AN APPLICATION OF LOGIT AND PROBIT MODELS OVER THE DEFAULT PROBABILITIES OF RETAIL BANKING MORTGAGE CREDITS

ENGİN KARACULA
107621002

ISTANBUL BILGI UNIVERSITY
INSTITUTE OF SOCIAL SCIENCES
FINANCIAL ECONOMICS MASTER PROGRAM

ASST. PROF. DR. ORHAN ERDEM

2009

AN APPLICATION OF LOGIT AND PROBIT MODELS OVER THE DEFAULT PROBABILITIES OF RETAIL BANKING MORTGAGE CREDITS

ENGİN KARACULA
107621002

TEZ DANIŞMANININ ADI SOYADI (İMZASI) :

JÜRİ ÜYELERİNİN ADI SOYADI (İMZASI) :

JÜRİ ÜYELERİNİN ADI SOYADI (İMZASI) :

TEZİN ONAYLANDIĞI TARİH :

Toplam Sayfa Sayısı: 28

Anahtar Kelimeler (Türkçe)
1) Lojit Modeli
2) Probit Modeli
3) Banka Kredileri

Anahtar Kelimeler (İngilizce)
1) Logit Model
2) Probit Model
3) Bank Credits

ÖZET

Günümüzde tüm dünyada ev sahibi olmak önemli bir konu halini almış durumdadır ve bu evlerin finansmanı yeterli derecede önemlidir. Bu çalışma ev kredilerindeki batık oranlarını etkileyen parametreleri ortaya koymaya çalışmaktadır. Bu parametreler banka müşterilerinin kişisel verilerini ve ev kredilerinin tipik özelliklerini içermektedir. Bu çalışma Türkiye'de yerleşik bir bankanın verileri ile yapılmıştır. Çalışma kapsamında veri toplama aşamasından sonra ev kredilerinin özelliklerini temsil eden parametreler seçilmiş, E-Views vasıtasıyla kredilerin batma koşulunu belirleyen parametreler ortaya çıkarılmıştır. Sonuç olarak, ana amaç ev kredilerindeki batma durumunun profilini lojit ve probit modeller vasıtasıyla ortaya koymaktır.

ABSTRACT

Recently, housing has been one of the core issues all over the world and finance of mortgages also has a significant importance. This study tries to state parameters that affect the default rates of mortgage loans. These parameters include personal information about the customers of a bank and the characteristics of the mortgages and the loans. Study is made with a data which include a bank's customers resident in Turkey. In content of this study, after data gathering period, parameters standing out with the representing ability of a mortgage loan is selected and with these parameters and by the helping of E-views, significant values that highly determine the default case of mortgage loans are found. Consequently, the main objective is to find out the possible default profile of mortgage loans by using logit and probit estimation methods.

TABLE OF CONTENTS

1. Introduction

Recently, housing has been one of the core issues all over the world. As a developing economy, Turkey has rapid population growth and urbanization, finance of housing (mortgages) becomes an important topic. In addition, it is not hard to say that subprime-mortgage crisis does not trigger the global crisis which is faced by the whole finance sectors in the world. Mortgage finance is highly effective on all markets because the necessity of people to house for living will never end. Banks as financers in this situation have to make such decisions that while a mortgage-based credit is given, the person financed must be suitable with all parameters such as income level, and because of this credit transaction the sector must not be affected if this loan does not perform.

Housing is one of the main necessities and in this mean a house can be identified as physical place that meets the demand of sheltering. On the other hand, it can be evaluated as an investment tool which also has contribution to the economies of the countries and also an indicator of social development. Because mortgage-sector is one of the leading sectors, governments prepare regulations in order to determine the factors that will affect finance of the mortgage sector. This is important because in a developing economy, people must have the opportunity of finding finance in corporate ways.

From Banks' point of view, while financing a mortgage, there is wide range of indicators. In this manner, financed loans and profile of the customers are in a wide relationship. The credit limit that will be assigned is one of the important parameters. Loan limit that will be used for mortgage-financing is determined with so many parameters. Some of these parameters are income level, appraisal and expertise value of the mortgage and social status of the person. In this mean, LTV (loan to value) and PTI (payment to income) ratios are distinctive parameters.

While so many parameters are determining the assigned loan, in the same manner, these parameters also define the default probability of the loan. In this way, in this project logit and probit models will be used in order to find out which parameter is effective at which rate. Shortly, logit regression analysis is a uni/multivariate technique which allows for estimating the probability that an event occurs or not, by predicting a binary dependent outcome from a set of independent variables. On the other hand, in order to explain the behavior of a dichotomous dependent variable, a suitably chosen Cumulative Distribution Function (CDF). The logit model uses the cumulative logistic function. But this is not the only CDF that can be used. In some applications, the normal CDF has been found useful. The estimating model that emerges from the normal CDF is known as Probit Model. The main difference between logit and probit models is that logistic function has slightly flatter tails. Moreover, logit and probit models give similar results, but the estimates of parameters of the two models are not directly comparable.

The main objective of this project is to explain the default probability of a loan in terms of the loan parameters and the profile of the customer. We will study with a Bank's mortgage data without giving specific information specifying the Bank and the customers. Our data is between 2005-2008 and consists of 65.000 mortgage borrowers and their loans including time to maturity, payment to income ratio, credit open date, mortgage amount, marital status, loan to value ratio, credit limit, interest rate, installment amount, income, gender, expertise amount, education and age parameters. With binary dependent variable models, we aim to find out which parameters are significant to demonstrate the overall mortgage data with which coefficients. We will do this by eliminating the insignificant parameters by concerning p values less than 5%. At the end, the remaining parameters will represent the data.

In this framework, we will first mention about general characteristics of mortgage sector over the world. In this section, highly known mortgage markets' summaries will be given and a comparison will be made. The next section covers the logit and probit models theory. In this part, how logit and probit models are working will be summarized. Then, the conceptual background of the study and data description will be mentioned. The used data's characteristics will be stated and successively, logit and probit models will be applied to the data in order to find out the significant acting parameters. Consequently, at the last stage, the main results will be outlined and conclusion will be stated. The parameters' meanings in case of a positive and negative coefficient step will be also given.

2. Interpreting Mortgage Sector

A mortgage is the transfer of an interest in property (or the equivalent in law - a charge) to a lender as a security for a debt - usually a loan of money. While a mortgage in itself is not a debt, it is the lender's security for a debt. It is a transfer of an interest in land (or the equivalent) from the owner to the mortgage lender, on the condition that this interest will be returned to the owner when the terms of the mortgage have been satisfied or performed. In other words, the mortgage is a security for the loan that the lender makes to the borrower stated in Housing Finance Systems Principles and Examples.

In the same paper, it is stated that in many countries, it is normal for home purchases to be funded by a mortgage. Few individuals have enough savings or liquid funds to enable them to purchase property outright. In countries where the demand for home ownership is highest, strong domestic markets have developed, notably in Ireland, Spain, the United Kingdom, Australia and the United States.

A mortgage loan is a loan secured by real property through the use of a note which evidences the existence of the loan and the encumbrance of that realty through the granting of a mortgage which secures the loan. However, the word mortgage alone, in everyday usage, is most often used to mean mortgage loan.

A home buyer or builder can obtain financing (a loan) either to purchase or secure against the property from a financial institution, such as a bank, either directly or indirectly through intermediaries. Features of mortgage loans such as the size of the loan, maturity of the loan, interest rate, method of paying off the loan, and other characteristics can vary considerably.

According to Anglo-American property law, a mortgage occurs when an owner (usually of a fee simple interest in realty) pledges his interest as security or collateral for a loan. Therefore, a mortgage is an encumbrance on property just as an easement would be, but because most mortgages occur as a condition for new loan money, the word mortgage has become the generic term for a loan secured by such real property.

As with other types of loans, mortgages have an interest rate and are scheduled to amortize over a set period of time; typically 30 years. All types of real property can, and usually are, secured with a mortgage and bear an interest rate that is supposed to reflect the lender's risk.

Upon making a mortgage loan for purchase of a property, lenders usually require that the borrower make a downpayment, that is, contribute a portion of the cost of the property. This downpayment may be expressed as a portion of the value of the property (see below for a definition of this term). The loan to value ratio (or LTV) is the size of the loan against the value of the property. Therefore, a mortgage loan where the purchaser has made a downpayment of 20% has a loan to value ratio of 80%. For loans made against properties that the borrower already owns, the loan to value ratio will be imputed against the estimated value of the property.

The loan to value ratio is considered an important indicator of the riskiness of a mortgage loan: the higher the LTV, the higher the risk that the value of the property (in case of foreclosure) will be insufficient to cover the remaining principal of the loan.

In the U.S., the process by which a mortgage is secured by a borrower is called origination. This involves the borrower submitting an application and documentation related to his/her financial history and/or credit history to the underwriter. Many banks now offer "no-doc" or "low-doc" loans in which the borrower is required to submit only minimal financial information. These loans carry a higher interest rate and are available only to borrowers with excellent credit. Sometimes, a third party is involved, such as a mortgage broker. This entity takes the borrower's information and reviews a number of lenders, selecting the ones that will best meet the needs of the consumer.

Loans are often sold on the open market to larger investors by the originating mortgage company. Many of the guidelines that they follow are suited to satisfy investors. Some companies, called correspondent lenders, sell all or most of their closed loans to these investors, accepting some risks for issuing them. They often offer niche loans at higher prices that the investor does not wish to originate.

If the underwriter is not satisfied with the documentation provided by the borrower, additional documentation and conditions may be imposed, called stipulations. The meeting of such conditions can be a daunting experience for the consumer, but it is crucial for the lending institution to ensure the information being submitted is accurate and meets specific guidelines. This is done to give the lender a reasonable guarantee that the borrower can and will repay the loan. If a third party is involved in the loan, it will help the borrower to clear such conditions.

In a survey about Mortgage Loans Industry and Market revised in 2008, in the U.K., there are currently over 200 significant separate financial organizations supplying mortgage loans to house buyers in Britain. The major lenders include building societies, banks, specialized mortgage corporations, insurance companies, and pension funds. Over the years, the share of the new mortgage loans market held by building societies has declined. Between 1977 and 1987, it fell drastically from 96% to 66% while that of banks and other institutions rose from 3% to 36%. The banks and other institutions that made major inroads into the mortgage market during this period were helped by such factors as:

- relative managerial efficiency;
- advanced technology, organizational capabilities, and expertise in marketing;
- extensive branch networks; and
- capacities to tap cheaper international sources of funds for lending.

In CSO Financial Statistics and Building Societies Commission annual reports, UK building societies had succeeded in greatly slowing if not reversing the decline in their market share by the early 1990s. In 1990, the societies held over 60% of all mortgage loans but took over 75% of the new mortgage market – mainly at the expense of specialized mortgage loans corporations. Building societies also increased their share of the personal savings deposits market in the early 1990s at the expense of the banks – attracting 51% of this market in 1990 compared with 42% in 1989.

A paper named Investing for the Next Millenium done by Building Societies Research deals with a foundation which explains that in the five years 1987-1992, the building societies collectively outperformed the UK clearing banks on practically all the major growth and performance measures. The societies' share of the new mortgage loans market of 75% in 1990-91 was similar to the share level achieved in 1985. Profitability as measured by return on capital was 17.8% for the top 20 societies in 1991, compared with only 8.5% for the big four banks. Finally, bad debt provisions relative to advances were only 0.4% for the top 20 societies compared with 2.8% for the four banks.

The Mortgage Loans Industry and Market Survey states that though the building societies did subsequently recover a significant amount of the mortgage lending business lost to the banks, they still only had about two-thirds of the total market at the end of the 1980s. However, banks and building societies were by now becoming increasingly similar in terms of their structures and functions. When the Abbey National building society converted into a bank in 1989, this could be regarded either as a major diversification of a building society into retail banking – or as significantly increasing the presence of banks in the residential mortgage loans market. Research organization Industrial Systems Research has observed that trends towards the increased integration of the financial services sector have made comparison and analysis of the market shares of different types of institution increasingly problematical. It identifies as major factors making for consistently higher levels of growth and performance on the part of some mortgage lenders in the UK over the years:

- The introduction of new technologies, mergers, structural reorganization and the realization of economies of scale, and generally increased efficiency in production and marketing operations – insofar as these things enable lenders to reduce their costs and offer more price-competitive and innovative loans and savings products;
- Buoyant retail savings receipts, and reduced reliance on relatively expensive wholesale markets for funds (especially when interest rates generally are being maintained at high levels internationally);
- Lower levels of arrears, possessions, bad debts, and provisioning than competitors;
- Increased flexibility and earnings from secondary sources and activities as a result of political-legal deregulation;
- Being specialized or concentrating on traditional core, relatively profitable

mortgage lending and savings deposit operations.

Within the European Union, the Covered bonds market volume (covered bonds outstanding) amounted to about EUR 2 billion at year-end 2007 with Germany, Denmark, Spain, and France each having outstandings above 200.000 EUR million mentioned in Covered Bond Outstandings.

In German language, Pfandbriefe is the term applied. Pfandbrief-like securities have been introduced in more than 25 European countries – and in recent years also in the U.S. and other countries outside Europe – each with their own unique law and regulations. However, the diffusion of the concept differs: In 2000, the US institutions Fannie Mae and Freddie Mac together reached one per cent of the national population. Furthermore, 87 per cent of their purchased mortgages were granted to borrowers in metropolitan areas with higher income levels. In Europe, a wider market has been achieved: In Denmark, mortgage banks reached 35 per cent of the population in 2002, while the German Bausparkassen achieved widespread regional distribution and more than 30 per cent of the German population concluded a Bauspar contract (as of 2001).

3. Mortgage Default Probability-Previous Literature

Harrison, Noordewier and Yavas (2004) debated on their paper about conventional wisdom in the mortgage industry holds that loan-to-value (LTV) ratios are positively correlated with mortgage default rates. Their paper offers a theoretical signaling model of why the correlation between LTV ratios and default risk is contingent upon the default costs of the borrower. Specifically, their model proposes that when default costs are high there exist a separating equilibrium in which risky borrowers will self-select into lower LTV loans to reduce the probability of facing a costly default, while safe borrowers will self-select into higher LTV loans as a signal of their enhanced creditworthiness. This adverse selection process gives rise to the possibility of higher default probabilities for lower LTV loans. Conversely, when default costs are low the conventional result, in which risky borrowers select higher LTV loans than safe borrowers, is obtained.

Interestingly, some empirical evidence supports this view of the world. Nearly 26 years ago, Campbell and Dietrich (1983) first reported the apparently counterintuitive finding that mortgage loans characterized by high LTV ratios at origination actually exhibit lower default rates over time compared to their low LTV counterparts.

They conclude that the pattern of coefficients on the original loan-to-value dummy variables in their model is consistent with the presence of adverse selection in the underwriting process, particularly with respect to mortgages with original LTV ratios below 85%. Recent studies from multifamily and commercial mortgage markets also fail to document any positive relationship between LTV and borrower risk. For example, Archer et al. (1999, 2002) investigate pools of mortgage loans securitized by the Resolution Trust Corporation (RTC) for the Federal Deposit Insurance Corporation (FDIC) during the early to mid-1990s and find no significant relationship between LTV ratios at origination and ultimate loan performance (i.e., the probability of default).

Similarly, Ambrose and Sanders (2001) find a lack of significance for LTV in their commercial mortgage performance investigation, and they argue this is entirely consistent with lenders using a compensatory model of credit evaluation, where risky borrowers upon any given dimension are held to more stringent standards along alternative dimensions.

Mortgage underwriters and academicians have conventionally subscribed to the view that loan-to-value ratios positively influence default rates. The argument usually made is the following: The greater the financial leverage (i.e., the higher the LTVratio), the greater the debt service requirement, and hence the higher the probability the borrower will ultimately encounter financial distress. Theoretical signaling models, such as those presented by Rothschild and Stiglitz (1976) or Brueckner (1992), lend credibility to this paradigm, while a number of empirical studies, such as Von Furstenberg (1969) or Deng, Quigley and Van Order (2000), provide evidence consistent with the proposition.

Given the importance of discerning default risk in the multitrillion dollar mortgage industry, it is not surprising that a rich theoretical and empirical literature has emerged on pricing this risk and determining the factors that contribute to a borrower's decision to default. The most relevant theoretical study for the current paper is Brueckner (2000), which addresses the signaling role of a borrower's loan-to-value ratio choice with respect to that borrower's default risk. In Brueckner's model, default occurs when the value of the asset plus the borrower's default cost falls below the loan balance. Risky borrowers are defined as those who have lower default costs, and thus are more likely to exercise the default option. Similar to the result of Rothschild and Stiglitz (1976), the unique equilibrium in Brueckner is a separating equilibrium in which risky borrowers obtain a larger loan than safe borrowers. In addition to the loan-to-value ratio, the existing literature identifies an array of other risk factors that lenders might utilize to screen risky borrowers from safe borrowers. For example, in Titman (1992) firms have private information about their production efficiency.

He shows that, under certain parameter conditions, "good risk" firms choose short-term financing while "bad risk" firms choose long-term financing. Similarly, Guedes and Thompson (1995) offer empirical evidence for a signaling model where firms issue either fixed- or adjustable-rate debt to finance a project and obtain support for a separating equilibrium where high-quality firms issue high-default-risk debt and low-quality firms issue low default- risk debt.

Milde and Riley (1988) use a production economy to show that a firm with a less risky investment project may choose larger or smaller debt financing depending upon the firm's production function.

Another line of research studies the prepayment risk of the borrower (e.g., Dunn and Spatt 1988) and analyzes such instruments as ARMs (Brueckner 1992, Rosenthal and Zorn 1993, Posey and Yavas 1999), coupon rates and points on FRMs (Yang 1992, Brueckner 1994a, LeRoy 1996, Stanton and Wallace 1998), and prepayment penalties and due-on-sale clauses (Dunn and Spatt 1985, Chari and Jagannathan 1989) to separate high-prepayment-risk and low-prepaymentrisk borrowers. Although the setup and focus are different, it is also worth mentioning that there is a considerable literature on both the relationship between the borrower's risk type and his/her choice of a collateral requirement under asymmetric information (e.g., Barro 1976)

Episcopos (1998) explored the use of artificial neural networks in the modeling of foreclosure of commercial mortgages at his article. His article has used an extensive database of individual loan histories to demonstrate the use of radial basis function networks in the prediction of mortgage default and to compare the network's prediction performance with that of the logit benchmark. The article summarizes that logit model suffers from length-based sampling bias since it treats both the foreclosed and current loans as completed payment records. He selected 1,619 loans including 175 default ones. In the article logit and radial basis function networks are compared. A model is used and defualt estimation capability of logit models is tested. In his model, the results show that in the training set out of 150 defaulting mortgages logit classified forty of them correctly for a correct classification rate of 27%. Out of the 1,162 nondefaulting loans the logit model predicted 98% correctly. In the holdout set, out of the 282 nondefaulting loans the logit model predicted 98% correctly, whereas the RBFN predicted 96% correctly. He found that type of apartment credited, borrower type, location of the mortgage, interest rate, debt coverage significant to demonstrate the default rates. In this study, similarly, interest rate and a different type of coverage ratio, LTV, are found significant. Our data does not contain type of the apartment credited, borrower type and location of the mortgage, so no comment on these parameters is done.

4. Logit and Probit Models Theory

4.1. Logit Models

Logit regression (logit) analysis is a uni/multivariate technique which allows estimating the probability that an event occurs or not, by predicting a binary dependent outcome from a set of independent variables according to the document of Library of IASRI, Logit and Probit Analysis.

Moreover, in an example of home ownership where dependent variable owns a house or not in relation to income, the linear probability Model is depicted it as;

$P_i = E(Y=1 \mid X_i) = \beta_1 + \beta_2 X_i$

where X is the income and Y=1 means that the family owns a house.

Let's consider the following representation of home ownership;

$$P_i = E(Y=1 \mid X_i) = \frac{1}{1+\exp[-(\beta_1 + \beta_2 X_i)]} = \frac{1}{1+\exp(-Z_i)} \qquad (1)$$

where $Z_i = \beta_1 + \beta_2 X_i$

This equation (1) is known as the (cumulative) logistic distribution function. Here Z_i ranges from $-\infty$ to $+\infty$; P_i ranges between 0 and 1; P_i is non-linearly related to Z_i (i.e. X_i) thus satisfying the two conditions required for a probability model.

In satisfying these requirements, an estimation problem has been created because P_i is non linear not only in X but also in the β's. This means that one cannot use OLS procedure to estimate these parameters.

Here, P_i is the probability of owning a house and is given by

$$\frac{1}{1+\exp(-Z_i)}$$

Then $(1-P_i)$, the probability of not owning a house, is

$$(1 - P_i) = \frac{1}{1 + \exp(Z_i)}$$

Therefore, one can write

$$\frac{P_i}{(1 - P_{i)}} = \frac{1 + \exp(Z_i)}{1 + \exp(-Z_i)} \qquad (2)$$

$P_i / (1 - P_i)$ is the odds ratio in favor of owning a house. Taking natural log of (2),

$$L_i = \ln[P_i / (1 - P_i)] = Z_i = \beta_1 + \beta_2 X_i \qquad (3)$$

That is, the log of the odds ratio is not only linear in X, but also linear in the parameters. At this point, L is called the Logit.

As P goes from 0 to 1, the logit L goes from $-\infty$ to $+\infty$. That is, although the probabilities lie between 0 and 1, the logistics are not so bounded. In addition, although L is linear in X, the probabilities themselves are not linear. The interpretation, on the other hand, of the logit model is as follows: β_2, the slope, measures the change in L for a unit change in X, i.e. it tells how the log odds in favor of owning a house change as income changes by a unit. The intercept β_1 is the value of the log odds in favor of owning a house if income is zero. Moreover, the linear probability model assumes that P_i is linearly related to X_i, the logit model assumes that the log of odds ratio is linearly related to X_i.

If advantages of Logit Model are mentioned, logit analysis produces statistically sound results. By allowing for the transformation of a dichotomous dependent variable to a continuous variable ranging from $-\infty$ to $+\infty$, the problem of out of range estimates is avoided. Moreover, the logit analysis provides results which can easily interpreted and the method is simple to analyze. It gives parameters estimates which are asymptotically consistent, efficient and normal, so that the analogue of the regression t –test can be applied.

Logit model analysis can be used to identify the factors that affect the adoption of a particular technology say, use of new varieties, fertilizers on farm. Another application may be in the field of marketing, it can be used to test the brand preference and brand loyalty for any product. Another example can be gender studies. Gender studies can use logit analysis to find out the factors which affect the decision making status of men/women in a family.

4.2. Probit Models

In order to explain the behavior of a dichotomous dependent variable we have to use a suitably chosen Cumulative Distribution Function (CDF). The logit model uses the cumulative logistic function. But this is not the only CDF that one can use. In some applications, the normal CDF has been found useful. The estimating model that emerges from the normal CDF is known as the Probit Model.

In home ownership example, the decision of the i^{th} family to own a house or not depends on unobservable utility index I_i, that is determined by the explanatory variables in such a way that the larger the value of the index I_i, the greater the probability of the family owning a house.

The index I_i can be expressed as $I_i = \beta_1 + \beta_2 X_i$ (4)

where X_i is the income of the i^{th} family.

In order to estimate β_1 and β_2, (4) can be written as

$I_i = \beta_1 + \beta_2 X_i + U_i$.

Probit analysis is a type of regression used to analyze binomial response variables. It transforms the sigmoid dose-response curve to a straight line that can then be analyzed by regression either through least squares or maximum likelihood.

The idea of probit analysis was originally published in *Science* by Chester Ittner Bliss in 1934. He worked as an entomologist for the Connecticut agricultural experiment station and was primarily concerned with finding an effective pesticide to control insects that fed on grape leaves (Greenberg 1980). By plotting the response of the insects to various concentrations of pesticides, he could visually see that each pesticide affected the insects at different concentrations, i.e. one was more effective than the other. However, he didn’t have a statistically sound method to compare this difference. The most logical approach would be to fit a regression of the response versus the concentration, or dose and compare between the different pesticides. Yet, the relationship of response to dose was sigmoid in nature and at the time regression was only used on linear data. Therefore, Bliss developed the idea of transforming the sigmoid dose-response curve to a straight line. In 1952, a professor of statistics at the University of Edinburgh by the name of David Finney took Bliss' idea and wrote a book called *Probit Analysis* (Finney 1952). Today, probit analysis is still the preferred statistical method in understanding dose-response relationships.

The main difference between logit and probit model is that logistic has slightly flatter tails. The normal or probit curve approaches the axes more quickly than the logistic curve. Moreover, qualitatively, Logit and Probit Models give similar results; the estimates of parameters of the two models are not directly comparable.

5. Data Description

In this project, a mortgage based credit data is used including dates between 2005 and 2008. Time to maturity, payment to income ratio, credit open date, mortgage amount, marital status, loan to value ratio, credit limit, interest rate, installment amount, income, gender, expertise amount, education and age parameters are used.

Loan to value is one of the key risk factors that lenders assess when qualifying borrowers for a mortgage. The risk of default is always at the forefront of lending decisions, and the likelihood of a lender absorbing a loss in the foreclosure process increases as the amount of equity decreases. Therefore, as the LTV ratio of a loan increases, the qualification guidelines for certain mortgage programs become more strict. Lenders can require borrowers of high LTV loans to buy mortgage insurance to protect the lender from the buyer default, which increases the costs of the mortgage.

The valuation of a property is typically determined by an appraiser, but there is no greater measure of the actual real value of one property than an arms-length transaction between a willing buyer and a willing seller. Typically, banks will utilize the lesser of the appraised value and purchase price if the purchase is "recent." What constitutes recent varies by institution but is generally between 1–2 years.

Low LTV ratios (below 80%) carry with them lower rates for lower-risk borrowers and allow lenders to consider higher-risk borrowers, such as those with low credit scores, previous late payments in their mortgage history, high debt-to-income ratios, high loan amounts or cash-out requirements, insufficient reserves and/or no income documentation. Higher LTV ratios are primarily reserved for borrowers with higher credit scores and a satisfactory mortgage history. The full financing, or 100% LTV, is reserved for only the most credit-worthy borrowers.

In addition, payment to income ratio is used. Payment to income ratio gives an idea to the mortgage lenders about the condition of the borrower’s financial health. If the payment to income ratio is low then it represents a feasible situation due to the reason that a lesser amount of debt is commonly seen as reasonable or prudent. Nevertheless, if there is not any debt to repay, higher amount of money can be saved for serving other purposes. Unfortunately, if the payment to income ratio is high, this suggests that at the end of the month, there won't be much money to be saved. So this parameter is critical on the ongoing life of the loan.

We claim that the group of variables included in our data provides information on the main mortgage behavior, hence, it may be considered as an adequate measure of observed quality. At this stage, the data contains the following parameters; time to maturity, payment to income ratio, credit open date, mortgage amount, marital status, loan to value ratio, credit limit, interest rate, installment amount, income, gender, expertise amount, education and age parameters. Some statistics on these attributes are summarized in Table 1.

Table-1: Descriptive Statistics for the whole sample

Explanation	Number of Observations	Minimum Value	Mean Value	Median Value	Maximum Value	Standard Deviation
Time to Maturity	65000	-80,00	1.802,26	1.493,00	7.259,00	1.252,59
Payment to Income Ratio	65000	0,01	0,38	0,37	1,36	0,18
Credit Open Date	65000	04.01.2005	21.05.2007	03.08.2007	02.04.2009	402,14
Mortgage Amount	65000	8.800,00	113.727,36	80.000,00	13.000.000,00	141.085,48
Marital Status	65000	0,00	0,86	1,00	1,00	0,35
Loan to Value Ratio	65000	0,02	0,62	0,66	3,33	0,17
Credit Limit	65000	8.000,00	80.196,53	60.000,00	16.000.000,00	105.920,18
Interest Rate	65000	0,00	1,34	1,34	2,50	0,21
Installment Amount	65000	200,00	1.870,64	1.257,00	188.639,00	2.420,37
Income Level	65000	500,00	6.778,88	3.550,00	700.000,00	13.338,82
Gender	65000	0,00	0,76	1,00	1,00	0,42
Expertise Amount	65000	11.000,00	134.411,54	96.800,00	15.500.000,00	161.643,61
Education	65000	0,00	1,99	2,00	5,00	1,13
Age	65000	19,00	40,75	39,00	88,00	9,26
Default	65000	0,00	0,02	0,00	1,00	0,12

In descriptive statistics, there are some values to be explained. First of all, time to maturity value with -80,00 means that one the credits had defaulted 80 days ago. Secondly, the value of 13.000.000,00 in the mortgage amount means that for one of the credits, Bank has taken that amount of mortgage for assurance. Furthermore, 15.500.000,00 value in expertise amount and 16.000.000,00 TL credit limit means with an expertise amount of 15.500.000,00, a loan with a value of 16.000.000,00 is given to a customer.

6. Logit and Probit Models over the Default Probabilities

First of all, we regressed the dependent variable (default case) to 14 variables we have. With 65.000 observations, time to maturity, payment to income ratio, credit open date, mortgage amount, marital status, loan to value ratio, credit limit, interest rate, installment amount, income, gender, expertise amount, education and age parameters are used in binary probit estimation method. Results are following.

Table-2: Binary Probit Results with 14 Parameters

Dependent Variable: TAKIP
Method: ML - Binary Probit (Quadratic hill climbing)
Date: 06/02/09 Time: 22:29
Sample: 1 65000
Included observations: 65000
Convergence achieved after 23 iterations
Covariance matrix computed using second derivatives

Variable	Coefficient	Std. Error	z-Statistic	Prob.
TIME_TO_MATURITY	-0.026488	0.002149	-12.32716	0.0000
PI	-1.278847	0.365039	-3.503316	0.0005
OPEN_DATE	0.005087	0.000277	18.35297	0.0000
MORTGAGE__AMOUNT	1.21E-05	1.98E-06	6.132240	0.0000
MARITAL_STATUS	-0.043761	0.163289	-0.267997	0.7887
LTV	1.523068	0.491916	3.096198	0.0020
LIMIT__TL_	7.21E-06	3.68E-06	1.956367	0.0504
INTEREST	-0.057043	0.224050	-0.254599	0.7990
INSTALLMENT_AMOUNT	-0.000409	3.64E-05	-11.23730	0.0000
INCOME	-4.95E-05	9.81E-06	-5.041008	0.0000
GENDER	0.057915	0.155252	0.373037	0.7091
EXPERTISE_AMOUNT	-3.40E-06	1.65E-06	-2.057004	0.0397
EDUCATION	-0.187504	0.052724	-3.556296	0.0004
AGE	0.005946	0.006051	0.982583	0.3258
C	-3727.754	203.0664	-18.35732	0.0000

Mean dependent var	0.015369	S.D. dependent var	0.123017
S.E. of regression	0.032268	Akaike info criterion	0.008238
Sum squared resid	67.66545	Schwarz criterion	0.010334
Log likelihood	-252.7380	Hannan-Quinn criter.	0.008887
Restr. log likelihood	-5162.496	Avg. log likelihood	-0.003888
LR statistic (14 df)	9819.516	McFadden R-squared	0.951043
Probability(LR stat)	0.000000		
Obs with Dep=0	64001	Total obs	65000
Obs with Dep=1	999		

P(Takip=1|X)=Φ(-3,737.754-0,026TMT-1.278PI+0.005OPD+1.21*10-5MA-0.043MS+1.523LTV+7.21*10-6LIM-0.057INT-0.001INS-4.9*10-5INC+0.057GEN-3.4*10-6EA-0.187EDC+0.005AGE)

In this equation, marital status, interest rate, gender and age parameters are insignificant with p-value of less than 5%. A point is R-squared value is 0.95 which is high which represent the ability of the sample representing the whole. To summarize results, if time to maturity parameter, payment to income ratio, marital status, interest rate, installment amount parameter, customer's income level, expertise amount and education level increases, the default probability decreases with changing coefficients, and if open date, mortgage amount, loan to value ratio, credit limit, gender and age increases, the default probability increases with changing coefficients. At this point, marital status, interest rate, gender and age parameters are insignificant with p-value of less than 5%. We estimate the equation without two of the insignificant parameters, marital status and gender.

Table-3: Binary Probit Results without Remaining 12 Variables

Dependent Variable: TAKIP
Method: ML - Binary Probit (Quadratic hill climbing)
Date: 06/02/09 Time: 22:44
Sample: 1 65000
Included observations: 65000
Convergence achieved after 46 iterations
Covariance matrix computed using second derivatives

Variable	Coefficient	Std. Error	z-Statistic	Prob.
TIME_TO_MATURITY	-0.030445	0.002476	-12.29691	0.0000
PI	-1.610330	0.427266	-3.768914	0.0002
OPEN_DATE	0.006629	0.000397	16.71663	0.0000
MORTGAGE__AMOUNT	1.34E-05	2.18E-06	6.149639	0.0000
LTV	1.487664	0.546256	2.723383	0.0065
LIMIT__TL_	1.24E-05	3.20E-06	3.881645	0.0001
INTEREST	-0.290396	0.240491	-1.207512	0.2272
INSTALLMENT_AMOUNT	-0.000459	4.20E-05	-10.91631	0.0000
INCOME	-6.53E-05	1.17E-05	-5.594620	0.0000
EXPERTISE_AMOUNT	-5.68E-06	2.28E-06	-2.491891	0.0127
EDUCATION	-0.236094	0.058716	-4.020928	0.0001
AGE	0.006357	0.006585	0.965351	0.3344
C	-4857.378	290.4980	-16.72086	0.0000

Mean dependent var	0.015369	S.D. dependent var	0.123017
S.E. of regression	0.031956	Akaike info criterion	0.007820
Sum squared resid	66.36449	Schwarz criterion	0.009636
Log likelihood	-241.1371	Hannan-Quinn criter.	0.008382
Restr. log likelihood	-5162.496	Avg. log likelihood	-0.003710
LR statistic (12 df)	9842.717	McFadden R-squared	0.953291
Probability(LR stat)	0.000000		
Obs with Dep=0	64001	Total obs	65000
Obs with Dep=1	999		

P(Takip=1|X)=Φ(-4,857.378-0,030TMT-1.610PI+0.006OPD+1.34*10-5MA-

+1.487LTV+1.24*10-5LIM-0.029INT-0.001INS-6.53*10-5INC-5.68*10-6EA-

0.236EDC+0.006AGE)

To summarize the results, if time to maturity parameter, payment to income ratio, interest rate, installment amount parameter, customer's income level, expertise amount and education level increases, the default probability decreases with changing coefficients, and if open date, mortgage amount, loan to value ratio, credit limit and age increases, the default probability increases with changing coefficients. At this point, interest rate and age parameters are insignificant with p-value of less than 5%. Estimation without age parameter is done again. Interest rate is insignificant and it doesn't eliminated because it is thought that interest rate of a loan is a main characteristics in determining the default case.

Table-4: Binary Probit Results without Age Parameter

Dependent Variable: TAKIP
Method: ML - Binary Probit (Quadratic hill climbing)
Date: 06/02/09 Time: 22:49
Sample: 1 65000
Included observations: 65000
Convergence achieved after 46 iterations
Covariance matrix computed using second derivatives

Variable	Coefficient	Std. Error	z-Statistic	Prob.
TIME_TO_MATURITY	-0.028648	0.002261	-12.67065	0.0000
PI	-1.975778	0.366778	-5.386858	0.0000
OPEN_DATE	0.006570	0.000382	17.20294	0.0000
MORTGAGE__AMOUNT	1.09E-05	2.22E-06	4.921118	0.0000
LTV	2.531281	0.352507	7.180801	0.0000
LIMIT__TL_	6.50E-06	2.25E-06	2.881925	0.0040
INTEREST	-0.279313	0.237210	-1.177494	0.2390
INSTALLMENT_AMOUNT	-0.000407	3.03E-05	-13.44594	0.0000
INCOME	-8.28E-05	6.52E-06	-12.69854	0.0000
EDUCATION	-0.240033	0.057669	-4.162217	0.0000
C	-4814.095	279.8054	-17.20515	0.0000

Mean dependent var	0.015369	S.D. dependent var	0.123017
S.E. of regression	0.032158	Akaike info criterion	0.007893
Sum squared resid	67.20950	Schwarz criterion	0.009430
Log likelihood	-245.5142	Hannan-Quinn criter.	0.008368
Restr. log likelihood	-5162.496	Avg. log likelihood	-0.003777
LR statistic (10 df)	9833.963	McFadden R-squared	0.952443
Probability(LR stat)	0.000000		

Obs with Dep=0	64001	Total obs	65000
Obs with Dep=1	999		

P(Takip=1|X)=Φ(-4,814.095-0,028TMT-1.9750PI+0.006OPD+1.09*10-5MA-

+2.531LTV+6.5*10-6LIM-0.279INT-0.001INS-8.28*10-5INC-0.24EDC)

To summarize the results, if time to maturity parameter, payment to income ratio, interest rate, installment amount parameter, customer's income level and education level increases, the default probability decreases with changing coefficients, and if open date, mortgage amount, loan to value ratio and credit limit increases, the default probability increases with changing coefficients. R-squared is still high enough and 0.95. At this point, with the remaining parameters only interest rate is not significant. Limit parameter's coefficient is very small, so a try is done in order to evaluate the effect of it.

Table-5: Binary Probit Results with Significant Parameters

Dependent Variable: TAKIP
Method: ML - Binary Probit (Quadratic hill climbing)
Date: 06/02/09 Time: 22:52
Sample: 1 65000
Included observations: 65000
Convergence achieved after 14 iterations
Covariance matrix computed using second derivatives

Variable	Coefficient	Std. Error	z-Statistic	Prob.
TIME_TO_MATURITY	-0.024844	0.001156	-21.48236	0.0000
PI	-1.596569	0.199682	-7.995559	0.0000
MORTGAGE__AMOUNT	1.24E-05	6.03E-07	20.51439	0.0000
LTV	1.278616	0.169894	7.525969	0.0000
INTEREST	1.123988	0.141914	7.920197	0.0000
INSTALLMENT_AMOUNT	-0.000303	2.07E-05	-14.58395	0.0000
INCOME	-6.13E-05	4.07E-06	-15.06626	0.0000
EDUCATION	-0.218784	0.029947	-7.305691	0.0000
C	-1.026780	0.243262	-4.220872	0.0000

Mean dependent var	0.015369	S.D. dependent var	0.123017
S.E. of regression	0.066880	Akaike info criterion	0.029395
Sum squared resid	290.6985	Schwarz criterion	0.030653
Log likelihood	-946.3498	Hannan-Quinn criter.	0.029785
Restr. log likelihood	-5162.496	Avg. log likelihood	-0.014559
LR statistic (8 df)	8432.292	McFadden R-squared	0.816688
Probability(LR stat)	0.000000		
Obs with Dep=0	64001	Total obs	65000
Obs with Dep=1	999		

P(Takip=1|X)=Φ(-1.026-0,024TMT-1.5960PI+1.278LTV+1.123INT-0.001INS-6.13*10-5INC-0.218EDC)

To summarize the results, if time to maturity parameter, payment to income ratio, installment amount parameter, customer's income level and education level increases, the default probability decreases with changing coefficients, and if mortgage amount, loan to value ratio and credit interest rate increases, the default probability increases with changing coefficients.

At this estimation, R-squared value with 0.816 is reached and all parameters are significant in p<0,05 ad p<0,01. The next table is with binary logit model estimation.

Table-6: Binary Logit Results with Significant Parameters

Dependent Variable: TAKIP
Method: ML - Binary Logit (Quadratic hill climbing)
Date: 06/02/09 Time: 22:55
Sample: 1 65000
Included observations: 65000
Convergence achieved after 14 iterations
Covariance matrix computed using second derivatives

Variable	Coefficient	Std. Error	z-Statistic	Prob.
TIME_TO_MATURITY	-0.046292	0.002286	-20.24718	0.0000
PI	-2.502507	0.424277	-5.898287	0.0000
MORTGAGE__AMOUNT	2.61E-05	1.38E-06	18.93902	0.0000
LTV	2.309587	0.306723	7.529873	0.0000
INTEREST	2.120430	0.257099	8.247538	0.0000
INSTALLMENT_AMOUNT	-0.000829	7.25E-05	-11.44216	0.0000
INCOME	-0.000107	1.47E-05	-7.263346	0.0000
EDUCATION	-0.374877	0.054148	-6.923191	0.0000
C	-1.892149	0.442584	-4.275230	0.0000

Mean dependent var	0.015369	S.D. dependent var	0.123017
S.E. of regression	0.065003	Akaike info criterion	0.028335
Sum squared resid	274.6139	Schwarz criterion	0.029592
Log likelihood	-911.8743	Hannan-Quinn criter.	0.028724
Restr. log likelihood	-5162.496	Avg. log likelihood	-0.014029
LR statistic (8 df)	8501.243	McFadden R-squared	0.823366
Probability(LR stat)	0.000000		

Obs with Dep=0	64001	Total obs	65000
Obs with Dep=1	999		

P(Takip=1|X)=F(-1.892-0,046TMT-2.502PI+2.309LTV+2.12INT-0.001INS-0.001INC-0.374EDC)

At this estimation, R-squared value with 0.823 is reached and all parameters are significant in p<0,05 ad p<0,01.

Result tables can be evaluated as below;

- If time to maturity parameter increases, the default probability decreases. This means, if loan will expire soon or if a few payments remain, credit default probability is high. This may be because of the last period crisis.

- If payment to income ratio increases, the default probability decreases. This may be expected as being the positive sign but the parameter is significant and this means that there is a reverse relationship between payment to income ratio and the default probability.

- Because the coefficient for the mortgage amount is very small and close to 0, it can be stated that there is not any significant effect of this parameter on the default probability.

- If loan to value ratio increases, the default probability increases. This means, if higher loans with low expertise amounts are assigned, this will end up with high default probabilities.

- If credit interest rate increases, the default probability increases. This means, loans assigned with higher interest rates have higher default probabilities.

- If installment amount increases, the default probability decreases. This may be expected as being the positive sign but the parameter is significant and this means that there is a reverse relationship between installment amount and the default probability. But the coefficient is very small for a significant effect.

- If customers' income level increases, the default probability decreases. This parameter is behaving as expected. This means, high level income gainers pay out their credits.

- If education level increases, the default probability decreases. This parameter is behaving as expected. This means, high level educated people pay out their credits.

7. Discussion

When estimated results are examined, it is seen that payment to income ratio's sign is negative. This parameter's sign is expected to be positive. In relation, income parameter's sign is negative and not to significant to consider. So, double negative sign may be commented as the effect of higher payment on default probability is positive.

Moreover, this study is done with active loans. The data don't include the credits that are successively paid and closed by not facing any default case. The data contains only snapshot active credits. So, there is always a probability of default for non-default credits.

It is thought that if the study is made for a time period at the past, with a data consisting default credits and normal credits, all paid by the customers at a time, the results will be more meaningful. If the results are compared with that time's non performing loan rate, stating the default case will be more accurate. But I cannot reach to opened and closed credit data. So this kind of study can be done.

Furthermore, there are continuously changing parameters in our data. For instance, income level of a customer may continuously change. But the data, and the Bank's decision strategy, is based on the income level which is stated by the customer during the credit application procedure. The effect of changing parameters is assumed to be zero because to gather a continuously changing data is too difficult to handle with.

8. Conclusion

In this study, we try to find out a profile for default cases of mortgage loans with characteristics of loans and customers. In order to determine dominant factors, at the beginning of the study, relevant parameters are chosen representing a mortgage loan and the credited customer. These parameters are time to maturity, payment to income ratio, credit open date, mortgage amount, marital status, loan to value ratio, credit limit, interest rate, installment amount, income, gender, expertise amount, education and age parameters. First of all, estimation with these 14 parameters is made. In first estimation, marital status and gender parameters became insignificant. In the next estimation, age parameter became insignificant. Then, in last estimation, time to maturity, payment to income ratio, mortgage amount, loan to value ratio, interest rate, installment amount, income and education parameters remain significant with an R-squared value of 0.82. To summarize the results, if time to maturity parameter, payment to income ratio, installment amount parameter, customer’s income level and education level increases, the default probability decreases with changing coefficients, and if mortgage amount, loan to value ratio and credit interest rate increases, the default probability increases with changing coefficients. While evaluating the results, if this study is done with a data consisting default credits and normal credits, all paid by the customers, it will be more meaningful as it is discussed in the discussion part. Furthermore, there are continuously changing parameters in this study, but because time effect on parameters cannot be continuously followed, these parameters’ values are accepted on snapshot basis and results are defined on that principle.

9. References

1. Ambrose, B. and A.B. Sanders, 2001. Commercial Mortgage-Backed Securities: Prepayment and Default. University of Kentucky.

2. Archer,W.R., P.J. Elmer, D.M. Harrison and D.C. Ling. 1999. Determinants of Multifamily Mortgage Default. Working Paper 99-2.Washington, DC.

3. Barro, R.J., 1976. The Loan Market, Collateral, and Rates of Interest. Journal of Money, Credit, and Banking 8: 439–456.

4. Brueckner, J.K., 1992. Borrower Mobility, Self-Selection, and the Relative Prices of Fixed- and Adjustable-Rate Mortgages. Journal of Financial Intermediation 2: 401–421.

5. Campbell, T.S. and J.K. Dietrich, 1983. The Determinants of Default on Insured Conventional Residential Mortgage Loans. Journal of Finance 38: 1569–1579.

6. Chari, V.V. and R. Jagannathan, 1989. Adverse Selection in a Model of Real Estate Lending. Journal of Finance 44: 499 508.

7. Covered Bond Outstanding, 2007, (http://ecbc.hypo.org).

8. CSO Financial Statistics and Building Societies Commission Annual Reports.

9. Deng, Y., J.M. Quigley and R. Van Order, 2000. Mortgage Terminations, Heterogeneity, and the Exercise of Mortgage Options. Econometrica 68: 275–307.

10. Dunn, K.B. and C.S. Spatt, 1985. Prepayment Penalties and the Due-on-Sale Clause. Journal of Finance 40: 293–308.

11. Guedes, J. and R. Thompson., 1995. Tests of a Signaling Hypothesis: The Choice Between Fixed- and Adjustable-Rate Debt. Review of Financial Studies 8: 605–636.

12. Housing Finance Systems for countries in Transition - Principles and Examples, 2005, United Nations, New York and Geneva, p. 45.

13. ISR/Google Books, Third Rev. Edtn. 2008. The Mortgage Loans Industry and Market: A Survey.

14. Kluwer Academic Publishers, Boston, 1998. Commercial Mortgage Default: A Comparison of Logit with RBFN.

15. Milde, H. and J.G. Riley., 1988. Signaling in Credit Markets. Quarterly Journal of Economics 102: 101–129.

16. Posey, L. and A. Yavas, 2001. Adjustable and Fixed Rate Mortgages as a Screening Mechanism for Default Risk. Journal of Urban Economics 29: 54–79.

17. Rosenthal, S.S. and P. Zorn, 1993. Household Mobility, Asymmetric Information, and

the Pricing of Mortgage Contract Rates. Journal of Urban Economics 33: 235–253.
18. Rothschild, M. and J. Stiglitz, 1976. Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information. Quarterly Journal of Economics 90: 629–649.
19. The Mortgage Loans Industry and Market: A Survey, ps. 15-16.
20. Titman, S., 1992. Interest Rate Swaps and Corporate Financing Choices. Journal of Finance 47: 1503–1516.
21. UBS Phillips and Drew, 1992. Building Societies Research: Investing for the Next Millennium.
22. Vasisht A.K., 2004. Logit and Probit Analysis, Library of IASRI.
23. Vincent, Kim, 2008. Probit Analysis.
24. Von Furstenberg, G.M., 1969. Default Risk on FHA-Insured Home Mortgages as a Function of the Terms of Financing: A Quantitative Analysis. Journal of Finance 24.
25. Yang, T.L., 1992. Self-Selection in the Fixed-Rate Mortgage Market. AREUEA Journal 20: 359–391.
26. Yavas, Noordewier, Harrison, 2004. Do Riskier Borrowers Borrow More? Real Estate Economics: ps. 385-411.