al.1982). The aim of linear probability is to search for a linear combination of explanatory variables. It assumes there is a linear relationship between the default rate and the factors. However, Pyndick et al. (1998), and Greene (1998), 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 is generally heteroscedasticity, it leads to inconsistent estimation problem.

According to Henley et al. (1996), 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. In this type of models the probability of default is logistically distributed. Logistic Regression, which is very similar to the probit model, has been also applied with success by Wiginton (1980), Gilbert et al.(1990) and Leonard(1993). This method was first proposed by Chesser (1974) for forecasting commercial loan non-compliance. The report of Steenackers et al. (1989), which concentrated on the origin of a credit scoring model for personal loans, is also based on a logistic regression model. In particular, the assumption that is made is that the probability of a loan to be good is dependent on the level of the characteristics of each applicant. The two main questions answered by this report are firstly which characteristics are proper to be used in the credit scoring model as variables that can discriminate between a good and a bad loan, and secondly how to obtain the score for each characteristic. The logistic modelling approach is commonly used to model the bank's lending decision. According to Collins et al. (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 discriminant analysis (Wiginton, 1980).

The paper of Henley et al.(1996) refers to the application of the k-nearest-neighbour (k-NN) method a technique in pattern recognition and nonparametric statistics, to the credit scoring problem. More specifically, this paper denotes the problem of choosing a suitable technique for distinguishing between a population of good and bad credit risks. Assessment was made between the performance of the k-NN method and a range of other classification techniques. Logistic and linear regression and decision trees were selected to represent the accepted credit scoring techniques. It was found that k-NN method performed well, achieving the lowest expected bad risk rate. It was claimed that a set with equal proportions of bad risks should be used to categorize future applicants regardless of the population bad risk rate. Furthermore, it was found that, given the power of technology, it is possible to classify an applicant within seconds and that k-NN method can provide reasons for turning down credit.

Duration or survival analysis estimates not only if, but also when a loan will default. Such models provide us also information about the profitability of customers on a product since they can deal not just with default risk but also other events that may affect profit like early repayment of a loan. Survival analysis was first suggested by Narain B.et al. (1992) and has been evolved by Stepanova M. et al. (2001). The report of Carling et al. (2001) analyzes the features that influence the time to maturity on consumer loans and develop a distribution of conditional expected durations of loans, representing how a loan application can be evaluated by calculating its expected earnings. The conversion of consumer loans from an active to a dormant status is examined and this transition is associated with the characteristics of the loan applicant. Last but not least, the expected return on each loan is calculated by using the predicted time to dormancy and it is compared with the revenue of a benchmark loan. Other examples of duration analysis are Kiefer (1988) and Diebold et al. (1990).

Santos Silva J. M. C. et al. (2000) developed a model for credit scoring in order to estimate the default probabilities using a data set on personal loans granted by a Spanish bank. This model is based on the beta-binomial distribution. The first problem that is raised is the sample selection. In particular, it examines the decision to accept or refuse the credit applications using data on the borrowers that is not available to the formation of the credit scoring model. The second issue that this paper deals with is that the repayment behavior of a borrower may change after he is classified as defaulter. In most cases the bank compels the borrower to repay his debt, and that may modify the borrower's behavior. However, the results of this report should be judged with great caution since the data set used is relatively poor.

According to the literature, there is no best method for estimating credit scoring models and new methods continue to evolve. Nevertheless, the intention of most of the models described above is to make a distinction between applicants who would repay from those who would default.

The modelling technique used in our research was the bivariate probit model, that is analyzed more extensively in the following Chapter 3.

CHAPTER 3

METHODOLOGY

3.1 Research Question

The intention of this research is to investigate the effects of the borrower's characteristics in the decision of granting consumer loans and in their performance. We use credit scoring model in order to examine the probability of obtaining a consumer loan as well as the probability of a loan to default in Greece for the period 2007 to 2009. The data sample consists of loan applications including the borrowers' characteristics (age, gender, annual income, marital status, nationality, region, type of residence), the loans' characteristics (loan duration, loan status, if there is a warrantor or not), and other variables indicating the relationship of the applicant with the bank (other banking transactions, other loans).

In order to estimate credit scoring models, several statistical methods are used which are also named score cards or classifiers. These models apply predictor variables (otherwise characteristics) from application forms or other sources such as customer's credit history and customer's data, so as to attribute estimates of the probability of obtaining a loan and the probability of default (Hand et al. 1997). Methods of statistical models include discriminant analysis (Dunn at al. 1976), linear probability models (Turvey 1991), probit models (Luftburrow at al. 1984) and logit models (Mortensen et al. 1988).

The probit and logit models are widely used by a lot of researchers such as Wihinton (1980), Schwartz et al. (1990), Roszbach (2004), Chang S. et al. (1998), Games F. et al. (2000), Hayden (2003) and Huyen D.T. (2006). These two models differ in the distribution function. The first of the models uses the normal distribution function whereas the later the logistic one. This difference has as a consequence the logistic model to have slightly flatter tails than the probit model does. According to Greene (2002), the two models' results are nearly identical. We use in our research the probit model.

The purpose of this research is firstly to examine the probabilities that we have mentioned above and secondly to explore the strength of the finding that banks' lending policies are not consistent with default risk minimization (Boyes et al. 1989). According to Boyes' et al. (1989) study, if the financial institutions' lending policies are compatible with default risk minimization one should find same signs for the parameter of one particular explanatory variable in the model of probability of obtaining a loan and in the model of probability of default. In other words the variables that increase the probability of a positive decision of taking a loan should also increase the possibility of failing to pay, or vice versa. But the results of his study and also the results of Roszbach (2004) and Jacobson (1998) prove the opposite. Specifically they found variables that bank uses to increase the likelihood of accepting the application and give the loan but also increase the risk of default (Roszbach 2004).

Taking into account that the credit scoring model is intended to provide precise valuation of every applicant's probability of default, loan officers should assign loan obtaining criteria in order to maximize profitability (returns) and minimize the default risk. Therefore it is worthwhile to see what is happening in the special case of Greek banking system and subsequently relate our results to the general frame that surrounds this scope. It is also very important to refer that the period of our research includes the latest global financial distress which affected Greece too. Taking into consideration this parameter our results will be even more interesting and commendable. Considering these financial conditions, we will also try to find if the criteria that affect the probability of obtaining a loan in these years change year to year.

3.2 Sample Data

The research sample period is from 2007 to 2009. The data are obtained from a commercial bank of Greece, denoted in our analysis as X bank, for confidence reasons. All the applications that we received were submitted in stores (the X bank has branches across the country) where the potential borrowers applied for instant credit to fund the purchase of a consumer good. The total number of observations from the available data set is equal to 14738. 8095 Out of 14738 applications were accepted and 6643 were rejected. The data set also includes the status of the accepted applications (loan status) which is divided into two categories the "good" loans and the "bad" loans. The monitoring day that the loan status is referred is the 3rd of August of 2010. The bank classifies as good loans the loans that are still active and the loans that have been completely taken up. The loans with more than 90 days payment delay are classified as bad loans. In the data set 6752 of the accepted loans are good loans and 1343 are bad loans which is approximately 19.9% (default rate).

The data set also includes the date that the potential customers submitted the application, applicants' characteristics such as gender, age, marital status, annual income, residence type, region, loan duration and some other information such as applicants' relationship with the bank.

3.3 Description of Variables

The variables that we used for our credit scoring model have common characteristics with the variables that Thomas (2000), Crook et al. (1992), Boyes et al. (1989), Roszbach (2004), Jacobson et al. (2003) , Dinh et al. (2007) and a lot of other researchers used in their studies. We used eleven variables in our research in order to assess the probability of obtaining a loan and the probability of default. These variables are the gender, age, annual income, other banking transactions, marital status, loan duration, postal code, nationality, residence type, underwriter and other loans.

Age measures borrowers' age in years. We classify this variable into three categories: 19-40 (young borrowers), 40-60 (middle-aged borrowers) and older than 60 years old (elderly).

Gender is divided in female and male and even though some studies (Schreiner 2004) showed that this variable lose its effect when it is correlated with other factors, we include it in order to see if there is any influence in the probability of obtaining a loan and the probability of default.

Marital status is the variable that indicates if the applicant is married, divorced or single. This variable shows how responsible, credible and mature one can be. Specifically, taking marital status into account we can observe how many dependants of the borrower there are and consequently how much the financial pressure is and what is borrower's ability to pay back his/her loan (Dinh et al. 2007).

Income is the variable that defines the annual income of the applicant. This variable is divided it into two classes: annual income less than 15000 euro (low-middle income) and annual income more than 15000 euro (high income). It is obvious that the income indicates to a great extend the financial health of the borrower and his/her ability to be reliable to his/her loan installments.

Postal code specifies where the applicant lives. This variable is divided into two groups: borrowers that live in the two big cities (Athens, Thessaloniki) and borrowers that live in other cities of the country (Kavala, Drama, Patra, and so on). People that live in big cities tend to have more job opportunities but in the same time their spending and cost opportunities are also more. Thus we will try to clarify the ambiguous effect of this variable.

Residence type declares whether the borrower lives in a house of his/her own (owner) or he/she is renting one (tenant) and this implies the borrower's financial wealth too.

Nationality shows if the applicant is Greek or not (other) and we will investigate whether there are payment consistency differences between them.

Banking transactions and other loans variables indicate the relationship of the applicant with the bank or other banks. Specifically the first variable shows if the borrower is already customer of the X bank or not whereas the second one specifies if he/she has other loans that should be paid.

Underwriter gives financial support and takes responsibility for paying any costs associated with the activity he or she underwrites. This creates some sort of security in the payment of the loan and therefore we will investigate if it affects the decision of obtaining a loan.

Loan duration measures the maturity of a loan in months. The specific consumer loan may have duration from 6-96 months. We categorize this variable into two groups: short term loans with duration less than or equal to 12 months and long term loans with duration more than 12 months.

All the above variables have been classified into different groups for the purpose of our research. It is important to mention that one group of each variable will not appear in the results. This happens because one group of the variable is treated as the benchmark for the other groups. For instance, the gender variable is divided into two classes: female and male. These two groups will transform into two dummy variables (indicator variables) and take values zero or one to indicate the absence or presence respectively of the effect. Consequently one of them must be omitted because otherwise their sum will be one resulting in perfect multicollinearity (dummy trap problem) (Suits 1957).

More specifically applicant characteristics (dummy variables) include:

AGE = Group1 (age1): 1 if the applicant's age is between 18 to 40 years old, 0 otherwise

Group2 (age2): 1 if the applicant's age is between 41-60 years old, 0 otherwise

Group3 (age3): 1 if the applicant's age is 61 or above, 0 otherwise

GENDER = Group1 (male):1 if the applicant is male, 0 otherwise

Group2 (female):1 if the applicant is female, 0 otherwise

MARITAL STATUS = Group1 (married):1 if the applicant is married, 0 otherwise

Group2 (divorced):1 if the applicant is divorced, 0 otherwise

Group3 (single): 1 if the applicant is single, 0 otherwise

INCOME = Group1 (lowincome):1 if the applicant has annual income less than /or 15000 euro, 0 otherwise

Group2 (highincome):1 if the applicant has annual income more than 15000euro, 0 otherwise

POSTAL CODE = Group1 (bigcity):1 if the applicant lives in Athens or Thessaloniki, 0 otherwise

Group2 (othercity):1 if the applicant lives in other cities of Greece, 0 otherwise

RESIDENCE TYPE = Group1 (tenant):1 if the applicant is renting a house, 0 otherwise

Group2 (owner):1 if the applicant owns a house, 0 otherwise

NATIONALITY = Group1 (nationg):1 if the applicant is Greek, 0 otherwise

Group2 (nation):1 if the applicant has other nationality, 0 otherwise

BANKING TRANSACTION = Group1 (bankbus):1 if the applicant has other banking transactions with the bank, 0 otherwise

Group2 (nobankbus):1 if the applicant has not other banking transactions with the bank, 0 otherwise

OTHER LOANS = Group1 (wotherloans):1 if the applicant has other loans to pay, 0 otherwise

Group2 (otherloans):1 if the applicant has not other loans to pay, 0 otherwise

UNDERWRITER = Group1 (wunderwr):1 if the applicant has underwriter, 0 otherwise

Group2 (underwr):1 if the applicant has not underwriter, 0 otherwise

LOAN DURATION = Group1 (loandur1):1 if the applicant wants a shot term loan with duration less than/or 12 months, 0 otherwise

Group2 (loandur2):1 if the applicant wants a long term loan with duration more than 12 months, 0 otherwise

The dependent variables are:

LOAN STATUS1 (loanst) = 1 if the loan is accepted, 0 if the loan is rejected

LOAN STATUS2 (loanst2) = 1 if the loan is good/does not default, 0 if the loan is bad/default

The general models of our research are the functions of the observed variables with the explanatory variables.

$$Obtaining\ consumer\ loan = f(Applicant's\ characteristics)$$
$$Consumer\ loan\ default = f(Applicant's\ characteristics)$$

3.4 Econometric Model

In the first model, as we mentioned above, we are interested in modelling the loan status of each individual in our sample which is whether one applicant obtains a loan or not. The applicants differ in age, marital status, race, income and other observable characteristics, which we denote as x. The goal is to quantify the relationship between the individual characteristics and the probability of obtaining a loan.

Taking values of zero and one, the binary dependent variable y, makes the linear regression of y on x not suitable. The most important reason is that the implied model of the conditional mean puts inappropriate restrictions on the residual of the model. Moreover, the fitted value of y from a simple linear regression is not limited to lie between zero and one. Therefore we overcome the problems with the linear model by choosing binary choice models (otherwise named univariate dichotomous models) which are specifically constructed to model the 'choice' between two discrete alternatives (Greene, Econometric Analysis,5th edition 2002).

Suppose that we model the probability of observing a value of one as:

$$\Pr(y_i = 1 | x_i, \beta) = 1 - F(-x_i'\beta), \quad (3.1)$$

F is a continuous, strictly increasing function that takes a real value and returns a value of ranging from zero to one. The choice of the function F determines the type of binary model. So the probability of observing a value of zero is:

$$\Pr(y_i = 0 | x_i, \beta) = F(-x_i'\beta). \quad (3.2)$$

The set of parameters β reflects the impact of changes in x on the probability. For example, among the factors that might interest us is the marginal effect of family condition on the probability of obtaining a loan. The problem at this point is to devise a suitable model for the right-hand side of the equation. (Greene, Econometric Analysis,5th edition 2002)

In our analysis we use the normal distribution giving rise to the probit model:

$$\Pr(y_i = 1 | x_i, \beta) = 1 - \Phi(-x_i'\beta) = \Phi(x_i'\beta) \qquad (3.3)$$

where Φ is the cumulative distribution function of the standard normal distribution.

It is possible to derive a binary choice model from underlying behavioral assumptions. This leads to a latent variable representation of the model. For example, let us look at the decision of a married female to obtain a loan or not. The utility difference between obtaining a loan and not taking one depends upon the stated income but also on other personal characteristics, like the woman's age, other banking transactions that might have, etc. Thus, for each person i we can write the utility difference between granting a loan or not granting one as a function of observed characteristics,x_i and unobserved characteristics, ε_i. (Verbeek 2008) Assuming a linear additive relationship formula, we obtain for the utility difference, denoted y_i^*,

$$y_i^* = x_i'\beta_1 + \varepsilon_i \quad for\ i = 1,2 \ldots N \qquad (3.4)$$

where y_i^* is the dependent variable, $\beta_1$are the coefficients, $x_i'$are the regressors variables, $\varepsilon_i$are the error terms and i is the volume of the observations.

The error terms ε_i are normally distributed with mean zero and variance σ^2.

So, the observed dependent variable is determined by whether y_i^* exceeds a threshold value:

$$y_i = \begin{cases} 1 & if\ \ y_i^* > 0 \\ 0 & if\ \ y_i^* \leq 0. \end{cases} \qquad (3.5)$$

The responding variable takes value of 1 if the loan was granted which means that the application was accepted and the value of 0 if it was rejected.

At this point we have to report that we do not model how one decides the amount of credit one applies for. Furthermore we take for granted that the individuals, who eventually take the loan, receive the exact amount of credit they applied for.

In the second model we want to find the probability of a loan to become "bad", in other words the probability of default. So we use again a binary dependent variable z_i in this case to denote "good" or "bad" loan (probit model). We also indicate an unobservable variable with the subscript *:

$$z_i^* = w_i'\beta_2 + u_i \quad for\ i = 1,2 \ldots N \qquad (3.6)$$

where $z_i^*$is the dependent variable, $\beta_2$are the coefficients, $w_i'$are the regressors variables, $u$are the error terms and i is the volume of the observations.

$$z_i = \begin{cases} 1 & if\ z_i^* > 0 \\ 0 & if\ z_i^* \leq 0. \end{cases} \qquad (3.7)$$

The dependent variable z_i takes the value 1if the loan is paid and 0 if loan is in delay. It is very important to mention that the distinction of a loan whether defaults or not is observed only if a loan is granted. For this reason, we have a censoring rule for (y_i, z_i) and an observation rule too (Jacobson et al. 2003).

Table 3.4.1 observation rule

	$z_i \leq 0$	$z_i > 0$
$y_i > 0$	$(0,.)$	$(0,.)$
$y_i \leq 0$	$(1,0)$	$(1,1)$

Thus we must now account for the sample selection rule. The problem that might arise in the model is that there are possibly factors which access the granting decision but do not appear clearly in the rule and these same factors affect the response in the default equation and then the default equation may produce biased predictions. Thus a predictor of default risk in a given population of applicants can be systematically biased because this given population is not made by a random sample. This population is constructed only by the applications that were being accepted (Greene, Sample selection in credit-scoring models 1998).

Therefore an estimation of the two univariate models (simple model) produces a biased set of coefficient estimates and eventually a biased estimate of the default probability. For this reason we will appoint and estimate a model of the default probability that account for the sample selection effect. We will use a bivariate probit specification to model this (Greene, Sample selection in credit-scoring models 1998). This model consists of two simultaneous equations, as we described previously.

$$z_i^* = w_i'\beta_2 + u_i \quad for\ i = 1,2 \dots N \qquad \text{(Default equation)}$$
$$z_i = 1\ if\ and\ only\ if\ z_i^* > 0, and\ 0\ else. \qquad (3.8)$$
$$y_i^* = x_i'\beta_1 + \varepsilon_i \quad for\ i = 1,2 \dots N \qquad \text{(Loan granting equation)}$$
$$y_i = 1\ if\ and\ only\ if\ y_i^* > 0, and\ 0\ else.$$

Where:

$z_i\ and\ w_i'\ are\ observed\ if y_i = 1$

$y_i\ and\ x_i'\ are\ observed\ for\ all\ applicants.$

Selectivity:

$[\varepsilon_i, u_i] \sim N2[0,0,1,\rho_{\varepsilon u}]$

The vector of attributes,x_i' are the factors used in approval decision. The probability of default given that a loan is accepted is:

$$P(z_i = 1|y_i = 1\ ) = \frac{\Phi_2(x_i'\beta_1, w_i'\beta_2, \rho)}{\Phi(x_i'\beta_1)} \qquad (3.9)$$

where Φ is the bivariate normal cumulative probability.

The probability of not default given that a loan is accepted is:

$$P(z_i = 0|y_i = 1\ ) = \frac{\Phi_2(-x_i'\beta_1, w_i'\beta_2, -\rho)}{\Phi(x_i'\beta_1)} \qquad (3.10)$$

If ρ is equal to zero then the selection is of no consequence and the univariate models are appropriate.

Taking into account all the above, we conclude that the types of observations are three: no loans, defaulted loans (bad) and not defaulted loans (good). The likelihood function will take the following form:

$$l = \prod_{no\ loans} pr(loan) \times \prod_{bad\ loans} pr(bad\ loan) \times \prod_{good\ loans} pr(good\ loan)$$
(3.11)

The loglikelihood (see in Appendix A) is:

$$\begin{aligned} \ln = & \sum_{i=1}^{N} (1 - y_i) \times \ln\left[1 - \Phi(x_i'\beta_1)\right] \\ & + \sum_{i=1}^{N} y_i \times (1 - z_i) \times \ln\left\{\Phi(x_i'\beta_1) - \Phi_2(x_i'\beta_1, w_i'\beta_2; \rho)\right\} \\ & + \sum_{i=1}^{N} y_i \times z_i \times \ln \Phi_2(x_i'\beta_1, w_i'\beta_2; \rho) \quad (3.12) \end{aligned}$$

where $\Phi(.)$ and $\Phi_2(.,.,\rho)$represent the univariate and bivariate standard normal c.d.f., the later with correlation coefficient ρ.

CHAPTER 4

EMPIRICAL RESULTS AND DICUSSION

4.1 Descriptive Statistics

The total number of consumer loans applications from X bank was 14738. 8095 of these applications were accepted and 6643 were rejected. We should to mention that the data sample did not have any missing values. The sample is consisted of personal characteristics (gender, age, marital status, residence type, nationality and postal code), bank related characteristics (other loans, other banking transactions, loan duration, underwriter) and financial characteristics (annual income). The period of our research is from 2007 to 2009.

Table 4.1.1 Total number of applications

			obtaining a loan		
			rejected	accepted	Total
date	2007	Count	1781	3809	5590
		% within date	31.9%	68.1%	100.0%
	2008	Count	2118	2461	4579
		% within date	46.3%	53.7%	100.0%
	2009	Count	2744	1825	4569
		% within date	60.1%	39.9%	100.0%
Total		Count	6643	8095	14738
		% within date	45.1%	54.9%	100.0%

Figure 4.1.1 Bar chart

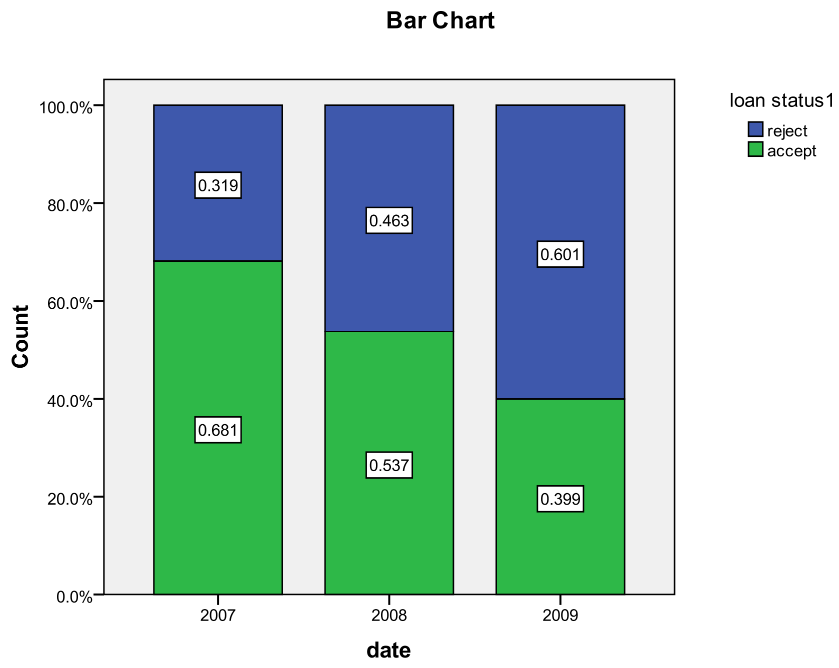


Table 4.1.2 Chi-square test

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	811.099[a]	2	.000*
Likelihood Ratio	821.339	2	.000*
Linear-by-Linear Association	810.933	1	.000*
N of Valid Cases	14738		

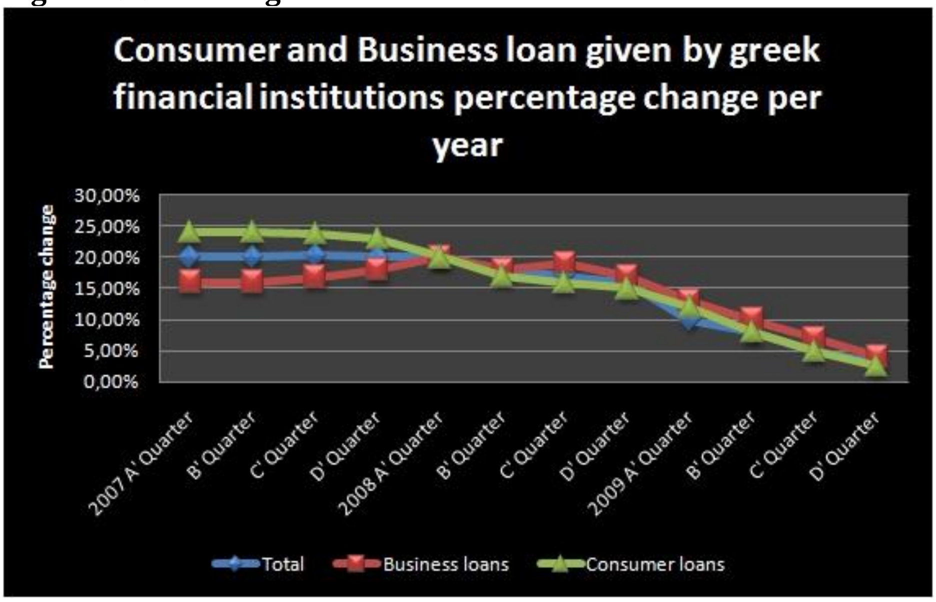
a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 2059.43.
* Significance level 0.05 (2-tailed).

Table 4.1.1 reports the number of applications in total were in 2007, 2008 and 2009 respectively. As we can see, the number of total applications slightly decreased and the number of consumer loans that were accepted decreased. More specifically almost 40% of the applicants obtained a loan in 2009 whereas the 68% in 2007. We observe the exact opposite results for the rejected loans. Figure 4.1.1 gives us a better look of the percentages of accepted and rejected loans together.

This decrease of the loans might be explained by the trial of financial institutions to minimize the volume of the consumer loans due to global financial distress. In periods with such conditions there is financial uncertainty and liquidity shortage therefore banks want to protect their capitals. As it is obvious the difference is bigger in 2009 when the financial distress starts to affect our country intensively. People, in their effort to protect themselves as well, they control their consuming behavior but on the other hand their resource of liquidity, in periods like this, is from banks. This is the reason why we notice a minor decrease in the number of applications.

In order to ensure that the number of applications has to do with the period we did a chi-square test, to see if this relationship is statically significant, with null hypothesis that the two variables (date and loanstatus1) are independent under the condition that the expected observations in the crosstabulation matrix are at most 20% less than five. As we can see in table 4.1.2 in our case the respective percentage is 0%, the probability is .000<0.05 so we reject the null hypothesis and we conclude that the two variables are dependent one another.

Figure 4.1.2 Loans given from Greek banks



Source: Bank of Greece

According to statistical data from bank of Greece we see in figure 4.1.2 the percentage change per year of consumer loans that are given by all Greek banks from 2007 to 2009.

The green line represents the consumer loans. We notice a major decrease through these years which is compatible with the results of our data sample.

Table 4.1.3 Total number of granted applications

			defaulting a loan		
			bad loan	good loan	Total
date	2007	Count	669	3140	3809
		% within date	17.6%	82.4%	100.0%
	2008	Count	465	1996	2461
		% within date	18.9%	81.1%	100.0%
	2009	Count	209	1616	1825
		% within date	11.5%	88.5%	100.0%
Total		Count	1343	6752	8095
		% within date	16.6%	83.4%	100.0%

Table 4.1.4 Chi-squared test

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	46.871[a]	2	.000*
Likelihood Ratio	49.962	2	.000*
Linear-by-Linear Association	24.031	1	.000*
N of Valid Cases	8095		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 302.78.
* Significance level 0.05 (2-tailed).

Table 4.1.3 indicates how many loans turn to be bad and how many loans turn to be good. The total number of loans that defaulted is 1343 which constitutes a default rate of 19.9% whereas the not defaulted loans are 6752. The total amount of good loans is larger than the total amount of bad loans. Taking into account that the financial institutions did not give so many consumer loans as before, we can infer that they might also take harder measures in the decision of providing a loan. This practice was followed as banks in financial distress try to minimize the default risk in order to avoid insolvency and illiquidity.

Comparing the percentages of successful loans in 2007 and 2009, which are 82.4% and 88.5% respectively, we result that as the crisis affects our country so the criteria to take a loan become more stringent and consequently that we have a better performance of loans.

We also run a chi-squared test to see if there is a statistically significant relationship between the quality of loans (loanstatus2) and the time period (date) and as we can see in table 4.1.4, taking into account that 0 cells have excepted count less than 5, the probability is $0.00<0.05$ so we reject the null hypothesis which is that the two variables are independent. Thus, the two variables are related to each other.

Table 4.1.5 Descriptive statistics

date			age	loan duration	stated income
2007	Mean		46.41	53.68	16063.67
	95% Confidence Interval for Mean	Lower Bound	46.08	53.03	13760.23
		Upper Bound	46.73	54.32	18367.11
	5% Trimmed Mean		46.19	53.68	12904.07
	Median		46.00	48.00	11565.00
	Variance		156.041	608.387	7.718E9
	Std. Deviation		12.492	24.666	87849.841
	Minimum		19	6	0
	Maximum		87	96	4661220
2008	Mean		46.05	65.19	17464.68
	95% Confidence Interval for Mean	Lower Bound	45.69	64.36	15548.81
		Upper Bound	46.41	66.01	19380.54
	5% Trimmed Mean		45.90	66.33	13973.54
	Median		46.00	60.00	12176.00
	Variance		155.300	816.376	4.373E9
	Std. Deviation		12.462	28.572	66128.345
	Minimum		18	6	0
	Maximum		85	96	3839630
2009	Mean		45.57	66.19	20188.71
	95% Confidence Interval for Mean	Lower Bound	45.21	65.37	16134.64
		Upper Bound	45.94	67.01	24242.79
	5% Trimmed Mean		45.45	67.41	15591.48
	Median		45.00	60.00	13780.00
	Variance		158.719	806.896	1.954E10
	Std. Deviation		12.598	28.406	139778.043
	Minimum		19	6	0
	Maximum		82	96	9107910

Table 4.1.5 gives details about descriptive statistics of the quantitative variables of the sample. The results are categorized by year in order to have a better view of the variables across the time. The age variable has a mean value 46.41 in 2007, 46.05 in 2008 and 45.57 in 2009 which denotes that people around forty age apply for consumer

loans. The standard deviation of the age is approximately 12 in every year which shows that the majority of people who applied for a consumer loan range from 30 to 50 years old. According to Boyle et al. (1992) and Thomas (2000), the older the customers the more risk averse they are and more impossible to fail to pay. Looking at the average of loan duration we observe an increase from year to year. More specifically the mean value is 53.68 in 2007, 65.19 in 2008 and 66.19 in 2009. This increase indicates that the applicants want a bigger period to pay back the loan as they seem to prefer long term loans. According to Roszbach (2004) and Jacobson et al. (2003) the long duration of a loan relates with an increase in probability of defaulting a loan. Moreover the average annual income is about 16.100 € in 2007, 17.500€ in 2008 and 20.200€ in 2009. At first glance, we see an increase in the income of the potential borrowers but looking at the standard deviations which are extremely high (87.850, 66.128 and 139.778 respectively) we conclude that the annual income varies widely among the sample mean. In other words there are applicants that have extremely high income (taking into account the maximum value) and others that have minor or no annual income (minimum value).

Table 4.1.6 Characteristic of applicants

Variables	Accept loan	Reject loan	Total
Age 18-40	2588	2803	5391
40-60	4054	3066	7120
> 60	1453	774	2227
Loan duration ≤ 12	375	138	513
> 12	7720	6505	14225
Income $\leq 15000€$	4923	4389	9312
$> 15000€$	3172	2254	5426
Oth.BankTr. Yes	6447	4457	10904
No	1638	2186	3834
Nationality Greek	7923	6318	14241
Other	172	325	497
Marital St. Married	4959	3503	8462
Divorced	2311	2450	4761
Single	825	690	1515
Oth.loans Yes	1639	0	1639
No	6456	6643	13099
Postal code Big city	4731	3700	8431
Other city	3364	2943	6307
Resid.type Tenant	1296	1261	2557
Owner	6799	5382	12181
Gender Female	2984	2490	5474
Male	5111	4153	9264
Underwriter Yes	854	1039	1893
No	7241	5604	12845

Table 4.1.6 shows the characteristics of all the applicants. As we can see most of them are in the age group of 40-60 years old. This age group is also the most preferable for taking a consumer loan. Furthermore banks were more willing to provide loans with bigger length which indicates that banks want to create long-term relationships with customers. As far as income is concerned, we observe that the applicants with low income (≤15000€) are twice as much the applicants that have high annual income. We also notice that the majority of the applicants who are being accepted have already relationship with the bank (such as deposits, credit cards) and have Greek nationality. Finally the most applications are done by men who are the majority of the applicants obtaining a loan as well.

The correlation matrix in Appendix B indicates that the variables that are positively correlated to the accepted loans are bank transactions, postal code, nationality loan duration and marital status by descending order whereas the variables that are negatively correlated are other loans, age, underwriter, income. For instance an applicant who has deposits in the X bank (other banking transactions) impacts positively in the decision of obtaining the consumer loan that applied for. We consider the correlation significant at the 0.01level (2-tailed). However it is very important to mention that the correlation coefficient does not control for other factors' effect and therefore we should conduct further examination to explore these relationships.

Table 4.1.7 Characteristics of borrowers

Variables	Good loan	Bad loan	Total
Age 18-40	2063	525	2588
40-60	3408	646	4054
> 60	1281	172	1453
Loan duration ≤ 12	360	15	375
> 12	6392	1328	7720
Income $\leq 15000€$	4040	883	4923
$> 15000€$	2712	460	3172
Oth.BankTr. Yes	5457	990	6447
No	1295	353	1648
Nationality Greek	6618	1305	7923
Other	134	38	172
Marital St. Married	4175	784	4959
Divorced	1881	430	2311
Single	696	129	825
Oth.loans Yes	1184	455	1639
No	5568	888	6456
Postal code Big city	3974	757	4731
Other city	2778	586	3364
Resid.type Tenant	1024	272	1296
Owner	5728	1071	6799
Gender Female	2529	455	2984
Male	4223	888	5111
Underwriter Yes	677	177	854
No	6075	1166	7241

In the Table 4.1.7, we can see the performance of the accepted loans in good and bad loans. Moreover we can see how the characteristics of borrowers are classified into these two categories. Taking into account the age of the borrowers we notice that people in the group of 40-60 years old performed better as far as repayment is concerned (18.95% default rate). Surprisingly, the borrowers who apply for short-term loans fulfilled their obligations to a greater extent than those who take longer duration loans. The default rate of short term loans is 4.16% whereas the default rate of long terms loans is 20.8%. Furthermore, we look at the annual income of the borrowers and note that there is a slightly difference between the proportion of good loans whether they have high income or not. The married borrowers also have approximately the same default rate with single borrowers (18%) whereas the divorced of them have bigger default rate (about 23%). At last we observe that women failed in a smaller extent to repay their loan in comparison with men.

Taking a look in the correlation matrix, in the Appendix B, we notice that the variables banking transactions, loan duration, other loans and marital status are positively correlated with the not default loan (loanstatus2) in 0.01 level of significance (2-tailed) though the income, age and underwriter are negatively correlated with the event of not default a consumer loan. We have to clarify again that there is a need of investigating the relationship of the factors in more depth in order to have a better and more correct view of their influence.

4.2 Probability of Obtaining a Loan through years 2007-2009

Does the probability of granting a loan change through years? In this section our goal is to examine the answer of this question. Firstly we separate the full data sample by year into three subsamples. The first subsample is consisted of all applications having been made in 2007, the second one respectively refers to applications having been made in 2008 and the third one to those made in 2009. It is important to have a sight of how the variables affect the probability of obtaining a loan in each period taking for granting that in periods of financial distress the financial decisions change. Thus it is worthwhile to have a view of how the financial institutions handle the critical decision of giving consumer loans and if something in their strategy changes through these years.

Table 4.2.1 Binary probit model 2007

Method: ML - Binary Probit (Quadratic hill climbing)
Sample: 1 5590
Included observations: 5590
Convergence achieved after 31 iterations
Covariance matrix computed using second derivatives

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	6.867453	5274.69	0.001302	0.999
AGE1	-0.30056	0.06003	-5.0068	0.0000*
AGE2	-0.19497	0.056829	-3.43074	0.0006*
BANKTR	0.409294	0.038833	10.53971	0.0000*
BIGCITY	0.062392	0.037958	1.643707	0.1002
MARRIED	0.152368	0.039908	3.818014	0.0001*
MALE	-0.01069	0.039003	-0.27404	0.7841
HIGHINCO	0.022557	0.042082	0.536019	0.5919
LOANDUR2	-0.54957	0.097511	-5.63598	0.0000*
NATIONG	0.496486	0.099606	4.98449	0.0000*
OTHERLOA	-6.81984	5274.69	-0.00129	0.999
OWNER	0.101358	0.049721	2.038511	0.0415*
UNDERWR	0.128946	0.055372	2.328707	0.0199*

Mean dependent var	0.681395	S.D. dependent var	0.465977
S.E. of regression	0.440627	Akaike info criterion	1.120952
Sum squared resid	1082.787	Schwarz criterion	1.136367
Log likelihood	-3120.06	Hannan-Quinn criter.	1.126324
Restr. log likelihood	-3498.3	Avg. log likelihood	-0.55815
LR statistic (12 df)	756.472	McFadden R-squared	0.10812
Probability(LR stat)	0		
Obs with Dep=0	1781	Total obs	5590
Obs with Dep=1	3809		

* Significance level 0.05 (2-tailed).

Table 4.2.1 provides us with the results of the first regression. We run the probit regression model for the year 2007. The subsample consists of 5590 observations which are all the applications submitted this year. The results were ready after 31 iterations that were required for convergence and the second derivatives method which was used in order to compute the coefficient covariance matrix. The McFadden R-squared, which is the likelihood ratio index, equals to 0.10812, the probability (LR stat.) is zero with 12 degrees of freedom. The number that indicates the degrees of freedom actually is the number of restrictions under the test.

In general, the probit model assumes that the higher the value of "probit" the greater the probability to obtain a loan. Thus the positive coefficient values imply increase in the probability of obtaining a consumer loan relative to the applicants' characteristics and the negative coefficient values imply decrease in the probability correspondingly.

Looking at the table 4.2.1 we observe that eight out of twelve influencing factors are statistically significant at significant level 5%. The regressors variables: age1, age2, bank transactions, married, loan duration2, nationality, owner and underwriter are statistically significant factors because their probability is less than p=0.05. More specifically, for the variable age1 we reject the null hypothesis that its coefficient is equal to zero and we deduce that it is statistically significantly different than zero with value -0.30056. For the age2 we also reject the null hypothesis and we accept that its coefficient is equal to -0.19497. Similarly for the variable loan duration2 the value of the coefficient is equal to -0.54957, for the bank transactions is 0.409292, for the married is 0.152360, for the owner is 0.496486, for the nationality is 0.101358 and for the underwriter is 0.128946. As we can see the variables age1, age2 and loan duration2 negatively impact the probability of obtaining a loan whereas the variables bank transactions, married, owner, nationality, and underwriter impact positively.

For example a person who applies for a consumer loan at the age group of 18-40 or 40-60 is less likely to obtain a loan particularly if one is in the first group. This is because the older the applicant might be the more steady income might have and more responsible might be with his/her financial obligations. Similarly, the bigger the length of a loan in demand the less the probability to obtain the loan.

On the other hand if the applicant has other relationships with the bank (such as deposits, credit cards) he/she has higher probability to obtain a loan. This indicates that financial institutions prefer to provide loans to their customers. If the potential borrower is married, it is more likely to take a loan as the marriage clarifies a responsible person who can commit to someone. Moreover, being Greek the applicant is more possibly to obtain a loan. Banks are less hesitant to provide a loan to a person who has Greek nationality as it may seem a safer choice due to the obvious commitment with Greece and especially the city of his/her residence. An applicant that has a house on his/her ownership indicates that he/she has financial wealth. Finally it seems, although it was unexpectable, that if a potential borrower does not have an underwriter (underwr=no underwriter, wunderwr=with underwriter) he/she has higher probability to obtain a loan. However the other four variables (high income, male, big city and other loans) are statistically insignificant in explaining the banks' lending decision in this study.

Table 4.2.2 Binary probit 2008

Method: ML - Binary Probit (Quadratic hill climbing)
Sample: 1 4579
Included observations: 4579
Convergence achieved after 32 iterations
Covariance matrix computed using second derivatives

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	6.936279	5854.587	0.001185	0.9991
AGE1	-0.28814	0.065357	-4.40872	0.0000*
AGE2	-0.21215	0.060935	-3.48154	0.0005*
BANKTR	0.465578	0.047233	9.856997	0.0000*
BIGCITY	0.021734	0.041036	0.529632	0.5964
MARRIED	0.094136	0.04345	2.166561	0.0303*
MALE	0.034917	0.042251	0.826429	0.4086
HIGHINCO	0.109528	0.044376	2.468156	0.0136*
LOANDUR2	-0.63648	0.121257	-5.24904	0.0000*
NATIONG	0.280234	0.108397	2.585254	0.0097*
OTHERLOA	-7.18296	5854.587	-0.00123	0.999
OWNER	0.101623	0.054045	1.880349	0.0601
UNDERWR	0.219128	0.06063	3.614177	0.0003*

Mean dependent var	0.537454	S.D. dependent var	0.49865
S.E. of regression	0.456667	Akaike info criterion	1.177801
Sum squared resid	952.2135	Schwarz criterion	1.196054
Log likelihood	-2683.58	Hannan-Quinn criter.	1.184227
Restr. log likelihood	-3161.06	Avg. log likelihood	-0.58606
LR statistic (12 df)	954.9729	McFadden R-squared	0.151053
Probability(LR stat)	0		
Obs with Dep=0	2118	Total obs	4579
	2461		

* Significance level 0.05 (2-tailed).

Table 4.2.2 shows the results of the probit model for the time period 2008. The volume of this subsample is 4579. The results were ready after 32iterations that were required for convergence and the second derivatives method which was used in order to compute the coefficient covariance matrix. The McFadden R-squared equals to 0.151053, the probability (LR stat.) is zero with 12 degrees of freedom. The number that indicates the degrees of freedom is the number of restrictions under the test.

We notice that eight out of twelve influencing factors are statistically significant at significant level 5%. The regressors variables: age1, age2, bank transactions, married, highincome, loan duration2, nationality and underwriter are statistically significant factors because their probability is less than p=0.05. More specifically, for the variable age1 we reject the null hypothesis that its coefficient is equal to zero and we deduce that it is statistically significantly different than zero with value -0.28814. For the age2 we also reject the null hypothesis and we accept that its coefficient is equal to -0.21215. Similarly for the variable loan duration2 the value of the coefficient is equal to -0.63648, for the bank transactions is 0.465578, for the married is 0.094136, for the high income is 0.109528, for the nationality is 0.280234 and for the underwriter is 0.219128.The variables age1, age2 and loan duration2 negatively impact the probability of obtaining a loan whereas the variables bank transactions, married, high income, nationality, underwriter impact positively.

The characteristic in the subsample of year 2008 that is also statistically significant in the banking decision to give a consumer loan is the high income. High income indicates financial wealth and therefore it constitutes an important criterion for the lending decision. On the other hand, in 2008 if a respondent has a home to his/her ownership it is statistically insignificant with probability 0.0601>0.05.

Table 4.2.3 Binary probit 2009

Method: ML - Binary Probit (Quadratic hill climbing)

Sample: 1 4569

Included observations: 4569

Convergence achieved after 35 iterations

Covariance matrix computed using second derivatives

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	7.098828	6592.015	0.001077	0.9991
AGE1	-0.35814	0.064927	-5.51598	0.0000*
AGE2	-0.21262	0.061567	-3.45353	0.0006*
BANKTR	0.470358	0.054833	8.578065	0.0000*
BIGCITY	0.048336	0.04288	1.127252	0.2596
MARRIED	0.067979	0.046705	1.455498	0.1455
MALE	-0.06215	0.044445	-1.39846	0.162
HIGHINCO	0.099868	0.044345	2.25206	0.0243*
LOANDUR2	-0.52528	0.12042	-4.36211	0.0000*
NATIONG	0.200511	0.139229	1.44015	0.1498
OTHERLOA	-7.7112	6592.015	-0.00117	0.9991
OWNER	0.082959	0.062168	1.334422	0.1821
UNDERWR	0.199268	0.065448	3.044648	0.0023*

Mean dependent var	0.399431	S.D. dependent var	0.489835
S.E. of regression	0.428586	Akaike info criterion	1.069384
Sum squared resid	836.6899	Schwarz criterion	1.089077
Log likelihood	-2429.01	Hannan-Quinn criter.	1.076318
Restr. log likelihood	-3073.93	Avg. log likelihood	-0.53163
LR statistic (12 df)	1289.852	McFadden R-squared	0.209805
Probability(LR stat)	0		
Obs with Dep=0	2744	Total obs	4569
Obs with Dep=1	1825		

* Significance level 0.05 (2-tailed).

Finally, in table 4.2.3 we see the results of the third regression, having run the probit regression model for the year 2009. The subsample consists of 4569 observations. The results were ready after 35 iterations that were required for convergence and the second derivatives method which was used in order to compute the coefficient covariance matrix. The McFadden R-squared, which is the likelihood ratio index, equals to

0.2098005, the probability (LR stat.) is zero with 12 degrees of freedom. The number that indicates the degrees of freedom is the number of restrictions under the test.

The table 4.2.3 shows that in the period of 2009 only six out of twelve influencing factors are statistically significant at significant level 5%. The regressors variables: age1, age2, bank transactions, highincome, loan duration2 and underwriter are statistically significant factors having probability which is less than p=0.05. More specifically, for the variable age1 we reject the null hypothesis that its coefficient is equal to zero and we deduce that it is statistically significantly different than zero with value -0.35214. For the age2 we also reject the null hypothesis and we accept that its coefficient is equal to -0.21262. Similarly for the variable loan duration is the value of the coefficient is equal to - 0.52528, for the high income is 0.099868, for the bank transactions is 0.470358 and for the underwriter is 0.199268.The variables age1, age2 and loan duration2 negatively impact the probability of obtaining a loan whereas the variables bank transactions, high income, underwriter impact positively.

We observe that the characteristics, having been statistically significant in the lending decision in the two previous years, Greek nationality and ownership of a house are now statistically insignificant in the model. Hence for the provision of a consumer loan in the period 2009, there were very specific and less characteristics that influence the banks' decision.

Taking into consideration all the above results we conclude that the criteria for taking a consumer loan actually changed through time. It is obvious that as the financial conditions became more difficult, the factors that influence the lending decision were related to the financial wealth of the respondent. For instance we observe that the income of an applicant in 2007 was not a statistically significant factor and so it did not influence the probability of granting a loan. However, in 2008 as well as in 2009 the income was one of the factors that were statistically significant. On the other hand, applicants' characteristics, such as their nationality or marriage or even ownership of house, were statistically significant factors in 2007 but as the financial conditions were worsening these characteristics did not influence the banking lending decision. Furthermore, we notice that the important factors in the decision of providing a loan were becoming less in number, as the financial institutions started to focus in specific characteristics of the potential borrowers.

All these indicate that the banks, in order to ensure themselves about the solvency of the applicant in periods with liquidity shortage (minimize the default risk), attended strictly to characteristics with financial implication. In addition they wanted to keep their clientele therefore they financed their customers more easily and they preferred to provide liquidity for a short period of time as well.

4.3 Probability of obtaining a loan and probability of default

In this section we will try to examine the effects of the borrower's characteristics in the decision of granting consumer loans and their performance for the period 2007-2009. We used a bivariate probit model in order to see if the financial institutions' lending policies are compatible with default risk minimization or not and compare our results with the inferences of other studies with the same researching view.

Table 4.3.1 Bivariate probit MLE of $\boldsymbol{\beta_1}$ and $\boldsymbol{\beta_2}$

Variables	P(obtain a loan)			P(good loan)		
	$\widehat{\beta_1}$	St. error	t-stat.	$\widehat{\beta_2}$	St. error	t-stat.
INERCEPT	6.026333	0.041500	145.21	0.568598	0.037133	15.31
FEMALE	-0.001170	0.023437	-0.05	0.032492	0.008661	3.75*
MARRIED	0.131105	0.025314	5.18*	0.008812	0.009626	0.92
DIVORCED	-0.015739	0.049115	-0.32	-0.022695	0.018429	-1.23
AGE1	-0.309629	0.035658	-8.68*	-0.093693	0.012889	-7.27*
AGE2	-0.189582	0.031583	-5.68 *	-0.046414	0.011405	-4.07*
BIGCITY	0.054674	0.022675	2.41*	0.019125	0.008425	2.27*
LOWINCOM	0.008756	0.024239	0.36	-0.028119	0.008811	-3.19*
OWNER	0.067367	0.030681	2.20*	0.048710	0.011434	4.26*
BANKTR	0.320849	0.025169	12.75*	0.080432	0.010619	7.57*
LOANDUR1	0.593931	0.062037	9.57*	0.138305	0.020415	6.77 *
NATIONG	0.335610	0.063765	5.26*	0.044858	0.029266	1.53
UNDERWR	0.190437	0.033598	5.67*	0.039761	0.013802	2.88*
OTHERL	-6.747611	0.041500	-162.59 *	0.072687	0.016892	4.30*
Sigma σ	0.369589		0.003729		99.11	
Rho ρ	0.259840		0.045344		5.73*	

*Significance level 0.05 (2-tailed).

Table 4.3.1 shows the parameters, the standard errors and the t-statistics for the probability of obtaining a loan and the probability of a good loan (not defaulted). The critical values are 1.645, 1.96 and 2.575 for the 10, 5, and 1 percent significance levels and the sample has no missing values. The number of observations is 14738, the number of endogenous variables is 2 and the number of iterations is 53 in order the algorithm to converge.

The statistically significant variables, at significance level 5%, for the probability of obtaining a loan are bank transactions, loan duration1, age1, age2, underwriter, Greek nationality, married, big city, owner and other loans with t-statistics equal to 12.75, 9.57, -8.68, -5.68, 5.67, 5.26, 5.18, 2.41, 2.20 and -162.59 respectively. The variables, female, divorced and low income, are statistically insignificant in the probability of obtaining a loan in this study.

The most important factors in the probability of obtaining a loan are the variables loan duration1, nationality Greek and bank transactions participate to the model with coefficients 0.593931, 0.33561 and 0.320849 respectively. These variables affect positively the banks' lending decision. As we notice the banks support their clientele since it is very important to remain competitive in the fierce banking industry. Another reason is that they have a more detailed view of the financial conditions and behaviors of customers who had previously transactions with them. Furthermore, it seems that short term loans are more attractive to banks' decision, as the short terms loans ensure a faster repayment of the loan, which favors the increase of liquidity and in a matter, ensures their capital during a difficult financial period. Also a Greek citizen appears a preferable choice by the financial institutions due to the obvious commitment with Greece and especially the city of his/her residence. The absence of bonds with a country for a foreigner may constitute a negative factor in the decision of being granted a loan as it is widely known that in many cases foreigners abandon the country where they had temporarily lived, leaving behind their debts.

The variables with significantly smaller probability of being granted a loan are underwriter, married, owner, big city, other loans, age1 and age2 participate to the model with coefficients 0.190437, 0.131105, 0.067367, 0.054674,-6.747611, -0.309629 and -0.189582 respectively. Consequently, if the applicant is married he/she has higher probability to obtain a loan as it is considered to be more responsible to his/her financial obligations. Furthermore, if the applicant has a house of his/her own, he/she has better chance to take a loan. A way to interpret this sign of the parameter would be that people who have not monthly financial obligations, like paying a rent for their house, are more responsible and able to pay their loan back and also it seems to have financial wealth. People who live in one of the two big cities of Greece (Athens and Thessaloniki) attributed positively the banks' lending decision as they might have bigger income in comparison with people who live in other cities of the country. Surprisingly, if an applicant has no underwriter, to ensure that he/she pays back the loan in case of applicant's default to pay, seems to have more possibilities to obtain a consumer loan. Rather than reasoning we also observe that not having other loans to pay worsens applicants' chances to obtain a loan. Finally, the results show that the younger the applicant the smaller the chance to obtain a loan. As we expected this factor affects negatively the probability of granting a loan since the majority of young people usually does not have stable job or steady income.

As far as the decision of a loan to be good (not defaulted) is concerned, the seventh column of the table 4.3.1 indicates that the statistically significant variables, at significance level 5%, are the female, age1, age2, big city, low income, owner, bank transactions, loan duration1, underwriter and with no other loans with t-statistics equal to 3.75, -7.27, -4.07, 2.27, -3.19, 4.26, 7.57, 6.77, 2.88 and 4.30 respectively.

The variables that affect the most the probability of a good loan are loan duration1 and other loans participate to the model with coefficients 0.138305 and 0.072687 respectively. The other variables that affect positively the event of a good loan are owner, female, big city, underwriter and bank transactions with coefficients, 0.011434, 0.032492, 0.019125, 0.039761 and 0.010619 respectively, whereas the variables age1, age2 and low income and affect negatively the event of a good loan participate to the model with coefficients -0.093693, -0.046414 and -0.028119 respectively.

As we can see, the short term loans constitute a positive weight in the probability of not defaulting. This can be explained by the fact that a short term can be paid back more rapidly. More specifically it is less possible expeditious changes to occur in a short period of time and especially during a financial distress where changes' frequency levels up. According to the results the borrowers that were already customers to a bank, having other bank transactions (such as deposit, credit cards) seem to have better performance. As we expected the factors owner and big city influence positively the probability of not defaulting as these characteristics are related to a financial wealth and responsible individual. Similar arguments could be applied to female borrowers.

Furthermore, the existence of an underwriter creates the insurance and in some cases the illusion that the borrower is capable of repaying his/her loan whereas in fact it is possible only to comprise the condition demanded to be granted it. For this reason the absence of an underwriter seems to give a better view about the financial reliability of the borrower which is compatible with our study's results. Finally, a borrower with no other financial obligations to pay (no other loans) has higher probability to meet his/her debts which is verified in our case.

The most important factor, from the statistically significant negative ones, is that the younger borrowers seem to have smaller chances to repay their loans probably due to their unsteady income or their compulsive behavior. Furthermore and not at all surprisingly borrowers with low annual income (lower than 15000€) tend to have difficulties to pay their loans which is consistent with common sense especially in period with financial distress.

The last parameter that we have to review is the correlation coefficient. The correlation coefficient is statistically significant with t-statistic equals to 5.73. Thus the selection bias is a big problem in the estimation of probability of default which verifies our choice to use the bivariate model in this study. The value of the correlation coefficient is 0.259840 that is to say the loan granting equation is positively correlated with the defaulting equation. In particular this suggests that unexplained tendencies to extend credit are actually associated with higher frequencies of not default in our sample (Boyes 1989).

Taking all these into account we notice that there exist some differences in the statistically significant factors between the banks' decision to provide a loan and the event of not defaulting a loan. More specifically the variables married and nationg were important in the first equation whereas they were not significant in the default equation. Moreover, although the variables; female and low income were not important in the granting equation, they were in the second one. This witnesses some inefficient use of information in the evaluation of applicants.

On the other hand we notice that all the statistically significant variables in both equations have the same sign that is the variables that increase (decrease) the probability of obtaining a loan also increase (decrease) the probability of a good loan. This leads us to infer that the financial institutions' lending policies are actually compatible with default risk minimization. A lending policy that attempts to seek out accounts, that they have low default probabilities, and has as overall motive the profit maximization. An exception is comprised the variable other loans which has different sign in the two equations. This could be explained by the significantly large observations in our data sample of applicants with no other loans and therefore we will not take it into consideration.

In a period of financial distress the major concern of financial institutions is to reduce their credit exposure which signifies reduction of the number of loans granted. Evidently, this extraordinary policy of banks made the criteria to grant a loan to become much more strict and consistent in a way that the probability of default a loan to reduce. A lending policy that attempts to seek out accounts, that they have low default probabilities, and has as overall motive the profit maximization. This policy will keep the financial institutions solvent and able to offer liquidity in a period which is mainly characterized by liquidity shortage, financial uncertainty, fear reaction and consequently rapidly financial changes.

Finally, we compare our results with the studies of Boyes' et al. (1989), Jacobson's et al. (2003) and Roszbach's (2004). We should point out that Boyes' et al. (1989) study contains a sample of 4632 applications that were processed between 1977 and 1980, and Jacobson's et al. (2003) and Roszbach's (2004) studies contain 13338 applications that were processed between September 1994 and August 1995. Comparing our results with these studies we observe that our research based on the data sample of a Greek bank does not confirm the conclusions in Boyes' et al. (1989), Jacobson's et al. (2003) and Roszbach's (2004) studies that banks do not appear to be minimizing default risk. On the contrary our results verify that banks follow a policy of default risk minimization. Subsequently, our research does not confirm that non-systematic tendencies to grant loans are associated with greater default risk either, as the correlation coefficient in our model is equal to 0.259840. Secondly, we found that some variables that are statistically significant in only one equation which testifies some inefficient use of information in the evaluation of applicants and is a fact as well in Boyes' et al. (1989), Jacobson's et al. (2003) and Roszbach's (2004) studies.

CHAPTER 5

SUMMARY AND CONCLUSIONS

5.1 Summary

Consumer credit constitutes a significant instrument in the financial planning of households and simultaneously is among the most profitable investments in lenders' asset portfolios. Despite the wide variety of banking services, lending to the public constitutes the core of the income of banks and other lending institutions. However, increases in the amount of loans also causes increases in the number of defaulted loans. Therefore, the primary problem of any lender is to differentiate between “low risk” and “high risk” debtors prior to granting credit. The main method used in assessing credit risks is the credit scoring analysis, which consists of the valuation of the financial history and financial statements of the applicant credit background. The aim of credit analysis is to evaluate the borrower’s probability of repayment, to determine the financial strength of the borrower, and to minimize the risk of non-payment to a cut off level. Credit scoring, which is broadly applied in consumer lending, is a statistical approach to predicting the probability that a credit applicant will default or become delinquent. Credit scoring system is a computerized process producing a score according to various relevant quantitative and qualitative characteristics of the borrower to assess loan applications. The overall idea of credit scoring is based on the statistical probabilities, or in other words the combinations of the borrowers’ characteristics differentiating good from bad. In such way a score is generated in order to act as an estimate of the risk level of each new loan when the banks’ officers have decided whether to accept the loans or not.

In this study we developed an optimal specification of the credit scoring model, and we focused on an analysis of the determinants of obtaining a consumer loan as well as the probability of a loan to default in Greece for the period 2007-2009. The onset of the recent financial crisis that erupted in September, 2008 caused an extraordinary reduction of business and consumer loans, mortgages and securitized products. Nevertheless, the financial results of the crisis in the Greek Banking sector became observable at the end of 2009. After a variety of government actions to promote the liquidity and solvency of the financial sector, almost every bank in Greece reacted concurrently similarly by decreasing dramatically the amount of loans granted. Obviously, this policy of all banks entailed that henceforth the criteria to grant a loan would be much more strict and consistent. Therefore it was worthwhile and interesting to see what was happening in the special case of Greek banking system during the latest global financial distress.

5.2 Results of research

Initially we examined, in our research if the probability of granting a consumer loan changes through years. Having in mind that during a period of financial distress the financial decisions change rapidly, we considered the analysis of the variables' affection in the probability of granting a loan per year very interesting. Separating the full data sample into three different time periods (2007, 2008, 2009), we performed a probit model for each sub-sample and indeed the empirical results suggest that the probability of obtaining a loan does change through time.

Our results indicate that during the different lending period, there are different numbers of variables which impact banks' decision to grant a consumer loan. More specifically as the financial conditions were becoming more difficult, the factors that influence the lending decision were related to the financial wealth of the respondent. For instance, the annual income of an applicant in 2007 was not a statistically significant factor and so it did not influence the probability of granting a loan whereas in 2008 the income was one of the factors that were statistically significant and in 2009 as well. Then again some factors such as nationality were statistically significant but as the financial conditions were worsening stopped to influence the probability of granting a loan, Furthermore, the important factors in the decision of providing a loan were becoming less in number, as the financial institutions started to focus in specific characteristics of the potential borrowers.

All these indicate that the banks, in order to ensure themselves about the solvency of the applicant in periods with liquidity shortage (minimize the default risk), attended strictly to characteristics with financial implication. In addition they wanted to keep their clientele therefore they financed their customers more easily and they preferred to provide liquidity for a short period of time too. Consequently the results show that it is important to review the borrowers' creditworthiness periodically, as the changes in economic condition could affect the criteria on lending policy, which is congruent with McAllister's et al. (1994) findings.

Afterwards we examined the effects of the borrower's characteristics in the decision of granting consumer loans and their performance for the period 2007-2009. We used a bivariate probit model in order to see if the financial institutions' lending policies are compatible with default risk minimization or not and compare our results with the inferences of other studies with the same researching view.

The results in this research show that there exist some differences in the statistically significant factors between the banks' decision to provide a loan and the event of a good loan. More specifically, although the variables, married and nationg, were important in the obtaining equation they were not in the default equation. Although the variables, female and low income, were not important in the granting equation they were in the second one. This testifies some inadequate use of information in the evaluation of applicants which is a fact as well in Boyes' et al. (1989), Jacobson's et al, (2003) and Roszbach's (2004) studies.

Furthermore according to the results, the variables that increase (decrease) the probability of obtaining a loan also increase (decrease) the probability of a good loan, which indicate that the financial institutions' lending policies are actually compatible with default risk minimization. This policy will keep the financial institutions solvent and able to offer liquidity in a period which is mainly characterized by liquidity shortage and financial uncertainty, therefore banks tend to follow it. Evidently the results of our research does not confirm the conclusions in Boyes' et al. (1989), Jacobson's et al. (2003) and Roszbach's (2004) studies that banks do not appear to be minimizing default risk.

5.3 Limitations of research

It is important to mention that there are some limitations in our research which are related to the data set, the estimation techniques and the variables we used. To begin with, we only used the data set from one commercial bank of Greece. The results of our research have limited implications and robustness and thus do not depict the lending behavior of all commercial banks in the consumer lending in Greece.

Furthermore we did not use in this research data information about the amount of consumer loans therefore we cannot make inferences about the influence of loan size to the banks' lending decision and to default risk. As a result, even if banks are risk averse, we cannot examine their relation between risk minimization and profit maximization.

As far as model is concerned, the potential exposure to future risk is not taking into account as the credit scoring is usually a static model in nature. Also the model does not account for the time to delinquency or default therefore it might, for instance, lose its predictive power during a recession if the characteristics entered into the model or the underlying customer population is sensitive to the economic cycle.

In addition, we should point out that there are observations (such as other loans) with very small variation, resulting to an extremely high coefficient, which cannot be explained by logical inferences.

5.4 Recommendations for future research

We could improve the research results and generalize the research findings, by using a larger and more extensive data from other commercial banks. Furthermore there are a number of characteristics that can be added to the models in order to enrich the models' performance. For instance we could include variables such as occupation of applicants or even better educational level of applicants, changes in applicants' annual income, the amount of credit and the time that a borrower defaults to pay.

It has been suggested that bigger loans are more preferable because they offer higher expected earnings (Jacobson et al. 2003) thus it is very important to examine whether the lending policy is compatible with profit maximization or not. Having in mind our findings, that banks' lending policy is compatible with default risk minimization, it will be very interesting to further investigate if it is compatible to profit maximization as well. If it is the case, we could also examine the relation between risk minimization and profit maximization. Moreover from a profit or utility maximizing perspective, it is not only important to know if but also when a loan will default. Traditional credit scoring models predict default risk and therefore fail to take into account the multi-period nature of loans contracts so a survival analysis could allow a more realistic evaluation of the return on a loan (Roszbach 2004).

BIBLIOGRAPHY

Altman E.I., Avery R. B., Eisenbeis R. A., Sinkey J.F. "Application of Classification Techniques in Business, Banking and Finance." *JAI Press*, 1981.

Avery R.B., Calem P.S., Canner G.B., *"Consumer Credit Scoring: Do Situational Circumstances Matter?"* Switzerland: Bank for International Settlements, 2004.

Boyes J. William, Hoffman L. Dennis, Low A .Stuart. "An Econometric Analysis of the

Bank Credit Scoring Problem." *Journal of Econometrics 40*, 1989: 3-14.

Boyle M., Crook J. N., Hamilton R., Thomas L. C. (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, 1992: 75-90.

Carling K., Jacobson T., Roszbach K., "Dormancy Risk and Expected Profits of Consumer Loans." *Journal of Banking and Finance 25* (Elsevier Science B.V.), 2001: 717-739.

Chang S., Feelders A.J and G.J. McLachlan. "Mining in the presence of selectivity." *Knowledge Discovery and Data Mining*, 1998: 199-203.

Coval J., Shumway T. "Do behavioural biases affect prices?" University of Michigan, Working Paper, 2000.

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

Crook J.N, Hamilton R. and Thomas L.C. "A comparison of a credit scoring model with a credit performance." *The service industries journal*, 1992.

Dinh T. H. T, Kleimeier S. "A credit scoring model for Vietnam's retail banking market." *Intrnational review of financial analysis*, 2007: 471-495.

Dunn D.J., Frey T.L. . "Discriminant analysis of loans for cash grains farms." *Agricultural Finance Review*, 1976: 60-66.

Games F., Hill M. "Consumer credit scoring models:does the underlying probability distribution matters?" *CEMAPRE.* 2000.

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

Greene W. H. *Econometric Analysis,5th edition.* Upper Saddle River: NJ: Prentice Hall, 2002.

Greene W. H. "Sample selection in credit-scoring models." *Japan and the World Economy*, 1998: 299-316.

Hand D.J., Henley W.E.. "Statistical classification methods in consumer credit scoring : a review." *Journal of the royal statistical society:series A(statistic in society)*, 1997: 525-541.

Hasan I., Zazzara C. "Pricing Risky Bank Loans in the New Basel 2 Environment." *Journal of Banking Regulation, 7(3-4)*, 2006: 243-267.

Hayden, E. "Are credit scoring models sensitive to different default definitions? evidence from the austrian market." *SSRN Working Paper*, 2003.

Henley W.E., Hand D.J. *"A k - Nearest - Neighbour Classifier for Assessing Consumer Credit Risk",* Royal Statistical Society, UK: The Statistician, 1996.

Huyen D.T. Thanh. "Credit scoring for vietnam's retail banking market :implementation and implications for transactional versus relationship lending." *Maastricht research school of Economics of Technology and Organi-*, 2006.

Ivashina Victoria, Scharfstein David. "Bank Lending During the Financial Crisis of 2008." Harvard Business School, 2008.

Jacobson Tor, Roszbach Kasper. "Bank lending policy, credit scoring and Value-at-Risk." *Journal of Banking anf Finance*, 2003: 615-633.

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

Kočenda Evžen, Vojtek Martin. *"Default Predictors and Credit Scoring Model".* CATEGORY 12: EMPIRICAL AND THEORETICAL METHODS, CESIFO WORKING PAPER NO. 2862, 2009.

Lewis, E. M., "An Introduction to Credit Scoring", (2nd Eds.), Fair, Isaac and Co., 1992
Luftburrow J., Barry P.J., Dixon B.L., "Credit scoring for farm loan pricing ." *Arcicultural finance review*, 1984: 8-14.

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

Mortensen T., Watt D. L., Leistriz F.L., "Predicting probability of a loan default." *Agricultural finance review*, 1988: 60-67.

Narain B., Thomas L.C., Crook J.N., Edelman D.B.. "Survival Analysis and the Credit Granting Decision in Credit Scoring and Credit Control." *Oxford University Press*, 1992: 109-122.

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

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

Rosenberg, E., Gleit A. “Quantitative methods in credit management: A survey”, *Operations Research*, 42, 1994: 589-613.

Roszbach K."Bank lending policy, credit scoring and the survival loans." *The review of Economics and Studies*, March 29, 2004.

Santos Silva J. M. C., Murteira J. M. R. "Estimation of Default Probabilities Using Incomplete Contract Data." 2000.

Schreiner, M. "Scoring arrears at a micro lender in Bolivia." *Journal of microfinance*, 2004: 65-88.

Schreiner, M. “The next breakthrough in micro credit?” CGAP Occasional Paper No. 7, 2003

Schwartz K., Menon L.R., Gilbert K.B. "Predicting bankruptcy for firms in." *Journal of Business, Finance and Accounting*, 1990: 161-171.

Steenackers A., Goovaerts M.J. “ *A Credit Scoring Model for Personal Loans.”* North Holland: Elsevier Science Publishers B.V., 1989.

Stepanova M., Thomas L.C. “*Survival Analysis Methods for Personal Loan Data.”* Operational Research Society, 2001.

Suits Daniel B. "Use of Dummy Variables in Regression Equations." *Journal of the American statistical association*, 1957: 548-551.

Thomas, L.C. "A survey of credit and behavioural scoring : Forecasting financial risk of lending to consumers." *Internation journal of forecasting*, 2000: 149-172.

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

Thomas L.C., Ho J., Scherer W. T., "Time Will Tell: Behavioural Scoring and the Dynamics of Consumer Credit Assessment." *IMA Journal of Management Mathematics*, 2001: 89-103.

Turvey, C.G. "Credit scoring for agricultural loans: a review with application." *Argicultural finance review*, 1991: 43-54.

Verbeek, Marno. *Modern Econometrics , 3rd edition.* Rotterdam: John Wiley & Sons, Ltd, 2008.

Wiginton, J.C. "A note on the comparison of logit and discriminant models of consumer credit behaviour." *Journal of financial and quantitative analysis*, 1980: 757-770.

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

APPENDIX A

Combining (3.5)-(4.7) and the independent variables, the likelihood function in equation (3.11) becomes:

$$l = \prod_{i=1}^{N} pr(y_i{}^* \leq 0)^{(1-y_i)} + \prod_{i=1}^{N} pr(y_i{}^* > 0, z_i{}^* \leq 0)^{y_i \times z_i}$$
$$+ \prod_{i=1}^{N} pr(y_i{}^* > 0, z_i{}^* > 0)^{y_i \times z_i} \qquad (A.1)$$

Substituting for (3.7), (A.1) implies the following loglikelihood function:

$$lnl = \sum_{i=1}^{N} (1 - y_i) \times \ln[pr(\varepsilon_i \leq -x_i{}'\beta_1)] + \sum_{i=1}^{N} y_i \times (1 - z_i)$$
$$\times \ln[pr(\varepsilon_i > -x_i{}'\beta_1 \cap u_i \leq -w_i'\beta_2)] + \sum_{i=1}^{N} y_i \times z_i$$
$$\times \ln[pr(\varepsilon_i > -x_i{}'\beta_1 \cap u_i \geq -w_i'\beta_2)] \ (A.2)$$

Because of the symmetry property of the bivariate normal distribution, the last line in (A.2) can be rewritten as:

$$pr(\varepsilon_i > -x_i{}'\beta_1 \cap u_i \geq -w_i'\beta_2) \leftrightarrow \Phi_2(x_i'\beta_1, w_i'\beta_2; \rho) \quad (A.3)$$

Since

$$\mathrm{pr}(\mathrm{y_i}^* > 0, \mathrm{z_i}^* \leq 0) = 1 - \mathrm{pr}(\mathrm{y_i}^* \leq 0) - \mathrm{pr}(\mathrm{y_i}^* > 0, \mathrm{z_i}^* > 0)$$

The loglikelihood function can be written as:

$$\ln = \sum_{i=1}^{N} (1 - y_i) \times \ln\,[1 - \Phi(x_i'\beta_1)]$$
$$+ \sum_{i=1}^{N} y_i \times (1 - z_i) \times \ln\{\Phi(x_i'\beta_1) - \Phi_2\,(x_i'\beta_1, w_i'\beta_2; \rho)\}$$
$$+ \sum_{i=1}^{N} y_i \times z_i \times \ln \Phi_2(x_i'\beta_1, w_i'\beta_2; \rho) \ (A.4)$$

APPENDIX B

Correlations

Statistics=Correlation Coefficient,Type=Spearman's rho

	loan status1	income	loan status2	gender	age	citizenship	bank transactions	marital status	loan duration	other loans	postal code	residence type	underwriter
loan status1	1.000	-.054**	.	.006	-.120**	.076**	.142**	.080**	.069**	-.320**	.028**	.017*	-.076**
income	-.054**	1.000	-.045**	-.126**	.155**	-.098**	-.075**	-.091**	-.002	.070**	-.084**	-.064**	.184**
loan status 2	.	-.045**	1.000	-.028*	-.079**	.022*	.066**	.041**	.075**	.150**	.019	-.008	-.038**
gender	.006	-.126**	-.028*	1.000	.007	.001	-.009	-.012	.034**	.011	-.061**	-.053**	-.075**
age	-.120**	.155**	-.079**	.007	1.000	-.077**	-.067**	-.349**	-.013	.072**	-.013	-.194**	.070**
nationality	.076**	-.098**	.022*	.001	-.077**	1.000	.047**	.030**	-.030**	-.046**	-.019*	-.172**	-.063**
bank transactions	.142**	-.075**	.066**	-.009	-.067**	.047**	1.000	.066**	.013	-.087**	-.010	.032**	-.065**
marital status	.080**	-.091**	.041**	-.012	-.349**	.030**	.066**	1.000	-.007	-.040**	-.050**	.171**	.018*
loan duration	.069**	-.002	.075**	.034**	-.013	-.030**	.013	-.007	1.000	.045**	.011	-.003	-.036**
other loans	-.320**	.070**	.150**	.011	.072**	-.046**	-.087**	-.040**	.045**	1.000	-.032**	-.026**	.042**
postal code	.028**	-.084**	.019	-.061**	-.013	-.019*	-.010	-.050**	.011	-.032**	1.000	.070**	-.075**
residence type	.017*	-.064**	-.008	-.053**	-.194**	-.172**	.032**	.171**	-.003	-.026**	.070**	1.000	-.028**
underwriter	-.076**	.184**	-.038**	-.075**	.070**	-.063**	-.065**	.018*	-.036**	.042**	-.075**	-.028**	1.000

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).