LAPPEENRANTA UNIVERSITY OF TECHNOLOGY
School of Business
Strategic Finance

Matti Tanninen

Determinants of credit risk in secured loans – Evidence from the auto loan industry

Examiners: Associate Professor Sheraz Ahmed
Professor Eero Pätäri

ABSTRACT

Author: Matti Tanninen
Title: Determinants of credit risk in secured loans – Evidence from the auto loan industry
Faculty: School of Business
Major: Finance
Year: 2013
Examiners: Associate Professor Sheraz Ahmed and Professor Eero Pätäri
Master's Thesis: LUT School of Business 107 pages, 9 figures, 34 tables, 4 appendices

Key Words: Credit risk, Default, Payment problems, Overdue, Auto loans, Instalment loans, Logistic regression,

The purpose of this study is to examine attributes which have explanation power to the probability of default or serious overdue in secured auto loans. Another goal is to find out differences between defaulted loans and loans which have had payment difficulties but survived without defaulting. 19 independent variables used in this study reflect information available at the time of credit decision. These variables were tested with logistic regression and backward elimination procedure. The data includes 8931 auto loans from a Finnish finance company. 1118 of the contracts were taken by company customers and 7813 by private customers. 130 of the loans defaulted and 584 had serious payment problems but did not default. The maturities of those loans were from one month to 60 months and they have ended during year 2011.

The LTV (loan-to-value) variable was ranked as the most significant explainer because of its strong positive relationship with probability of payment difficulties. Another important explainer in this study was the credit rating variable which got a negative relationship with payment problems. Also maturity and car age performed well having both a positive relationship with the probability of payment problems. When compared default and serious overdue situations, the most significant differences were found in the roles of LTV, Maturity and Gender variables.

TIIVISTELMÄ

Tekijä: Matti Tanninen
Tutkielman nimi: Luottoriskiin vaikuttavat tekijät vakuudellisissa lainoissa – tutkimus autolainoista
Tiedekunta: Kauppatieteellinen tiedekunta
Pääaine: Rahoitus
Vuosi: 2013
Tarkastajat: Tutkijaopettaja Sheraz Ahmed ja professori Eero Pätäri
Pro Gradu -tutkielma: Lappeenrannan teknillinen yliopisto, 107 sivua, 9 kuviota, 34 taulukkoa ja 4 liitettä

Hakusanat: Luottoriski, maksun laiminlyönti, maksuongelmat, maksun viivästyminen, autolaina, osamaksu, logistinen regressio

Tämän tutkimuksen tarkoituksena on löytää muuttujia, jotka pystyvät selittämään todennäköisyyttä, jolla vakuudelliset autolainat aiheuttavat luottotappioita tai niiden ottajat joutuvat vakaviin maksuvaikeuksiin. Lisäksi pyrimme löytämään eroja kahden edellä mainitun sopimusryhmän välillä. Selittävinä muuttujina käytämme 19 muuttujaa, jotka ilmentävät luottopäätöksen tekohetkellä olevia tietoja. Tutkimusmenetelmänä käytämme logistista regressiota ja muuttujien valinnassa backward elimination – menetelmää. Tutkimuksessa käytetty data sisältää 8931 vuoden 2011 aikana päättynyttä autorahoitussopimusta suomalaisesta rahoitusyhtiöstä. Luottojen pituus on ollut yhdestä kuukaudesta 60 kuukauteen. Sopimuksista 1118 oli yritysten ottamia ja 7813 kuluttaja-asiakkaiden. Sopimuksista 130 aiheutti luottotappioita ja 540 sopimuksessa asiakkaalla oli vakavia maksuvaikeuksia.

Tutkimuksen merkittävimmäksi selittäväksi muuttujaksi osoittautui LTV (lainan määrä verrattuna vakuuden arvoon), jolla todettiin olevan vahva positiivinen yhteys maksuvaikeuksien todennäköisyyden kanssa. Toinen merkittävä selittävä muuttuja oli luottoluokitus, jonka vaikutus maksuvaikeuksien todennäköisyyteen oli negatiivinen. Myös rahoitusajalla ja ajoneuvon iällä huomattiin olevan merkittävä positiivinen vaikutus maksuvaikeuksiin. Vertailtaessa sopimuksia, jotka aiheuttivat luottotappioita ja sopimuksia, joissa maksut olivat merkittävästi myöhässä, huomasimme suurimmat erot muuttujissa LTV, rahoitusaika ja sukupuoli.

ACKNOWLEDGEMENTS

This thesis was the final challenge before my graduation. This challenge ended my educational journey which started over 17 years ago in Mikkeli. I want to thank my parents for their unconditional support during that journey.

The writing of the thesis was rewarding and interesting over the eight month long process. I would like to thank Professor Sheraz Ahmed for his efforts as an instructor. I would also like to thank the company which provided me the data for this study and made this thesis possible.

Finally, I want to acknowledge the support of my girlfriend Laura who has been as patient as she always is during this thesis process.

Vantaa 21st of May 2013,

Matti Tanninen

TABLE OF CONTENTS

APPENDICES

1 INTRODUCTION

1.1 Personal motivation

“Risk comes from not knowing what you are doing.” –Warren Buffett

The credit risk is usually like a hidden truth behind one symbol. It is very simple and easy to illustrate the precise amount of credit risk with just one single character. It is much more laborious and tricky to describe which attributes cause the risk and what the weights of each attributes are. Unfortunately, the second option and more laborious one is precisely the one we should employ to succeed. Warren Buffett, one of the most successful investors in the world, has stated in many interviews that the actual source of the risk is the unawareness. Accordingly, if you do not have a clue from which attributes the credit risk symbol is formed and you believe in it blindly the symbol might be the source of the risk itself. Because of our professional curiosity we do not satisfy only to stare at the symbols. We want to know from which attributes the symbols are made of in our professional context.

Our professional experience is based on one Finnish finance company. Almost all our tasks are related to the controlling of the credit risk or the consequences of when credit risk has realized. The latter situation is more commonly known as a default or a serious overdue. Traditional credit risk models focus on determining the probability of default. According to Okumu, Mwalili and Mwita (2012, pp. 22-24) those models classify borrowers into different risk categories which predict their probability to default. Those categories are commonly used in the credit decision process because of practical reasons; there must be some quick way to analyse the risk of one application. Because of our professional curiosity we wanted to form our own model to explain default and overdue situations.

There exists also a special interest in one certain variable. This variable is down payment which is a hot topic in today’s Finnish media especially in the context of residential mortgage loans. Actually mortgage house loans and instalment loans for car purchases are very much similar, as Heitfield and Tarun (2004, pp. 474) noticed in their research “What Drives Default and Prepayment on Subprime Auto Loans?”, both are secured with collaterals which are a part of everyday life for most people. Both also include fixed-coupon amortization schedules and carry fixed interest rates. In car loans the value of the collateral varies more aggressively than in house loans. Usually it goes down with an unknown speed. In mortgage house loans the value of the collateral might actually increase. All in all, the situation is very similar in both loans; the credit risk arises from the difference between the amount of the loan and the value of the collateral.

Basically the only way to reduce the credit risk in both previously mentioned loans is to take more down payment. That is a critical variable in both types of loans. Minister Antti Tanskanen and his working group introduced for Finnish Ministry of Finance in October 2012 that it would be necessary to limit residential mortgage loans to be maximum 80% of the value of the collateral object. The other way around this means, that the down payment in such loans should be at least 20%. Because of this topical and emotive conversation, it was seen to be essential to take a special notice to the meaning of the down payment in auto loan contracts.

1.2 Scientific relevance

From the scientific point of view the motivation to this study is the dilemma between granted loans, occurred defaults and parties which have caused the defaults. As Figure 1 shows the lending portfolio which includes loans from commercial and public institutions to consumers and companies, grew more than 150% in ten years in Finland (Statistics Finland 2012). Note that in Figure 1 there are all kinds of loans included not just auto loans. Auto loans are tricky to separate from the whole lending portfolio because finance houses have no incentives to reveal what kind of purchases they have granted loans for. We have to remember that the biggest finance houses give consumption loans for several purposes.

Another reason why car loans are tricky to separate is that some car purchases are made with ordinary bank loans where the car is not collateral. Those loans are also included in Figure 1. The point of Figure 1 is to describe how the overall lending activity has risen.

Figure 1: Lending portfolio in Finland 2002-2012. (Statistics Finland 2012; For precise amounts see appendix 1)

In addition to growth in granted loans the defaults have increased as well. Figure 2 shows statistics of how the amount of default register marks have grown during the last six years.

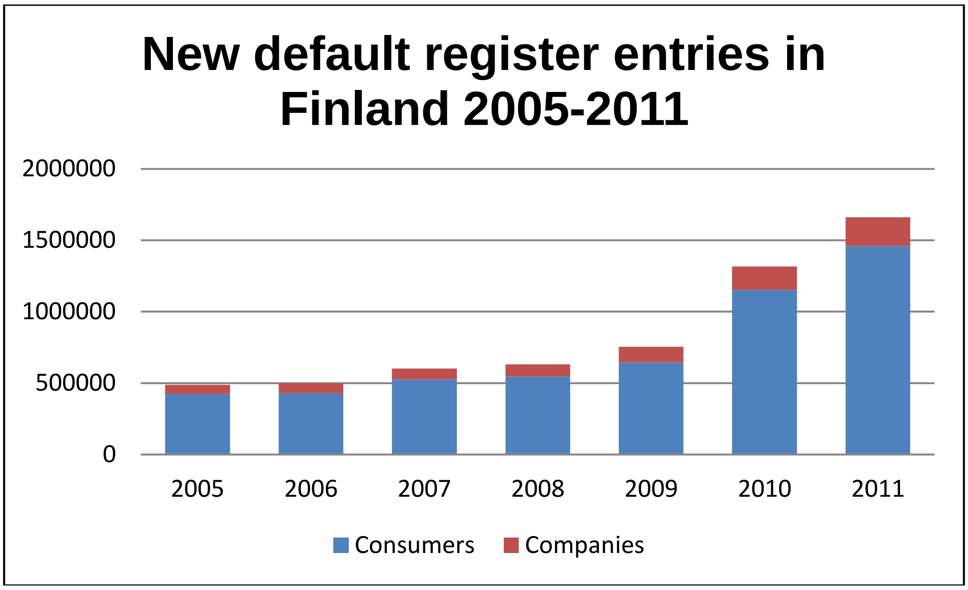


Figure 2: New payment default entries for private customers and companies 2005 – 2011 (Suomen Asiakastieto Oy, 2012; For precise amounts see appendix 2)

In Finland individuals and companies can get default entries several times. This means that every loan which ends to default can cause a new entry. On the other hand this means that one single default mark does not necessarily mean that the individual or the company is in the situation of total insolvency. It just signals that the obligor has got problems with at least one of its liabilities. For private customers the amount of new entries per year has grown more than 240% in seven years and for companies more than 200% (Suomen Asiakastieto 2012, Appendix 2). It is interesting that, at the same time when lending portfolio and the amount of defaults have more than doubled, the amount of private customers with default marks has grown only 8%. Corresponding number for companies is 30%. (Suomen Asiakastieto 2012, Appendix 3). It seems that the very same individuals and companies get default register entries over and over again. This also gives relevance for this research because defaults do not seem to be random events at all and thus it should be possible to forecast them even at some level.

In credit risk literature it is very common to estimate the probability of default (PD) by using dataset that includes information of credit derivatives. The current financial crisis, or more presice the premises of that, focuses our sight into the roots of the credit risk and to analyse the structure of the loan portfolio and single contracts. In addition we must keep in mind that the credit risk is still the most important risk factor in banking in the vast majority of countries. According to Virolainen (2004, pp.8) credit losses from defaulted credits were the main reason for difficulties during the banking crisis in the early 1990s in Finland. As Smith (2011, pp. 8) noticed, in publication of federal reserve bank of Philadelphia, the current financial meltdown has made us overlook the defaults on loans for motor vehicles. These findings inspired us to make a statistical analysis of the credit risk in auto loan context.

1.3 Research problems

The purpose of this study is to find out if some of the information, available at the moment of credit decision, can explain the default or overdue situation. In other words we try to find out if there is any information in our professional context which gives a hint of the future payment problems. We divide payment problems to two categories; default and serious overdue. Default means that the contract has caused credit losses for the finance company. Serious overdue means that there is a delay in the payments more than 60 days during six months period before the end of the contract. The definition of the serious overdue comes from our target company. Datasets which include defaulted contracts include also contracts with serious overdue situations and naturally contracts with no payment problems at all. However in those datasets contracts with overdue values are treated exactly same way as the contracts with no payment problems at all. In those datasets we are only interested in if the contract has defaulted or not. Datasets which predict the probability of serious overdue do not include any defaulted contracts. Another thing which we try to solve is if there are any remarkable differences between loans which default and loans which have serious payment difficulties but survive without defaulting. That question comes from the data provider.

The research problems that are examined in this study can be expressed as following questions:

1. Are there variables which can forecast the probability of default or serious overdue?

2. Are there remarkable differences between contracts which have defaulted and contracts which have had payment delays?

1.4 Structure and limitations

In many previous studies the comparison of different methods and also the comparison of the prediction accuracy of different methods have been the main goal. However in this quantitative study the main goal is to solve the predictive power of different variables with logistic regression. The variables are chosen by our professional curiosity and expertise.

The first Chapter introduces the motivation to this study and also the backgrounds for this paper. The second Chapter presents the theoretical backgrounds. It introduces basic concepts of credit risk, payment problems and auto loan context. The third Chapter introduces seven previous studies which are divided into two groups. Four of them explain payment problems at general level and the rest of them concentrate especially on auto loan defaults. In the Chapter 4 we introduce facts of the unique dataset used in this research and also the special features of the two dependent and 19 independent variables. The Chapter 5 is for the tests and the results. That Chapter is divided into three sections. In the first one we introduce the results of the single regressions. In the second one we introduce the results from backward elimination procedure and also the final models. In the third section we present our subjective variable ranking which includes also our interpretations. The last Chapter concludes this study, giving suggestions for managerial implementation and suggests some topics for further research in the area of default and payment problems estimation.

There are some limitations in this study. Firstly we must state that our data includes only loans which have ended during year 2011. That is because it was the most recent dataset available and thus the most useful information for the target company. Because we have the most recent data available in our study we do not have more up-to-date data which we could have used as a benchmark to our models. Thus we have to do that part later. Secondly we define a serious overdue as a situation where a payment has been late more than two weeks during six months before the full payment of the loan. It is possible that there has been overdue situation earlier in the lifetime of the loan than only six months before the full payment. In other words the dataset used includes only recent payment problems. One limitation is that the results can be generalized only in our target company. Some of the variables are unique and they have not been even introduced in detail due to the respect for the data provider’s requests. If necessary our results might be possible to generalize in Finland but in that case it is important to remember that this study concerns only secured instalment loans where the collateral is an automobile.

2 THEORY

2.1 The definition of risk

At the end of the day the only constant thing in business environment is the change (Jorion 2001, pp. 4). The question which leads us to define the word risk is; What is the result of the change?

The risk itself means the possibility of some kind of harm in the future. In other words, it describes the volatility of unexpected outcomes. That is why the definitions of different risks focus on describing those possibilities. The scientific research of risks focuses on describing uncertainty which is consequence of possible harm. Thanks to the modern way of risk management, which was developed in the 18th century, nowadays it is possible to define risk factors, harms and probabilities and hedge against them (Alhonsuo, Nisén and Pellikka 2009, pp. 40; Jorion 2001, pp.3).

Usually risks are divided into two categories: business and non-business risks. The business risk refers to the situation where it is possible for the company to have an impact on the possible harm in the future. Therefore it is related to the core competence of the firm which is after all the only thing which creates the competitive advantage to the firm and adds the value for the shareholders. (Jorion 2001, pp. 3-4) Those business risk sources are crucial part of understanding the business logic of one firm. If the change is the only constant thing in the business environment and one can manage it well, is it not the most important single component for success? One good example of the business risk is innovations of competitors that can change the balance in the market. Non-business risks are usually harms in the future which cannot be influenced by the company so firms do not have a control of them. Non-business risks include strategic risk which results from the changes in the political or economic environment in the country where the firm operates. (Jorion 2001, pp.4) Example of this kind of risk is a rapid disappearance of the Soviet Union in the early 1990s which revolutionized the political and economic environment in the Finnish industries (especially in the export market).

From where does risk arise from? Some risks are produced by humans such as business cycles, political changes or inflation (Jorion 2001, pp. 7-8). They are easier to forecast than other risks because the human nature seems to have some kind of logic. Or at least scientist tries to figure it out constantly. Karl Marx said “History repeats itself, first as tragedy, second as farce”. Marx probably meant that in most cases the unforeseen change was actually a visible thing because of the tragedy in the past. In this light the changes which are results of human behavior are the ones which are possible to forecast by using a historical data. Other risks are made by no-one, such as unforeseen natural phenomena. (Jorion 2001, pp. 8)

2.2 Credit risk and credit losses

Credit risk consists of two sources; default risk and credit spread risk. The default risk means that the borrower does not meet a part or all of his obligations. (Choudhry 2004, pp. 2) The credit spread risk refers to the “mark-to-market” approach which Kimmo Virolainen introduced in the discussion paper “Macro stress testing with a macroeconomic credit risk model for Finland” made for Bank of Finland (Virolainen 2004). It refers to an unexpected decrease in the credit quality, for example a sudden drop in the credit rating. Stephanou and Mendoza (2005) included value risk in their credit risk definition in their working paper for the World Bank. They actually meant almost the same thing as Virolainen (2004) because they refer with the word “value” to the opportunity cost of not pricing the loan correctly because of recently decreased risk rating. So the realization of the spread risk or value risk does not necessarily mean the loss of the principal or interest. All in all, the credit risk refers to the situation when expected cash flows are threatened and it is actually still the most important risk in banking in the vast majority of countries (Virolainen 2004).

There is a dilemma concerning the credit risk in the financial industry. It comes from the fact that financial institutions are more willing to borrow money to a party which is already their customer. This is natural because in such case the lender has information of payment behavior of the customer with whom they have an ongoing relationship. The dilemma comes into the picture when we start to think about a diversification. This dilemma has brought portfolio thinking as a part of measuring the credit risk. (Jorion 2001, pp. 313) A tool against this diversification problem is limits which are usually set for the biggest clients to make sure that one single customer does not get too heavy position in the loan portfolio. However, those limits do not remove the dilemma. There exists an incentive to loan for same customer more and more and thus create an imbalanced loan portfolio.

Default by definition refers to the situation when borrower is not able to meet his obligations. The actual moment of the default differs highly in different contexts. Some creditors consider the status of the loan as default when the payments of the borrower are late more than 90 days. According to the International Settlement standard for Banks: The client is in default if any payment connected to the loan is overdue more than 90 days (Kocenda and Vojtek 2009, pp. 6). That is called a technical default which refers for example to a company which has not honored its payment obligations but has not yet reached the stage of the bankruptcy either. Same technical default fits for a private customer who has not paid as agreed but who is not yet at a stage of official insolvency or a loan arrangement either. After a proper default the lender gets the recovery amount which is usually expressed as a percentage of the total loan amount. It consists of money from foreclosure, liquidation or restructuring of the defaulted borrower. (Choudhry 2004, p.2) In Finland this usually means going through the execution procedure where the lender must probate its receivables. There are laws which control the default situations in the case of bankruptcy (of a company) or insolvency (of a private customer). If the obligor cannot honor all of his payments the obligor loses control of his assets. In that stage, an independent agent starts to settle the payment obligations by using available assets as good as he can to settle obligations. The bankruptcy code ensures that all creditors are treated equally. (Schönbucher 2003, p.1)

In this ongoing study we define a default as a situation when the collateral has been taken back. This is because our target company usually writes credit losses for accounting after the collateral is realized. In such case the receivable is in thread and has no collateral anymore so according to the accounting norms it should be written as a credit loss. Those credit losses are the dependent variable in this study. This means that the recovery rate mentioned above does not have any influence on the dependent variable at all. In this study we measure realized credit risk by calculating how much each contract has caused credit losses to the target company. This numerical value of the realized credit risk is easy to pick from the accounting database and add to the research material. The interesting question is; which attributes have caused those credit losses? We think this through by using Figure 3 which is made by Stephanou and Mendoza (2005, pp. 7) to their working paper for the World Bank.

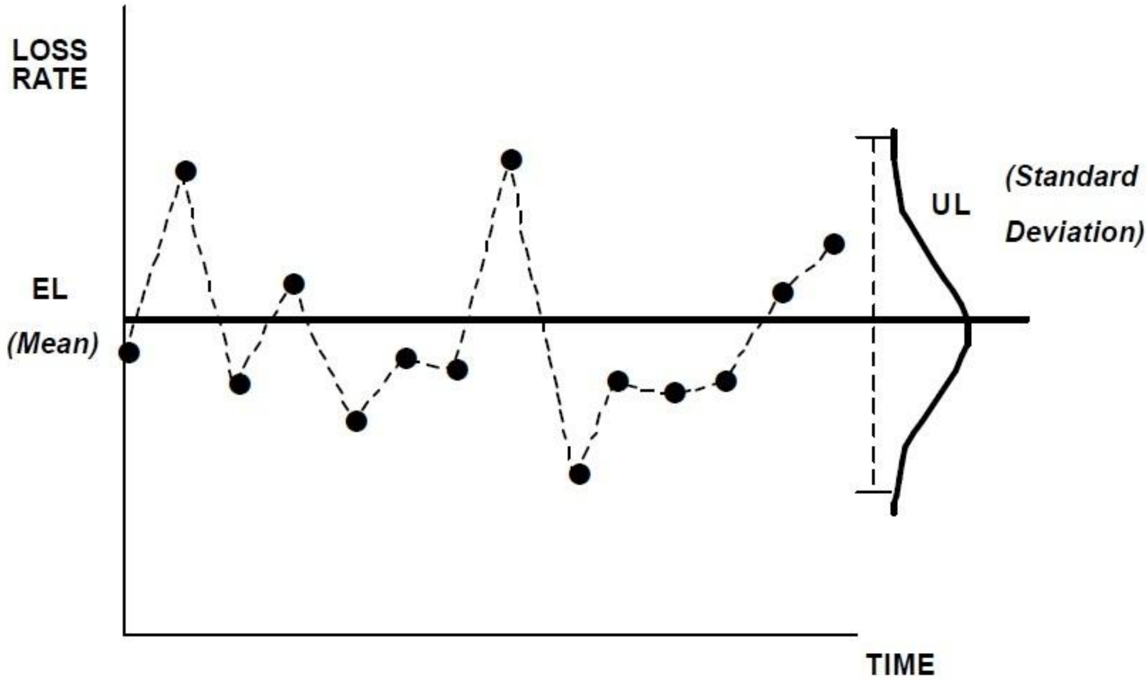


Figure 3: Expected and Unexpected Loss (Stephanou & Mendoza, 2005)

As we can see from Figure 3 the EL (expected loss) ratio represents the mean of the credit losses and thus it refers to the quality of the whole loan portfolio. Therefore it is a basic and fixed feature of the lending activity. The EL actually does not constitute any risk by itself because if losses are always at expected level there is no uncertainty. Therefore that amount should be considered as a natural cost of doing business. The actual risk arises from the attribute UL (unexpected loss) which constructs of the standard deviation of the expected losses. That is the area where the real uncertainty exists. (Stephanou & Mendoza, 2005 pp. 6-7) As we stated in the last chapter (2.1) the risk arises from the change which result is not visible. In this regard the UL ratio is the one which tells the true amount of risk.

Credit losses fluctuate naturally over time and that fluctuation is possible to be measured by statistical methods. PD (probability of default) is usually specified on a one-year basis. It tells for example that financial institution expects 1 % of its loans to end up default. EAD (exposure at default) tells us how much the borrower has owned at the moment of default and LGD (loss given default) refers to the amount of borrowed money the lender will lose in the event of default. The LGD is usually presented as a percentage of the EAD and when determining it one should ask: Do we have collateral to liquidate? How much time does recovery process take and how much work it requires? (Stephanou & Mendoza, 2005 pp. 7-12) Sometimes it is possible to reorganize the assets of the borrower and through that they offer a partial payment for the lender (Hull 2007, pp. 293). This obviously increases the recovery rate and thus decreases the LGD. Using attributes mentioned former it is possible to calculate the amount of the expected loss EL as follow:

$$EL = PD \times EAD \times LGD \qquad (1)$$

The variation of the credit losses is caused by EAD and LGD (see Formula 1; Stephanou & Mendoza 2005). Those ratios focus our attention to the amount of principal loaned to one borrower and also to the attributes which affect to the recovery rate. In our study all loans are secured with collateral. That is why the recovery rates are quite high immediately after default and through execution procedure they get even higher. But how about the EAD? In our case (car loan context) the only way to control the exposure at default is a down payment since there are no other tools to control the financed amount. In this light we should pay a special attention to the down payment or the LTV (loan-to-value) ratio in our own data analysis.

2.3 The lending process

A technical perspective to the lending process is that there are two principal parties and relatively straightforward series of actions involving them. These actions lead from the original loan application to the repayment of the loan or to the situation of default. (Kocenda & Vojtek 2009, pp. 1)

When credit granting has grown over a time the decision making has had pressure to become faster. In the credit industry there are usually two different points when creditors inspect their clients; screening and monitoring. The screening process starts usually when a credit application arrives. The traditional way of screening is just to base the credit decision to the expertise of the credit analyst who looks at the former credit events and tries to estimate the risk of default by comparing the details of new applicant with previous ones. (Hand & Henley 1997, p. 524)

According to Rose and Hudgins (2010, pp. 522), the lending game has become a sales position. Lenders consider quick credit decision making process as a competitive advantage (Alhonsuo et al. 2009, pp. 233). The ball is in customers’ court. The customer has the power to choose whether to fill the lending application or not. And if he does, the decision must come quickly. In a car loan industry the sales position is very visible because there is a person with sales incentives in between the end customer and the finance company. His target is to get the deal done and repatriate his bonus. If one finance company does not give a positive decision immediately he will offer the application to another firm in no time. This arise the problem of the asymmetric information, not only between the finance company and the end customer, but also between the retailer party and the finance company. This should be taken into account in the screening process with reasonable criticism.

Due our practical experience we divide the screening process into two categories; screening of private customers and screening of business customers. Private customers' screening bases on credit status inquiry, which is offered in Finland for example by the credit bureau Suomen Asiakastieto Oy. From the database of Suomen Asiakastieto it is possible to get information of the creditworthiness of the customers. Basically they give information if customer has caused defaults for other financial institutions earlier or not. In addition screening process usually includes some wage information and of course information of earlier payment behavior (or monitoring information) if available. Altogether; information from credit information register has the biggest part in preliminary screening of the private customer.

The screening process of company customers bases naturally on financial statements. In financial statement there are three most important things to solve out; solvency, liquidity and financial flexibility. Solvency means that company's assets exceed its liabilities. Maybe the most known ratio for solvency measurement is the current ratio which basically indicates the coverage that short-term creditors would have if current assets were liquidated.[1] Company is liquid when it can avoid undue costs by paying bills on time. Financial flexibility refers to company's financial leverage, dividend policy, asset efficiency and profitability which should be in line with its estimated growth in sales. (Maness and Zietlow 2002, pp.23) In addition to financial ratios the age of the company and the default information of persons in charge are vital details in practice. In company

[1] $\text{current ratio} = \frac{\text{current assets}}{\text{current liabilities}}$

screening, the register of previous defaults is usually taken into account in rating made by credit bureau. Rating does not necessarily tell any direct information of previous defaults but it tells the probability to cause new ones.

After all, consumer loans are far more complicated to evaluate than corporate loans. Initially this sounds irrational but it is true. It is quite easy for consumers to hide crucial information of themselves, concerning for example health or future employment prospects, when they make a credit application. For corporate loans there exists lots of regulated information available such as financial statement. (Rose and Hudgins 2010, pp. 596) Even if they do not give all the information needed there still are lots of signposts which tell the overall situation of the company. But what do we have for consumer loans? Maybe some kind of salary report which has no regulations what so ever so there might be lots of variation between different kind of reports and thus the reliability of them is doubtful. How about report of consumer’s health or development of his marital status? It would be not appropriate even to ask that kind of statements. This refers to the term of moral hazard. Usually the longest possible maturity and the highest possible last payment in the contract reflect something that consumer does not want to tell. Luckily, consumer loans are usually smaller when compared to corporation loans and therefore usually the better diversification covers the lack of information.

We sum up the lending process by introducing six steps for lending made by Rose and Hudgins (2010, pp.522-524):

1. Finding suitable customers
2. Evaluating customer’s character and sincerity of need
3. Site visit and credit record check
4. Evaluating customers financial condition
5. Assessing collateral (if needed) and signing the contract
6. Monitoring the compliance of the contract

The problem of asymmetric information focuses on step one and two which is discussed earlier. Those steps, in our context, are executed by retailer companies of cars. It means that the customers are found by them and they also filter the information which goes further to the credit analyst in the finance company. This information has a crucial part to find out whether the borrower has a serious intention to repay or not. Previously we introduced the database of credit bureau as a crucial part of credit decision making in Finland. This refers to the step three in the list of Rose and Hudgins (2010). Step four and five are related to the screening process which ends to the decision of whether the loan will be granted or not, with or without additional collaterals.

In the list of Rose and Hudgins (2010) they have monitoring as the last step of the lending process meaning that the credit analyst cannot but the loan contract on the shelf and forget it after the contract is signed. Monitoring or predicting usually means the credit worthiness estimation by using behavioral or performance scoring which refers to already existing information of applicant (Hand & Henley 1997, p. 524). From this point of view a credit analyst should observe the performance of each loan and monitor the compliance of the contract to get information for future credit decisions.

2.4 Credit ratings

The idea of the whole rating system is that the rating itself is a very straightforward opinion of the creditworthiness described only with a few symbols. Even though the rating itself is usually presented in very simple way the assumptions, considerations, judgments and reasoning behind the rating might be complex and usually those are also public. In the international level there are three major rating agencies; Standard and Poor’s Rating Services, Moody’s Investors Service and Fitch Ratings (Wyss 2009, pp. 534). The best rating the firm can get from Moody’s or S&P’s is Aaa or AAA which means that the firm has almost no chance of default. (Hull 2007, pp.289) In Finnish level the biggest rating agencies are Suomen Asiakastieto Oy and Bisnode Finland Oy. Every agency has their own way, criteria and scales to assess the creditworthiness of firms and their probabilities to default. Some of them even offer forecasts of potential recovery rates in the event of default.

The demand for credit ratings from creditors' side is understandable but why do companies which are not creditors request a credit rating from the agency? Investors, especially public funds, want to see established opinion of the quality of some security or a firm before their investment decision. They want an opinion from some reliable party which is not the issuer or underwriter of the security. How reliable are those ratings? It is a fact that in the long run, securities with higher credit ratings have had lower default rates when compared with low rating securities. But in the end of the day ratings are just opinions. This means that a rating does not remove the need for the investor to understand what he is buying or to where he is investing. (Wyss 2009, pp. 534) That is to say; rating is only a part of the screening process, not the process itself.

One noteworthy issue concerning the reliability of the credit ratings is the earning logic of the rating agencies. Most of the big ones make their turnover, in practice, by selling the ratings to the companies. This means that companies and institutions pay to the agencies to get rated. Accordingly the agencies do not charge the users of the ratings. Actually the end result of the rating process might be public. Does not this earning logic make ratings slightly unreliable? This is a reasonable question especially in the aftermath of the latest financial crisis. There are also advantages with that logic. When companies have incentives to get a good credit rating they are willing to give some non-public, detailed information or confidential data, about their business which investors would not otherwise get (Wyss 2009, pp. 534). When smaller agencies are concerned the earning logic is different. Companies do not have a huge demand for rating from agency which is not internationally remarkable so they are not willing to pay to get analysed. On the other hand, they do not have any incentives to reveal any extra information to the agency if they do not think the bureau is a significant one. In that case also the information used in the rating process is only public information which is available for investors anyway. Because of this the only additional value agencies can afford to subscribers of ratings is the analyzing work and that is pretty much the source of income for those rating firms. For example in Finland, according to our phone discussion with the customer servant of Suomen Asiakastieto the earning logic bases on the payments from the subscribers. Only way to collect income from rated companies is to give them right to use “The strongest in Finland” slogan in their home page or in other marketing material. Slogan means basically that the company has a rating AA+ or AAA so it has very high creditworthiness. However, the main earning logic is to provide ratings to subscribers using public data.

How a rating is assigned? Basically agencies try to answer for one question: What is company’s ability and willingness to repay their obligations in the future, relatively. Analysts consider a wide range of business and financial risks that may interfere with full payment and try to make a forecast of company’s future position. That position they compare with other businesses to evaluate the relative credit risk of the firm. Most agencies use a combination of quantitative and qualitative analysis so they do not just analyse historical data and try to figure out future position only by staring at a rear-view mirror. In addition, after rating is made it is not static. It will be reviewed and updated on regular basis. Agencies give messages and warnings to the market about the direction in which the rating may move.

2.5 Credit scoring

To make comparing of credit worthiness possible creditors must give some numerical value of credit worthiness to the applicants. Usually that is made by scoring. Credit scoring refers to the formal process of estimating how likely applicants are to default with their repayment. (Hand & Henley 1997, p. 523) Credit scoring models are statistical devices such as scorecards or classifiers, which use predictor variables to estimate the probability of delinquency. The most commonly known traditional scoring model was the multiple discriminant credit scoring analysis which was made by Altman in 1968. (Altman 1968; See chapter 3.1) Nowadays probabilities are usually formulated with statistical methods like linear regression, probit regression, logistic regression, discriminant analysis, neural networks or decision trees. The final decision is usually made by comparing the probability with the adequate threshold. (Hand & Henley 1997, p. 524)

The reason for development of the credit scoring is the demand for faster decisions, whether or not to grant a credit, together with a possibility to use computers to automate the decision making process. The real use of credit scoring began in the 1960's when credit card business became significant creating a demand for an automatic decision making system. (Kocenda &Vojtek 2009, pp. 2) Although originally scoring models classified applicants by default potential based only on an ordinal ranking. However they were the original precursors to the later numerical PD (probability of default) estimations. (Stephanou & Mendoza, 2005 pp. 8) Because of the development of credit scoring the loan delinquency rates have lowered twenty to thirty percent compared to credit companies which use only credit analysts' judgment in making credit decision. Credit scoring has also increased the borrower acceptance rates as well as decreased the average time of credit decisions. Over a time scoring systems have also turned out to be objective and avoid personality clashes between lenders and borrowers. (Rose and Hudgins 2010, pp. 603) Those systems process only with numbers, not with intuitions or feelings. Straightforwardness is an advantage and a disadvantage at the same time.

These days nearly all lenders use credit scoring to evaluate credit applications. The advantages of credit scoring models are their ability to handle a large volume of credit applications fast. That is why credit card companies such as VISA and Master Card are the heavy users of the scoring systems. They need a high amount of credit decisions in minimum time with minimum labor. Same concern insurance companies which nowadays get most of the applications through internet. Lenders have a cutoff level and if applicants' credit scores fall below it, the credit is likely to be denied. The most important variables used in the credit score evaluation for consumers are the credit bureau ratings, home ownership, income level, number of deposit accounts owned and occupation. A credit company can give weights for different attributes in the analysis and change those weights due the continuous testing. Testing is a crucial part of proper credit scoring system because the economic change is constant and abrupt. (Rose and Hudgins 2010, pp. 599-600)

2.6 Credit contracts

In this study we define a word credit as an amount of money which is loaned to a customer by financial institution and which must be repaid, with interest, usually in contracted instalments (Hand & Henley 1997). Terms loan and credit go hand in hand in the literature. Alhonsuo et al. (2009, pp. 229-230) defines credits as the major concept and loans as one of the minor sections. The most important difference between those two terms is that credit refers to some kind of trust and it does not necessarily need any money involved to occur. A loan instead refers usually to a contract where someone gets money from some other party and has an obligation to pay it back afterwards.

There exist many different credited relationships and contracts. Historically loans are the oldest way of borrowing money. It is a bilateral contract between the borrower and the lender which includes a sum (principal) and an agreed payoff stream with an interest payment and a certain maturity. (Schönbucher 2003, p.10) Traditional loan contracts are fixed term and the loan is supposed to be totally repaid after a certain time. In such case the instalments are calculated to include both principal and interest and usually the amount of instalments is given ahead. In such case the whole contract is settled by following the agreed payoff stream. (Hand & Henley 1997, p. 524)

Most of the credits (86% of the data) included in this study are granted for consumers. Consumer loans in Finland usually includes mortgage loans for buying a house, consumption loans for purchases and loans guaranteed by government for studying (Alhonsuo et al. 2009, pp. 229). Over the past couple of generations lots of people have adopted the way to borrow money to supplement their income and enhance their lifestyle. This phenomenon has made loan grating to increase with explosive way (as Figure 1 shows in the Introduction part). The cyclical nature and individuals' bizarre attitude on interest rates make consumer loans very profitable but also very risky business. The cyclically sensitiveness arises from the fact that consumers tend to reduce their borrowings when the pessimistic attitude against the future raises. Of course this holds also vice-a-versa. The bizarre attitude on interest rates (or interest inelastic) comes from the fact that individuals seems to be far more interested in monthly payments required by a loan agreement than the actual interest rate charged. This gives the finance companies opportunity to charge quite sticky interest rates. (Rose and Hudgins 2010, pp. 589-594)

Among normal consumers, discussion of consumption credits has sometimes demeaning tone. That is of course because high interest rates of them (when compared with secured loans) but also because the word “credit” refers so strongly to credit cards and thus to the misusing of them. (Peura-Kapanen 2005, pp. 46) This is interesting because in the large scale, also mortgage loans for houses are a part of the consumption credits and those loans seem to be a natural part of the everyday life for most of the people. One reason for a bad reputation and a negative discussion in media might be predatory loans which are also known as subprime loans with high interest rates and other expensive covenants. The subprime loan refers to lending for customers with limited financial resources and short or poor credit histories. Those customers have usually lower and quite volatile income level and fewer assets when compared with prime borrowers. Sometimes lenders even encourage borrowers to grant loans even if they notice their financial incapacity. (Tarun and Heitfield 2004, pp. 457)

Corporate loans (14% of the observations in the dataset of this study) can be divided into three categories by the maturity: Short (under 1 year), medium (1-5 years) and long (more than 5 years). The separation of the corporate loans can be made also by the purpose of the loans to three categories: investing, asset and operational loans. (Alhonsuo et al. 2009, pp. 238) We define corporate loans in our data as medium maturity operational loans because the purpose of them is to get a vehicle for the company and the maximum maturity which our target company offers is 60 months.

2.7 The nature of Finnish auto loan market

The most usual way of getting a loan for a car is an instalment loan. In such context instalment loan is defined as a loan where the ownership of the purchased items transfers only when borrower has met all of his payment obligations. Usually the retailer of the purchased item (for example automobile) makes the instalment contract with the end customer. Then the retailer transfers the contract to the finance company which pays the principal to the retailer. After that the end customer pays the agreed instalments to the finance company. (Finanssialan keskusliitto 2010, pp.3) Usually in Finnish car loans the interest is fixed term. The interest rate is like a part of the whole deal, in same way as for example tires of the car or other equipment, so the riskiness of the customer himself does not determine the cost of the capital as much as in other kind of loans.

In 2011 car retailing was responsible for 14% of the whole trade turnover in Finland (Statistics Finland 2011). It means approximately 17.5 billion euro per year. Traditionally in Finland banks have financed more of those car purchases than finance companies. In the year 1985 commercial banks offered more than 90% of consumer loans. Between years 1985 - 2010 finance companies entered into the credit market. In 2010 commercial banks offered only 69% of consumer loans. (Finanssialan keskusliitto 2010, pp. 5) This phenomenon is also visible in the car financing industry.

Howells and Bain stated in the year 1998 that in Scandinavia nearly all intermediaries which are not banks themselves are closely connected with banks or they are direct subsidiary of some bank. They called this Scandinavian approach to financial services as the “all-finance” approach which is a consequence of financial market deregulation and integration in the 1980s. The phenomenon, stated by Howells and Bain (1998), is visible still more than a decade later. Most of the finance companies operating in the auto loan market in Finland are owned by some bank which has outsourced their consumer loan services to a subsidiary company. There exist also a few finance companies which have other extra incentives to finance cars such as companies owned by a car manufacturer. Those finance companies have basically two extra incentives for financing, in addition to making profitable business; to advance the selling of the parent company by providing loans and also to get opportunity for customer relation management. That is actually the biggest difference in those two groups because the finance companies owned by the banks are usually intended only to satisfy customers’ need for purchases so they do not have much interest for the purchased product.

There exist also finance companies without any connection to banks or manufacturers of purchasable items. The reason why most of them have a bank or other remarkable organization as a mainstay is the price of the money which they have to borrow from the open market. Reason for money borrowing is that non-banking institution is not available to take deposits from the public so they have to get money for lending activity other ways. A large well known and stable bank or manufacturer behind the finance company decreases the cost of the borrowed capital by decreasing the risk of default and thus makes easier to operate profitable credit business.

2.8 Secured instalment loans for automobiles

In this study we investigate instalment loans which are secured with the ownership of the collateral object. The features of such loans are described next.

Secured loans are said to be the oil of the economy and the engine of the growth. That is because collaterals encourage lenders to offer such loans that would not otherwise be available. (McCormack 2004) In a secured loan there is a collateral object, an automobile in this study, which belongs to a lender in the situation of default. The collateral decreases the risk level of the lender because they can cover the lost principal (or at least some of it) by selling the collateral. Through that way they can offer lower interest rate for borrowers and make a transaction profitable for both parties. For example in the real estate market, where the loan amounts are high, the collateral is the component which makes transactions possible. Collaterals are a kind of answer for the problem of the asymmetric information which exists always when a credit decision is made. (Hyytinen & Pajarinen 2005, pp. 25) The collateral also encourages obligor to exercise the agreed payoff schedule, basically because a rational obligor does not want to lose the collateral asset.

Another way to secure a loan is a personal guarantee. It means that someone else than the obligor himself commits to meet the payment obligations if the original obligor cannot do so. (Alhonsuo et al. 2009, pp. 234) Secured instalment car loans have a privilege against other loans which have no collateral included. When firm goes to a bankruptcy or private customer goes into a loan arrangement, secured instalment loans are defined as B-loans by the arrangement trustee. The B-loans have a privilege against C-loans which have no collaterals. It means that the lender has a right to get its collateral out of the bankrupt's estate without any principal cut. (Juridicial system in Finland 2012)

An instalment loan refers to a transaction where the buyer pays the price in agreed instalments. Usually such loans are employed to buy big-ticket products such as automobiles, home appliances or furniture. (Rose and Hudgins 2010, pp. 591) Instalment loans are one good example of fixed term loans mentioned before because in the instalment loan the agreed payoff stream includes all of the obligations borrower has. Usually the seller keeps the right of the ownership until all the agreed payments are settled. As a define instalment loan must lead to the transfer of the ownership at the end of the maturity. (Finnish law of instalment trade 18.2.1966/91 §1) Otherwise it refers to some kind of rent or leasing contract.

Loans included in our data are secured instalment loans but they have some special features which are introduced next. Collateral is an item with value which gives a support to the borrower’s ability to repay the loan (Rose and Hudgins 2010, pp. 695). In our research all loans have collateral which is a car in most cases but can also be a motorcycle, a caravan, a camper van or a light truck. The security is not the car itself but it is the legal right for the lender to take the car back for finance company if borrower does not meet his obligations. In practice the customer is the official holder, not the owner, of the car as far as he can meet his payment obligations. When all payments are done the lender sends required documents to the customer for the official ownership registration.

Usually instalment loans are annuity loans where all instalments are same sized. In the loan portfolio of the target company there exist also some contracts with larger last payment. The bigger last payment is also called a salvage value which means in this context only a bigger payment in the end of the contract and has no suggestion to the value of the car. This possibility is made because of some of the cars are so expensive that suitable monthly payment requires so heavy down payment that many customers cannot afford it. By transferring, for example 20% of the principal, to the end of the contract it makes possible to amortize only a part of the whole principal at a time and to leave the rest of it to the future. This helps the borrower to get a monthly payment at a reasonable level. Of course customers have to pay interest for the whole principal all the time so this possibility might be comparatively expensive. On the other hand it makes possible to buy an expensive car and amortize first for example 80% of the principal and the rest of it in the future. From the finance company point of view this kind of contract is a two-sided question. If the finance company decides to finance also the bigger last payment in smaller instalments it has good information of the customer's payment behavior for the new credit decision because the customer has already paid most of the contract. Especially if the contract has been made in reasonable way in the beginning, the last payment should be less than the value of the car in the end of the original contract. In that situation the value of the collateral makes it easy for the finance company to accept a new credit contract even if the historical payment behavior of the customer is a little bit poor. On the other hand, a larger last payment transfers cash flows in to the future which obviously adds risk from the lender's point of view.

A down payment is usually required for all kind of secured loans. The down payment decreases the financed amount and thus makes it more equivalent with the value of the collateral. Our target company requires down payment to all of its secured instalment loans (a few exceptions exist). Usually required down payment is between ten and thirty percent calculated from the market value of the automobile but in practice it can be almost anything between zero and ninety percent. Sometimes the portion to be financed is only a small part of the value of the car and in that case the finance company only supports the solvency of the customer for a few months. We have discussed in this paper several times of the LTV (loan-to-value) and its role in the credit market. LTV is the amount of loan compared to the value of the collateral. It can also be calculated by taking the amount of down payment from the market value of the collateral.

Last special feature which we introduce is a prepayment of the instalment loan. It means that the obligor has a right to meet all of his payment obligations before the maturity whenever he wants. In that case the finance company loses interest payments from the rest of the maturity because it is not allowed to charge them from the customer.

3 PREVIOUS STUDIES

In the pages that follow, we introduce previous studies concerning default situation measurement and credit risk evaluation from many different viewpoints. First we introduce a few articles measuring default probabilities at a common level and in the second section of this chapter we delve into the details of the car loan default estimations. In the end of this chapter is table 1 which captures all previous studies discussed in this chapter.

3.1 Common models to explain default

The classic of the default probability estimations is from the late 60's. It was first published in 1968 by Edward Altman titled "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy" and has since achieved considerable scholarly and commercial success. Edward Altman calculated his model by using a sample of 66 publicly traded manufacturing companies. Half of the companies failed and went bankruptcy during years 1946 – 1965. Half of them did not go bankruptcy and they were still in existence in 1966. From this sample Altman developed the Altman's Z-score formula which is basically a formula to estimate the probability of one firm to go to bankruptcy in next two years. Originally Altman got 22 variables which he tested separately at first. Interestingly the most significant ones did not end up in the final discriminant function which was a result of numerous computer runs analyzing different ratio-profiles to find out which does the best job estimating the PD (Formula 2):

$$Z = 0.012 \cdot X_1 + 0.014 \cdot X_2 + 0.033 \cdot X_3 + 0.006 \cdot X_4 + 0.999 \cdot X_5, \quad (2)$$

In Formula 2 the explanatory variables are simple accounting ratios:

X_1: Working capital/ Total assets
X_2: Retained earnings / Total assets
X_3: Earnings before interest and taxes / Total assets

X_4: Market value of equity / Book value of total liabilities
X_5: Sales / Total assets

The interpretation of the Altman's Z-score formula says that if the Z-score of the company is more than 3.0, the company is unlikely to default. The "alert-mode" should start when the company is in the level 2.7 - 3.0 and in the level 1.8 - 2.7 the possibility of default is significant. If the company gets Z-score under 1.8 the probability of financial embarrassment is very high. Altman used companies' ratios as explanatory variables and had a motive which came remarkable for whole the PD estimation science:

"The question becomes, which ratios are most important in detecting bankruptcy potential, what weights should be attached to those selected ratios, and how should the weights be objectively established." (Altman 1968, pp. 591)

In our study variables present some quality of the loan contract or customer, not some business indicator like in Altman's model. But still, as Altman said; the most important thing is first to find out variables which does the best overall job together and then find out correct weights for them. Good example of this logic is "Sales/ Total assets" ratio which would not have appeared at all in the model based on the statistical significance measures. However, that ratio had a unique relationship to other variables in the Altman's formula so it ranked second in its contribution to the overall discriminating ability of the model.

Altman tested his model with six ways. First he estimated bankruptcies and non-bankruptcies from initial group of firms. The model estimated 95% of firms' end-statuses correctly. Next he tested the model with values taken from same companies but two years prior to bankruptcies. The model estimated bankruptcies with 72% accuracy and non-bankruptcies with 94% accuracy correct. Finally he end up using five years old data for same firms and the forecasting accuracy of bankruptcy declined to 36%. It seemed that after the second year, the discriminant model became unreliable. Three other tests he made were related to secondary samples. He took ratios from other firms and put them to the initial discriminant formula. Surprisingly he got better accuracies than with initial data. Altman's main findings were:

1. All of the observed variables showed some deteriorating when bankruptcy approached
2. The most significant deteriorating in the majority of these ratios occurred between the third and the second years prior to bankruptcy.

In 2010 Altman and his colleagues Rijken, Balan, Forero, Watt and Mina released an updated version of the classic Z-Score model. In 2012 Altman and Rijken tested that model successfully in their study: "Toward A Bottom-Up Approach to Assessing Sovereign Default Risk: An Update". It was published in International Research Journal of Applied Finance in the year 2012. Altman et al. (2010) formulated a new Z-Metrics™ approach to estimate the median probability of default for one and five year horizons for nonfinancial companies by using a sample of more than 260 000 observations (financial statements, macro economic data and market prices). That model was a logical extension of the Altman Z-Score technique. It was not the first update of that paper but because of its topicality, we state a few main points of it. In the paper published in 2012, Altman and Rijken measured the default probabilities of listed companies in Europe and U.S.A 2009-2010 by using previously introduced Z-metric model. Their goal was to solve out the sovereign risk in those areas. The motivation for their research was the current financial crisis which speed and depth surprised, strange to say, especially the credit rating agencies.

To form a new model, Altman et al. (2010) used multivariate logistic regression which they formulated by using dozens of variables representing accounting ratios, operating firm specific information and also some macroeconomic indicators. After all, they selected 13 variables to produce a credit score for each public company which they later converted as PD (probability-of-default) ratios. The model outperformed not only the credit agency ratios but also the old Z-Score model when tested with out-of-sample data in 2012. In the Z-metric rating system they had 15 rating categories from the top “ZA+” rating to the lowest quality “ZF-“ rating. As a result of Altman et al (2010) study they made a model which included market variables (and fundamental), using trend and static measures combined with macroeconomic variables.

Another way to investigate default situations is to observe the duration of credit contract before default and to find out reasons for that event. Okumu, Mwalili and Mwita (2012) used survival analysis to find out if the gender of the borrower has explanatory power on the survival time of the contract. They did their test in Kenya because especially there the financial institutions tend to use only credit scorings to rate their customers whether they are good or bad loan applicants. In Kenya and the larger African continent both practitioners and scholars showed insignificant attention to the credit risk analysis so Okumu et al. (2012) decided to investigate it. Survival analysis is a statistical method for estimating the time to some events such as default of a credit contract. Models do not only estimate the probability of default but also the most likely point of time for default to happen.

In the test of Okumu et al. the time-to-default (T) was defined as a random variable. It was countered from the beginning of the loan contract. The objective of Okumu et al. study was to use a product-limit survival model to generate default probabilities at several points in time for two risk groups (males and females). Their data was from one lending commercial bank from Kenya and it included 500 personal loans with maturity of 30 months from the period January 2007 to June 2010 (half of observations were men and half were women). Okumu et al. used Kaplan-Meier estimator to form survival curves for the risk groups. To find out whether they are statistically different or same they used the log-rank test.

The results showed that 11 out of 250 male borrowers defaulted with average time of 15 months. From 250 female observations 13 defaulted with average time of 13 months. Four male borrowers and six female borrowers settled their loan before the maturity. The survival curves were statistically similar with 95% confidence interval which tells us that it is not reasonable to classify borrowers on the basis of gender because it does not seem to affect credit risk.

As mentioned earlier the most commonly used method for default models is the logistic regression which was also used by Kocenda and Vojtek (2009). Another method they used was Classification and Regression Trees (CART) which is a little bit less known method in this context. The goal of Kocenda and Vojtek was to compare the methods in terms of efficiency and power to discriminate between the low and high risk clients. The data of Kocenda and Vojtec included 21 variables, mostly socio-demographic but also financial and behavioral variables, and 3403 observations. The dataset represented retail loans taken from a bank which operated in Czech Republic. Loans were used mostly to real property purchase and reconstruction during 1999-2004. A noteworthy detail is that almost 50% of the loans in the data defaulted. Koceda and Vojtek defined loan as a bad or a defaulted one if the borrower was more than 90 days overdue with any payment connected with the loan (so the definition of the default was much tighter than ours).

At first, Koceda and Vojtek took in the logistic regression all variables with information value more than 0.1. The reason for such a low requirements was that they wanted to include as much socio-demographic variables as possible, even they tend to exhibit lower information values. They made a forward-backward stepwise model selection by using the Akaike Information Criterion (AIC) to select the best model and the maximum likelihood method to estimate coefficients. To test the performance of their model they split randomly their data into two categories. Two-third of the dataset they used for development of the model and one-third to validation. The quality of the models they tested by using the Receiver Operating Curve (ROC) and the GINI coefficient.

The results of the logistic regression showed that the resources of borrower and the level of education were the most significant variables to explain the probability of default. The length of the customer relationship between the borrower and the lender was the most important behavioral characteristic which highlights the importance of the monitoring. The marital status variable was also significant and its role was explained by the dual income effect of couples who are married. From the lender's point of view the married customers are less risky because their partner is seen as a guarantor for the loan. The purpose of the loans was also a reasonable variable to take in to the model because it showed signs that loans taken to build a new house were less risky than loans to renovate an old one. This was explained by the superior financial potential those new-house-builders have to the renovators. Interestingly Koceda and Vojtek found both credit scoring variables insignificant.

The advantage of the Classification and Regression Trees (CART) approach in estimating the PD is that it is very easy to explain to the management and it has an ability to deal with missing observations. The CART is a non-parametric approach. It consists of several layers of nodes. The first layer includes a root node and the last layer includes leaf nodes. It is a binary tree so every node (expect the leaf nodes) is related to two other nodes. Interestingly the variables which the CART approach found most significant are exactly the same which logistic regression found most important.

After all both methods were robust, parametric and non-parametric, revealing that the most important explanatory variables were: Customer's educational level, the amount of resources customer owns, marital status, the length of the customer relationship with the bank and the purpose of the loan. Their results stated that the socio-demographic variables should not been ignored from a proper credit scoring model.

3.2 Models specified to determine auto loan defaults

In previous studies there are also studies made especially to explain defaults in auto loans. One of those studies is “Empirical Examination of Drivers of Default Risk in Prime Auto Loans” made by Kocagil and Demir (2007) where they tried to find out the empirical drivers of the default risk in the case of the U.S. prime auto loans. The data included details from about 500 000 auto loans originated in the first-quarter 2000 and financed by the major prime auto loan issuers from the United States. The Quantitative financial research analyst PhD Ahmet E. Kocagil and asset-backed securities specialist Ebru Demir who made this analysis for FitchRatings focused on two key areas; bias resulting from a particular regressor “new or used car” and the presence of adverse selection in subvented loans.

The variable which indicates whether the financed object is a brand new or a used one causes usually biased result because the default risk equation and the used car selection equation seem to correlate with each other highly. This refers to a self-selection which means that the variable which defines if the car is a new or a used one is a endogenous one. Endogenous in this context means, that the choice to buy a new or a used car includes hidden information of the customer’s income level. Due that fact Kocagil and Demir used a two-staged analysis where they had a proxy indicating whether the financed car was a new or a used one. Other thing where FitchRatings’ team took notice was subvented loans. In those contracts car manufacturers want to provide usually lower APRs (annual percentage rates) for borrowers with a low credit risk and gain sales through that way. Usually subvented customers have high bureau scores and that is why they usually do not cause as much defaults as other contracts with same attributes. However, according to Kocagil and Demir, subvented contracts still cause more that 25% of defaulted loans. To address this problem Kocagil and Demir decided to enter a subvented loan variable in to the regression model along with the APRs variable. In this way they were able to point out subvented contracts whit low bureau scoring which seemed to have higher default risk than an average contract.

In their regression model Kocagil and Demir used many explanatory variables for example loan-to-value ratio (LTV), credit bureau score, APR, the subvented loan indicator, the proxy for the new/used car indicator and also characteristics which gave information of possible delinquency and charge off. The FitchRatings' team carried multivariate analysis through both logit and probit regression estimations (maximum likelihood estimations) which both gave similar model discrimination power and coefficient significance. In other words they tried to explain the probability of whether the default situation occurs or not. So there were many continuous variables to explain one variable which can have only two results; yes or no. As a result Kocagil and Demir found out that the LTV ratio (loan-to-value) was the most influential risk factor. According to Kocagil and Demir the LTV ratio seems to be a powerful explanator for defaults in secured consumer loans with instalments but they also highlighted its huge role especially in auto loan context. Another main result was the function of the APR variable which was huge, no doubt of it, but the information inside that variable and also the noise resulting from it made its role a little mystery. The weight of the APR variable dropped in the estimation when scholars added the new/used proxy variable into the model.

In addition to default, prepayment has also been a popular explanation target in auto loan models. For example, Agarwal, Ambrose and Chomsisengphet tried to form a model to estimate the probability to prepayment or default of the automobile loans in 2008. They made their study (Agarwal et al. 2008) "Determinants of Automobile Loan Default and Prepayment" for the Federal Reserve Bank of Chicago and they tried to explain default or prepayment by using information of the customers' consumption choices. A self-selection as a given attribute in the auto loan market is widely recognized. Some parties see it as a reason for biased results (see previous study of Kocagil and Demir) and thus try to eliminate its influence. Other parties (for example Agarwal et al.) use it as a main explanatory variable in the whole test.

Agarwal et al. used data which included over 20 000 loans from one financial institution which offers direct auto loans to customers. The different between direct and indirect auto loan is that in indirect auto loans there is an auto retailer between the financial institution and the end customer. In the direct auto loan financial institution and the end customer have their negotiation about the loan together. In the study of Agarwal et al. (2008) they got auto loans issued with 48 month and 60 month maturities as well as fixed rates. They took their data from the United States and observed the performance of the loans from January 1998 to March 2003. In their data there were 4730 observations with prepayment and 534 defaults. Loan details included the value of financed object (ranges from $4,625 to $108,000), the age of automobile, financed amount, LTV (loan-to-value), monthly payments, interest rate, time of origination, payoff time, time for prepayment or default, auto model details (auto model dummy), new/used car, the credit score of the borrower and the unemployment rate in the county of residence of the borrower (a proxy for local economic conditions).

A credit risk of one contract was measured by credit scores (FICO score system in this case). Agarwal et al. did not find almost any differences between contracts with new or used cars. The median FICO value for a customer with a used car was 722. Respectively the median FICO value for a new car contract was 726. The most significant different was the median loan amount which was about two and a half times bigger for new automobiles as compared with value for used automobiles. In their regression analysis the dependent variable was able to get three different values: 0 for current, 1 for prepayment and 2 for default. They regressed the dependent variable against a variety of independent variables which gave information about an economic environment and personal risk factors of customers.

Agarwal et al. constructed a few variables by themselves. First they constructed a prepayment premium which was based on the perfect market assumption which says that a single person should act rationally and pay the loan early if the interest rate is fixed and common interest rates decreases. This means that a decline in the prevailing three year treasury note rate should have a positive impact on prepayment behavior. Another constructed variable they made was CLTV (monthly loan-to-value ratio) which took in to account the decline in the market value of each car model and thus the increase in the LTV (loan-to-value) amount. They expected that an increase in LTV during the loan period refers to the higher probability of default. In other words, they expected that a loan performance is possible to explain with the quality of financed car. As a statistical method Agarwal et. al used a duration analysis to estimate the conditional probability of default or prepayment at a time t. They added predictor variables such as the LTV value and the FICO scores to the analysis because of their supposed capability to explain the termination event. After the probability function was made they determined a survivor and distribution function by using the maximum likelihood method.

The results of Agarwal et al. highlighted the effects of macroeconomic conditions on default and prepayment probabilities. For example weakening macroeconomic conditions in the year 2000 caused more defaults and fewer prepayments. The age of the borrower also had some explanation power. Borrowers below the median age of 40 had a higher probability of default than the older borrowers and the older borrowers had higher probability of prepayment. The constructed variable called Prepayment premium was also significant independent variable and had a positive relationship with prepayment (and surprisingly with default also). Monthly payment and unemployment rate had also a positive impact on both prepayment and default. Monthly income had a positive impact on prepayment but a negative impact on default, as expected. The results showed also that the FICO score and the new car dummy were negatively related to the default probability as for LTV and the used car dummy were positively related to the default probability. Finally, 31 dummy variables which denoted different auto brands revealed interesting results. Loans for premium brands, such as Lexus and BWM had a higher probability of prepayment, while loans for most economy automobile makes had a lower probability of default. Some of luxury car brands did have positive impact on both default and prepayment. Therefore we can state that the quality attributes of the car brand do have explanatory power.

The study of Heitfield and Sabarwal “What Drives Default and Prepayment on Subprime Auto Loans?” was published in Journal of Real Estate Finance and Economics in 2004. They used a novel data including subprime auto loans to estimate a competing risk model of default and prepayment. Their perspective was to compare the subprime auto loans with residential mortgages which, in general, seem to be quite similar nature contacts. Heitfield and Sabarwal had same problem than so many other scholars before in the auto loan papers; The lack of data. Fortunately the development of the loan-backed securities made data of auto loan pools available because some of the finance companies are issuers of backed auto loan securities which are publicly traded. Heitfield and Sabarwal focused to research that kind of financial institutions because public security trading forces them to release performance data of their loans. Using Moody's reports which combine information from SEC filings concerning backed auto loan securities they managed to get a data including 3595 observations which were basically monthly details of auto loan pools. The data presented performance of 124 pools issued by 13 finance companies which were specialized to grant subprime auto loans. Those finance companies offered loans also for other kind of customers than consumers and they originated loans basically using a network of franchises and car retailers. This reminds of the way how the target company of this ongoing research originates its loans as well. The data did not introduce information of single loans but it showed the monthly statistics of loan pools. From those monthly statistics Heitfield and Sabarwal picked monthly rates of defaults and prepayments. The sample included pools issued during 1994-1996 and they included 3.3 million auto loans with average maturity of 60.1 months (in our ongoing research the maximum maturity is 60 month).

As a research method Heitfield and Sabarwal used the discrete outcome interpretation of duration models. Basically they tried to find out why some loans drop out before the end of the month t when assumed that they survived in the end of month t-1. They tried to solve this issue by making hazard rates (which index functions depend on a pool's age, calendar time and the pool’s issuer) using a simple multinomial logit specification and monthly survival functions. Results revealed that the prepayment rates increase quickly with loan age but are not affected by prevailing market interest rates. This means that the subprime lenders seem to have no interest to refinance their loans when prevailing interest rate levels decline a little bit. The default rates seemed to be driven largely by shocks to household liquidity. For example a rapid increase in the unemployment rates seemed to have a huge effect. However the hazard function seemed to be quite flat and there was no significant decline toward the end of a loan's life.

Heitfield and Sabarwal also found remarkable differences between different lenders. Lenders which offered highest interest rates seemed to suffer more defaults than others but also, surprisingly, they also met lower rates of prepayment. Conclusion of that phenomenon is that borrowers seem to be a heterogeneity group of customers and lenders seem to focus on different customer segments. Thus interest, default and prepayment rates vary between different lenders. Heitfield and Sabarwal described the LTV ratio to be more important variable to cause falling in the hazard rates of residential mortgages than in auto loans. This is because in auto loans the value of the collateral drops quickly and in residential loans it drops hardly at all. Naturally the same effect concerns the LTV ratio but other way around. In auto loans the LTV ratio might stay still over a time, and even increase, while in residential loans it decreases in normal situation when borrower make his monthly payments and collateral keeps it value and even increases it. This finding confirms the important role of the LTV ratio, when auto loans are concerned. But it makes us wonder the statement of Minister Antti Tanskanen which is discussed in our introduction part. It stated that the maximum LTV ratio for residential loans should be 80% which means a requirement of 20% down payment. In our sample the average down payment in car loans was approximately 30% (because the average LTV in the whole dataset was 69.16%). Of course in residential loans the principals are in other level but still the 80% LTV ratio sounds slightly oversized assuming that the value of the house should not ever drop as fast as the value of the car in normal market circumstances.

Table 1: Summary of the previous studies

Scholars	Objective	Sample	Method	Explanatory issue	Main explanators	Notice
Altman 1968	Explains probability of a bankruptcy in the next two years	66 companies with assets of more than $1 million (Moody's Industrial manual 1946 – 1965)	Discriminant analysis	Bankruptcy of a firm	Working capital/ Total assets, Retained earnings / Total assets, EBIT / Total assets, Market value of equity / Book value of total liabilities, Sales / Total assets	Model was not suitable for estimating financial companies' probability of default
Altman et al. 2010 (2012)	Probability of default – metric for companies at global level for 1 and 5 year horizons	260 000 observations (financial statements, macro economic data and market prices). 1989-2009 The U.S and Canada	Logistic regression	"Credit event" = a formal default or a bankruptcy	Market, fundamental and macro economic variables	A logical extension of the Altman Z-Score technique
Okumu et al. 2012	Can a gender explain default?	500 loans for consumers. Kenya 2007-2010	Survival analysis	Default probabilities at various points in time	Gender	Gender had no explanation power
Kocenda and Vojtek 2009	How to determine between low and high risk debtors?	3 403 retail loans for consumers. Czech Republic 1999-2004	Logistic regression & Classification and Regression Trees	Probability to end up as a good or a bad loan	Educational level, Customer's assets, Marital status, Customer relationship and The purpose of the loan	Both credit scoring variables were insignificant
Kocagil and Demir 2007	What are the determinants of default risk in prime auto loans?	500 000 auto loans. The U.S 2000	Logit and probit regression	Default (yes or no)	LTV, APR, Credit bureau score, Term at origination, Subvention and New/used car dummy	APR's explanation power dropped due the proxy variable made to correct endogeneity of a used car indicator
Agarwal et al. 2008	Determine default or prepayment of an auto loan by using information of consumption choices	20 000 auto loans.1998-2003 The U.S.	Survival analysis	Default or prepayment	Macro economic conditions, Age/ quality of the car, LTV, FICO, Age of the customer, State, APR and Monthly payment	Loans for most economy automobiles had a lower probability of default
Heitfield and Sabarwal 2004	Determine default or prepayment of subprime auto loans. Comparing with residential mortgages	Auto loan pools (3.3 million loans during 1994-1996)	Multinominal logit and survival fuctions	Default or prepayment	LTV, Duration, Interest rates and Unemployment	Surprisingly, the unemployment rate seemed to be positively correlated with prepayment

4 DATA AND ANALYSIS

The research concerning the PD (probability of default) is usually divided into two categories: market based and structural research. In our study the approach is definitely the structural approach which refers to finding out the characteristics of defaulted loans by using historical data and statistical methods. The market based approach investigates for example credit default swap spreads and the volatility of equity market value to infer the likelihood of default. (Stephanou & Mendoza, 2005 pp. 8)

In our data we have both private customers and corporate customers. However, we analyse them separately because of practical issues. Practical issues in this case mean for example different variables. In datasets which include only companies there are no variables like age or gender. Also we want to know precisely what drives both company customers and private customers to payment problems. Even loans for corporate customers are usually (in the previous literature) defined as non-retail loans we define them differently. Because instalment loans which we investigate are offered to both kind of customers in same manners and credit decisions are made with same tools and with same credit policy whether the applicant is a firm or a consumer, we define all our loans as retail loans.

4.1 Data

We received our data from a finance company that operates in Finland. Our data consists only of secured auto loans where the collateral is an automobile, a light truck, a motorcycle or a motor caravan and it has 8931 observations. Observations in the data are single contracts, not a pool of them like in the study of Heitfield and Sabarwal (2004). All those loans have ended during year 2011 so it is the most up-to-date data available for this kind of study. Because the maximum maturity for those auto loans is 5 years, and there is no minimum maturity, contracts in the data are made between years 2005 and 2011. From the beginning of this research it was clear for us to study both segments; loans for companies and loans for private customers. That was because the instalment loans as a product and also the credit decision process are very similar for both segments so there was no reason to focus only on one of them. We decided to separate those groups into two different datasets because those groups have a couple of unique variables (gender and age). In this point we must state that the group of companies includes other kind of organizations as well; three registered associations, four business associations and one foundation. Because the amount of other kind of organizations was so small, we decided to keep them as a part of group 'companies'. We also decided to keep its name as 'companies' because other kind of organizations represents only 0.7% of the observations in company loan dataset. We thought that the awareness of the fact that there are also a few observations which do not fulfill the definition of a company is enough.

In this stage we introduce our abbreviations for our four models which we are going to form in this study. From now on, we use these abbreviations when we discuss of these models.

PRIVATE.DEF = Explains the probability of default with data including only private customers.

PRIVATE.OVER = Explains the probability of serious overdue with data including only private customers.

COMPANY.DEF = Explains the probability of default with data including only company customers.

COMPANY.OVER = Explains the probability of serious overdue with data including only company customers.

4.1.1 Dependent variables

Our dependent variables are default and overdue. In our original dataset there were 7813 loans for private customers and 1118 loans for company customers. 105 (1.3%) of private customers' contracts defaulted and 477 (2.6%) recorded a serious overdue. In company loans 25 (2.2%) defaulted and 107 (9.6%) recorded an overdue value. Generally the credit contract ends only when it becomes repaid. In the situation of default the finance company repays the loans itself, in a sense, and writes the amount of repayment, which the collateral does not cover, as a credit loss. In this practical part of the study we define the default as a situation when the credit loss is made to the accounting. For sure there have been default situations during year 2011 also when the contracts have ended due to the insolvency of a customer but if the value of the collateral has covered the remaining loan it has not caused any straight credit loss for the company and thus it is not considered as a default situation in this research. Or at least the actual amount of credit loss has been insignificant consisting only some collection fees or sales loss. All in all, we can state that our default definition in this practical part is very strict and takes only in to the account the most serious defaults. We have to remember that for example Basel II defines a default as a situation when the obligor is past due more than 90 days (Bank for International Settlements 2004). That definition is more like our definition for the serious overdue which is another dependent variable in this study.

The serious overdue variable, or just overdue as we name it in the dataset means that payments have been more than 60 days late during six months period before the contract has ended. In this point we have to emphasize that contracts in this study which have recorded an overdue value have not defaulted. This framing is reasonable because we wanted to investigate also if there are remarkable differences between contracts which have defaulted and contracts with have just recorded overdue value. So we try to find out the reasons why some of the troubled contracts end up to default and some manage to survive against all odds (the research question 2).

4.1.2 Independent variables

We took 19 independent variables into our study. Those variables are chosen mainly because of our own professional experiences of credit risk analysis and our own interests. A few variables we also chose because we wanted to have some variables which we can compare with the earlier studies.

LTV was the first variable which we decided to have in this research. That was because we were interested, in the very beginning of this study, in the role of the down payment as a key variable in the whole loan industry (see the suggestion of Minister Antti Tanskanen in Introduction part). LTV is a continuous variable and it means Loan-To-Value. It compares the amount of the loan to the value of the collateral. The other side of the coin is the down payment which is usually the gap between the value of the collateral and the amount of the loan. So, if the LTV ratio is for example 90% the amount of down payment is 10%. The definition and nature of down payment is discussed in Chapter 2.8. LTV is very popular variable in the previous studies. For example Kocagil and Demir (2007) found it as the most important explanatory variable when they tried to explain defaults in prime auto loans. Heitfield and Sabarwal (2004) and Agarwal et al. (2008) found it also as a remarkable explanatory variable. In the whole dataset the average LTV was 69.16%. In private customers' contracts the average LTV was 68.80% and in the company loans 71.64%. The average LTV percent for each research group are described in Table 2 below.

Table 2: The average LTV percents

LTV	Private	Companies
Default	84.76 %	82.70 %
Non-Default	68.52 %	71.38 %
Overdue	76.77 %	78.21 %
Non-Overdue	68.04 %	70.64 %

From Table 2 we can see that LTV seems to have more significant role as an explainer of default than overdue. In every model averages are as we expected: The bigger LTV value predicts payment problems. Especially in private customers' contracts and default estimation. All in all, the difference in the LTV averages between defaulted and non-defaulted observations is very visible. Same holds with the overdue estimation.

RatingNum describes the rating value which the contract has got when the credit decision is made. In the original dataset the rating variable was classified and it got values from A to D but we decided to convert it to be a continuous variable. The A rating is obviously the best one so it gets the biggest value (4). B rating gets number 3, C gets 2 and D gets number 1. The reason for making the RatingNum variable continuous was that, as a continuous variable it gave better explaining power and was easier to handle. For example from continuous variables it is possible to calculate average values. The average RatingNum in whole dataset was 2.73. In private customer contracts the average RatingNum was 2.75 and in company loans 2.58. The average RatingNum values for each research groups are described in Table 3.

Table 3: The average RatingNum values (scale 1-4)

RatingNum	Private	Companies
Default	2.28	2.04
Non-Default	2.76	2.59
Overdue	2.42	2.16
Non-Overdue	2.78	2.64

As we can see from Table 3, the average ratings are higher in private loans in all rows. Averages are as we expected: a lower rating number predicts payment problems. The differences between default and non-default and overdue and non-overdue averages are bigger in company observations which suggest that RatingNum is better variable to explain problems in company loans.

Maturity tells the running time of each contract and it is a continuous variable also. The maximum maturity of contract in this study is 60 months because the company which provides our data does not offer loans for longer time. The shortest maturity in our data is one month. Our experience from the target company says that the loans are usually made to be precisely 12, 24, 36, 48 or 60 months even though the maturity could be anything between 1 and 60 months. The average maturity in the whole dataset was 45 months. In private customer loans the average maturity was the same value as for the whole data (45) and for company loans it was little bit smaller 44.05 months. The average maturities for each research groups are described in Table 4 below.

Table 4: The average maturities in months

Maturity	**Private**	**Companies**
Default	53.90	56.16
Non-Default	44.81	43.77
Overdue	47.95	46.48
Non-Overdue	44.61	43.48

From Table 4 we can see that longer contracts seem to get in payment problems easier than short ones. That is how we expected. In company loans' defaults the effect of maturity is strongest and in company loans' overdues the effect is weakest. All in all, the Maturity variable seems to have a bigger role when it explains defaults.

LastPayment is a continuous variable and it tells the amount of the last payment which completes the payment schedule. In auto loans it is usually possible to transfer some of the principal to the end of the maturity and get monthly payments at a reasonable level even if the car is quite expensive. The bigger last payment is usually possible to get only for quite new cars. When older cars come into the picture it is important to take care that the final payment should not be more than the value of the car in the end of the contract. This is a thing which finance company should worry about. If the maturity of the contract is the maximum five years and the car is relatively old already in the beginning of the contract it is very hard to estimate the value of the car when the final payment is about to be paid. If obligor is not able to pay the last instalment it is important that the value of the collateral meets the outstanding loan. Even at some level. Of course the bigger last payment transfers the cash flows to the future and that is why it adds the risk from the finance company's point of view. That is why the opportunity of the bigger last payment is not offered for all kind of financial deals.

In the data the last payment gets value zero if it does not differ from earlier payments so in that case there is no last bigger payment at all. If the last payment is bigger than other ones, we have the actual amount of the last payment. The LastPayment variable is a continuous one. In private customer loans 1111 (14.2%) had a bigger last payment and the average last payment within those 1111 contracts was 5315 euros. When compared those last payments to the price of the purchased cars the average percent was 21%. In company loans 139 (12.4%) got last payment and within those contracts the average last payment was 7109 euros. When compared to the car price the average last payment in company loans was 22%. This suggests that the bigger last payment occurs more often in private loans when compared to company loans. However, when it occurs in company loans it is likely to be relatively larger than in private loans. The average last payments for each research groups are described in Table 5 below. The average numbers in the table below include all those zero values mentioned earlier so those average numbers are not comparable with previously stated average numbers. They are only comparable with each other.

Table 5: The average last payments (€)

LastPayment	**Private**	**Companies**
Default	1480.83	2202.07
Non-Default	745.85	853.65
Overdue	992.71	1035.09
Non-Overdue	729.56	833.96

From Table 5 we can see that a bigger last payment predicts payment problems. This is also how we expected. The LastPayment variable seems to have the biggest role in explaining defaults in company loans. In that segment the difference between default and non-default averages is the largest.

TimeFinance is one of the most interesting variables in this study. We did not found any evidences of it from previous studies and that is why it was so exciting to see its possible explaining capacity. The reason why earlier studies have not got this variable is probably the lack of data. We build this variable by calculating the time between the date when the credit decision is made and the date when the contract is financed. In a normal case a car retailer contacts finance company after the customer has chosen a car. The car retailer gets the credit decision in a few minutes via telephone or internet. After the credit decision is made the retailer and the customer sign the contract and the customer pays the down payment to the retailer. Next the retailer gives the car for the customer and sends the contract to the finance company. When finance company gets the original contract paper it pays the principal to the car retailer and sends a payment schedule to the customer.

Sometimes it takes only a few days to get the actual contract paper to the finance company after the credit decision is made and sometimes it takes a few weeks. That is because the contract paper must be physically in the finance company before the finance company pays the principal to the car retailer. Because of our practical experience we had a clue that the time between those two steps might have some explaining power to the probability of payment problems. If everything is fine with the customer and the contract, the car retailer should send the complete contract paper as fast as possible to the finance company because it is an assumption that they want their money out of the deal as fast as possible. Especially because usually car retailer gives the car to the customer when the contracts is signed and the down payment is paid. In the whole dataset the average finance time was 7 days. In private customer contracts the average finance time was 6 days and in company loans 8.5 days. The average TimeFinance values for each research groups are described in Table 6 below.

Table 6: The average TimeFinance in days

TimeFinance	Private	Companies
Default	9.10	9.24
Non-Default	6.40	8.44
Overdue	7.92	9.80
Non-Overdue	6.30	8.30

From Table 6 we can see that the average TimeFinance values really are longer in contracts which have ended up having payment problems. The most significant difference between average times we can see in private customer defaults where the difference between defaulted and non-defaulted contracts is almost three days.

CarAge describes simply the age of the car in years in the moment when the contract is made. CarAge is also a continuous variable. In the credit decision process car age plays quite a big role because the car is the collateral in the loan contract and due to common knowledge the value of the car depends highly on its age. There are no limit values of how old the financed car could be but the range in whole dataset is 0 - 25 years. In the whole dataset the portion of new cars (age 0) is 27.23% and the average age of financed car is 4 years. In private customer loans the portion of new cars is 1900 which means 24.32% and the average car age was the same 4 years as in the whole data. In company loans the portion of new cars was 47.60% and the average age of the financed cars was 2.6 years. The average car ages for each research group are described in Table 7 below.

Table 7: The average car ages in years

CarAge	Private	Companies
Default	5.03	4.00
Non-Default	4.30	2.59
Overdue	5.02	3.75
Non-Overdue	4.25	2.47

From Table 7 we can see that an older car predicts payment problems. The CarAge variable seems to have the biggest role in explaining defaults in company loans. In defaulted company loans the cars are four years old on average and in non-defaulted only a little bit more than two and a half years.

DealDate is a continuous variable with describes the day of the month when the contract is made. Our goal is to find out if there are differences in the quality of the contracts which are made in the beginning of the month and in the end of the month. The average deal date in the whole dataset was 16.34. In private customer loans the average deal date was 16.30 and in the company loans 16.67. The average deal dates are described in Table 8.

Table 8: The average deal dates

DealDate	Private	Companies
Default	17.77	16.36
Non-Default	16.28	16.67
Overdue	16.29	17.01
Non-Overdue	16.27	16.34

From Table 8 we can see that the average DealDate is very similar in all segments. Even though we can see that the private customer contracts which defaulted are made, on average, later than those which did not default, those differences are so small that we do not give much weight for that information. According to the average values, there seems to be very light positive relationship between the date of month and payment problem probability. However, that relationship is almost invisible.

DealMonth is similar variable as the DealDate variable and also continuous. It describes the month when the contract is made. With that variable we tried to figure out if, for example the quality of the contracts decreases in the end of the year when there is usually low-season in the car retail market. The average deal month in the whole dataset was 6.28.

In the private customer loans the average deal month was also 6.28 and in the company loans 6.27. The average deal months are described in Table 9.

Table 9: The average deal months

DealMonth	Private	Companies
Default	6.26	6.48
Non-Default	6.28	6.26
Overdue	6.34	5.79
Non-Overdue	6.28	6.31

Table 9 shows us that the relationships between events and non-events go completely across and there seems to be no logic at all. Also those averages are so close to each other and also close to the overall average that we cannot find any explaining power from them.

Clock2 is a dummy variable which is converted of the variable called 'Clock' which tells the exact time when the credit decision is made in the finance company. Normally office hours are from 8 a.m. to 4 p.m. After normal office hours it is still possible to get credit decision until 6 p.m. Our dummy variable gets value 1 if credit decision is made during last two hours of the working day (4 p.m.-6 p.m.) and value 0 if credit decision is made during normal office hours. With this variable we tried to find out if the quality of the credit decisions drop during those last two working hours.

In the original dataset 1483 credit decisions were made during two last working hours which is 16.60% of the total amount of credit decisions. In the original dataset 1.46% of contracts defaulted which means 130 out of 8931 contracts. From the contracts which credit decision was made during last two working hours 1.21% defaulted. This suggests that there is no difference in the quality of credit decisions between office hours and late hours when all contracts are observed. In private customer loans which credit decision was made during two last working hours 13 out of 1432 (0.97%) defaulted. In company loans 5 out of 141 (4.26%) defaulted. The average values for this dummy variable are described in Table 10 below.

Table 10: The average values for dummy variable Clock2

Clock2	Private	Companies
Default	0,12	0,18
Non-Default	0,17	0,12
Overdue	0,15	0,18
Non-Overdue	0,17	0,12

From Table 10 we can see that in private loans credit decisions which are made after working hours predicts smaller probability of payment problems. However, the differences are very small in private contracts. In company loans the differences are larger. There seems to be visible difference between decisions made during office hours and late hours and those late hours' decisions seem to predict payment problems. From the average values we can state that the quality of credit decisions which concern company loans seems to drop a bit after normal office hours.

Product2 is also a dummy variable. In this study we do not describe what kind of products these product dummies include (product2, product3 and product7) because of secrecy request from our data provider. However our hypothesis is that contracts with product3 should cause less defaults than contracts with product2 and contracts with product2 as collateral should cause less defaults than contracts with product7. In the Product2 variable the observation gets value 1 if the collateral in that certain loan is a product number two and 0 if it is not. In the original dataset 364 out of 8931 observations got value 1 in this variable. In private customer loans 67 out of 7813 (0.9%) observations got product2 as the collateral and none of them defaulted. In company loans 297 out of 1118 contracts (26.6%) got product2 as collateral and 5 of them defaulted (1.7%). The average values for this dummy variable are described in Table 11.

Table 11: The average values for the dummy variable Product2

Product2	**Private**	**Companies**
Default	0.000	0.200
Non-Default	0.009	0.267
Overdue	0.017	0.234
Non-Overdue	0.008	0.271

In private loans the average values does not tell much because there simply was so few Product2 observations. Instead in the company loans the effect of Product2 was more visible. According to the average values it seems that if the collateral is Product2 it decreases the probability of payment problems when company customers are concerned.

Product3 is also a dummy variable. It describes whether the collateral of the contract is product 3 or not. If it is, the value for that observation is 1, otherwise it is 0. In original dataset 5822 observations got product3 as the collateral which is 65.2% of the whole data. In private customers that percentage is 71.80% and 0.97% of them defaulted. In company loans 18.69% got product3 as the collateral and 3.83% of them defaulted. The average values for the Product3 dummy are described in Table 12 below.

Table 12: The average values for the dummy variable Product3

Product3	**Private**	**Companies**
Default	0.724	0.320
Non-Default	0.718	0.184
Overdue	0.761	0.215
Non-Overdue	0.716	0.181

From Table 12 we can see that if the collateral is Product3, it increases the probability of payment problems in both, private and company customers' loans. The effect is strongest in company customer defaults. In private customers' contracts the effect of the Product3 variable is actually very small.

Product7 is a dummy variable as well. It is formed exactly the same way as previously discussed variables Product2 and Product3. In the whole data 61 out of the 8931 observations (0.68%) got product7 as the collateral. In private customer loans 46 out of 7813 (0.59%) got product7 as the collateral and one of them defaulted, 5 of them got overdue value. In the company loans 15 out of 1118 loans got product7 as the collateral and none of them defaulted. Two of them got overdue value. The average values for the Product7 dummy are described in Table 13 below.

Table 13: The average values for the dummy variable Product7

Product7	Private	Companies
Default	0.010	0.000
Non-Default	0.006	0.014
Overdue	0.010	0.019
Non-Overdue	0.006	0.013

Remembering that the appearance of Product7 as collateral is very rare we can state that according to the average values, there is a light positive relationship between the Product7 appearance and the probability of payment problems.

MonthCost is a continuous variable which describes the monthly costs which the obligor must pay in addition to the principal amortization and the interest. It is usually called a processing charge. The monthly costs vary in the data from 0 to 15 euros and the average value in the whole data was 7 euros which was also the average for private customer loans. In the company loans the average value was little bit higher 7.13 euros. The average month costs for each research group are described in Table 14 below.

Table 14: The average month costs in euros

MonthCost	Private	Companies
Default	7.29	7.40
Non-Default	7.11	7.13
Overdue	7.32	7.27
Non-Overdue	7.09	7.11

From Table 14 we can see that clearly a bigger monthly cost foreshadows payment problems. This is a little bit surprising. One simple explanation could be that car retailers request bigger monthly payments from customers which they see more risky. However that explanation does not make sense because car retailers do not carry the credit risk in the contracts included in this study so they have no incentive to compensate higher risk with extra costs.

StartCost is similar continuous variable than earlier mentioned MonthCost but this variable measures the cost which is added to the contract when the loan is granted and this cost comes to the each contract only once. The average start cost in the whole data was 124 euros which was also the average start cost in private customer loans. In company loans the average start cost was little higher 124.5 euros. The average start costs for each research group are described in Table 15.

Table 15: The average start costs in euros

StartCost	Private	Companies
Default	126.12	142.00
Non-Default	124.28	124.06
Overdue	122.67	129.50
Non-Overdue	124.40	123,5.0

The average values in Table 15 shows similar information than MonthCost averages, when company loans are concerned. Bigger costs foreshadow payment problems. However the average values in private customers' loans tells us different message because bigger costs seems to predict higher probability of default but lower probability of overdue.

City is a dummy variable which describes the region where the obligor has lived (or the company has operated) when the credit contract is made. It is formed by using the postal code which is required when the credit decision is made. The dummy gets value 1 when the obligor lives in one of the five biggest cities in Finland measured by the number of inhabitants. Those five biggest cities are Helsinki, Espoo, Tampere, Vantaa and Turku (Population Register Centre 2012). From the original data 1946 (21.8%) lived in those five biggest cities. From private customers 1660 (21.2%) lived in five biggest cities, 1.27% of them caused credit loss and 5.42% got overdue value. In company loans 286 (25.6%) companies operated in those five biggest cities, 1.4% of them caused credit loss and 6.6% got overdue value. The average values of the City variable are described in Table 16.

Table 16: The average values for the dummy variable City

City	Private	Companies
Default	0.20	0.16
Non-Default	0.21	0.26
Overdue	0.19	0.18
Non-Overdue	0.21	0.27

Table 16 tells that according to the average values, if the obligor lives (or operates) in some of the five biggest cities, it decreases the probability of payment problems. The effect seems to be stronger in company loans.

Gender is also a dummy variable which is obviously relevant only for private customers. Our gender dummy gets value 1 when obligor is a male and value 0 if the obligor is a female. In our original dataset 63% of the obligors were males and 37% females. 1.45% of men defaulted and 5.91% got overdue value. From women 1.12% defaulted and 6.44% got overdue value. According to percentages we can state that women have relatively more overdue situations but after all they settle their loans better than men. The average values for the Gender dummy are described in Table 17.

Table 17: The average values for the dummy variable Gender

Gender	Private
Default	0.70
Non-Default	0.63
Overdue	0.61
Non-Overdue	0.63

Table 17 confirms our findings because the average value of the Gender dummy is higher in default than in non-default cell. It means that according to the average values men have bigger probability to end up default than women. When overdue situations are concerned the thing is other way around.

Age is a continuous variable and only available for private customers. The average age in our data for private customers was 46 years. The average age of men was 47 and the average age of women was 45. The average age in the customers who defaulted was 42 years and within those customers who got overdue value it was 43 years. The average ages are described in Table 18.

Table 18: The average ages in each research group

Age	Private
Default	42.43
Non-Default	46.35
Overdue	42.72
Non-Overdue	46.60

Table 18 shows us that higher age decreases the probability of default and overdue according to average values.

CarPremium is a dummy variable which gets value 1 if the collateral is a car which we have classified as a premium car. According to our own classification the premium cars are Lexus, Mercedes-Benz, BMW, Audi, Jaguar and Volvo. There might be other premium cars in the market as well but in our data we found only those marks which fulfill our definition of a premium car. From the original data 614 contracts included a premium car as collateral. That is 6.87% of all observations. From those contracts 1.79% defaulted and 11.56% got overdue value. In private customer loans 7.04% of contracts had the collateral which was a premium car. 1.27% of them defaulted and 5.27% got overdue value. In company loans 5.90% of them got a premium car as the collateral. 6.1% of them defaulted and 13.64% got overdue value. The average values for the CarPremium dummy are described in Table 19.

Table 19: The average values for the dummy variable CarPremium

CarPremium	Private	Companies
Default	0.07	0.16
Non-Default	0.07	0.06
Overdue	0.13	0.08
Non-Overdue	0.07	0.05

From Table 19 we can see that premium cars seem to increase especially private customers' probability of overdue and company customers' probability of default.

Price is a continuous variable which describes the selling price of the car. In our data the range of the prices was from 800 euros to 136 500 euros. In the whole data the average price was 18 556 euros. In the private customer loans the average price was 17 461 euros and in the company loans 26 215 euros. The average prices for each research groups are described in Table 20.

Table 20: The average prices of cars in different research groups (€)

Price	Private	Companies
Default	16 893	26 680
Non-Default	17 468	26 204
Overdue	17 196	24 146
Non-Overdue	17 486	26 428

From Table 20 we can see that the probability of default or overdue seems to have negative relationship with the price of the car. It seems that customers with lower payment capacity are more likely to buy cheaper cars. On the other hand, it would make sense that more expensive cars would increase probability of default which is actually the case in company customer defaults. The effect is strongest in company customers' contracts which got overdue value. This predicts that loans for cheaper cars are more likely to overdue. That makes sense because it is possible that the car price correlates with the short term solvency of the company due the down payment requirement. The down payment requirement is usually a percentage value so the actual amount in cash depends on the price of the car.

4.2 Methodology

In this chapter we introduce our main methods used in this study and we also give some reasoning for using them. All our tests in this research are made with SAS Enterprise Guide 4.2 software.

Originally in probability of default related studies the PD was estimated by using the discriminant analysis (Z-scores) but later methods like logistic and probit regression became more popular. Decades ago it was usual to see published researches which used the basic least squares (OLS) linear regression to analyse a dichotomous dependent variable. At those days they did not know any better way or did not have access to good software for alternative methods. The biggest problem with linear regression is that values which it can construct are not bounded at all. In the logistic regression we transform the probability to a form of odds so the values are bounded by 0 and 1. Through that way we get rid of the upper bound. By taking logarithm of odds we also get rid of lower bound. By setting the results to the same form as linear functions of explanatory variables we get the logit model which is introduced in Formula 4. One big difference between the ordinary linear regression and the logistic regression is that there exist no random disturbance term in the formula of logit model.

However, that does not remove the fact that there is random variation in the logit model also. (Allison 2001, pp. 13-14)

The main idea of all those methods which explain the probability of default is to define function S of the form (Stephanou & Mendoza 2005, pp. 8):

$$S(X_i) = \beta_1 X_1 + \beta_2 X_2 + \ldots \beta_n X_n \qquad (3)$$

The vector X in Formula 3 represents the relevant risk factors. Those factors can be for example some financial information of company, age or gender of private customer or the region where the obligor lives. The idea of logistic regression is introduced in Formula 4 (Kocenda and Vojtek 2009, pp. 11), which describes the relationship between the probability *p* and vector *x*. The variable w_i represents the weight or importance of one characteristic used in analysis.

$$\log\left(\frac{p}{1-p}\right) = w_0 + \sum w_i \log x_i, \qquad (4)$$

The main research method in our study is the logistic regression which we chose due its popularity in the previous studies (Heitfield and Sabarwal 2004, Kocagil and Demir 2007, Kocenda and Vojtek 2009 and Altman et al. 2010). In short logistic regression is a regression model where the response variable follows a binomial distribution and it describes the relationship between the response variable and one or more independent variables. Basically the goal of the logistic regression is to model the probability of occurrence of a binary or dichotomous outcome. (Lix, Yogendran, Burchill, Metge, McKeen, Moore and Bond 2006, pp.132) In logistic regression there are no assumptions about the distributions of the explanatory variables. Actually logistic regression does not make many of the key assumptions of linear regression, for example it does not assume that the relationship between a response variable and explanatory variables is a linear one. In addition it does not assume that the response variable or the residual values are distributed normally. Independent variables should not correlate very strongly with each other because it can cause problems in estimation. In addition, the sample should be quite large and provide sufficient numbers in both categories of the response variable. The more independent variables, the larger the dataset should be. For example goodness-of-fit test Hosmer and Lemeshow recommend sample sizes greater than 400 for logistic regression. (Bewick, Cheek and Ball 2005, pp. 117)

As a method to estimate coefficients of each variable we used the maximum likelihood method which is one of the three popular methods. The other ones are ordinary least squares and weighted least squares. As an optimization technique we used the Fischer Scoring algorithm. According to Allison (2001 pp.16) maximum likelihood estimators have usually good properties in large sample. They are also consistent, asymptotically efficient and asymptotically normal. The main reason to choose maximum likelihood method is still that those properties get better when the sample is bigger. Our data includes 8931 observations so it is relatively large.

When coefficients are formed it is important to find out an intuitive meaning for them. The interpretation is a little bit different than in linear regression because the meaning must be determined in the terms of odds. In the linear regression interpretations are simple. If the coefficient is 0.50, it means that the probability of event increases 0.50 when explanatory variable increases with one unit. In logistic regression, for example if the variable is Gender (1=male and 0=female) and the odds ratio for variable is, for instance, 1.5. It means that the odds for event to happen are 1.5 times higher for men than women. In other words this means that the odds for event to happened for men is 50% higher than for women. If the variable is continuous we must subtract one from the odds ratio and then multiply it with 100. Through that way we know the percentage change in odds when independent variable increases with one unit. (Allison 2001, pp.28-29)

The selection of variables has been important issue in multiple regression analysis for long time. There exist many techniques for choosing significant explanatory variables. The most popular ones are stepwise selection, forward selection and backward elimination. (Fan and Cheng 2007, pp. 814) We end up using backward elimination. In contrast to forward selection, this method begins with the full model and eliminates one variable at a time. The backward selection starts with a full model and forward selection with model which has no variables included. The backward elimination procedure eliminates first variable which improves the model the most by being eliminated. Process continues by deleting one insignificant variable at a time so far, that there are no insignificant variables left in the model. The process is also terminated if all but one variable has been deleted. (Xu and Zhang 2001, pp. 478)

To find out which variables we should eliminate in our stepwise backward procedure we used the Wald test. With the Wald test we can test statistical significance of each coefficient in the model. Basically it compares the square of the regression coefficient with the square of the standard error of the coefficient. The Wald test is asymptotically distributed as a chi-square distribution. (Lix et al. 2006, pp.137) As a measure of how much each independent variable and each model can explain of the variation of dependent variable we used Nagelkerke test. The Nagelkerke test reflects the coefficient of determination which is basically the proportion of variance that is possible to explain with regression model. It is a measure of how well the model predicts the dependent variable from the independent variables. (Nagelkerke 1991)

The goodness-of-fit of our models we measured with the Akaike information criterion and with the Hosmer-Lemeshow test. The Akaike Information Criterion method (AIC) is a method for model selection which is one of the most popular methods for comparing multiple models. It takes both descriptive accuracy and parsimony into account. In the AIC model the main goal is to estimate the information loss when the probability distribution f, which is related with the true model, is approximated by probability distribution g, which is related to model that is to be evaluated. When AIC is used for selection, the chosen model is the one that minimizes the Kullback-Leibler distance between the model and the truth. (Wagenmakers and Farrell 2004, pp. 192) The traditionally AIC is defined as:

$$AIC = -2 \ln L + 2K \qquad (5)$$

In Formula 5 (Burnham and Anderson 2002, pp.62), L is for the maximized likelihood function for the estimated model and K is the number of free parameters in the statistical model. The parameter K is also called as a measure of complexity. (Burnham and Anderson 2002, pp.62)

Another goodness-of-fit test in our study was the Hosmer-Lemeshow test. The Hosmer-Lemeshow test creates ten equal sized groups of subjects. Those groups are created based on their estimated probability and then divided into two categories based on the actual observed outcome variable. After that the Hosmer-Lemeshow test calculates the probability value which is computed from the chi-square distribution to test the fit of the logistic model. The null hypothesis of the Hosmer-Lemeshow test is that the model is fit. If the p value is less than 0.05, the null hypothesis will be rejected. It means that the model does not fit. (Kuss 2002, pp. 3792-3793)

5 RESULTS

5.1 Single estimations and correlations

At first we wanted to know how well each of our independent variables can explain overdue situations and defaults by themselves and also how they correlate with each other. We also wanted to know what the direction of the relationship between our dependent and independent variables is at individual level. In total we made 72 regression analyses where we used only one explanatory variable at a time. At this stage we made also correlation matrixes and observed them to find the highest correlations between explanatory variables. The complete correlation matrixes are available in appendix 4.

The results of single estimations for private customer loan defaults are described in Table 21.

Table 21: The results of single estimations for private customer defaults. Explaining defaults by using one explanatory variable at a time.

PRIVATE.DEF – Single Estimations				
Variable	**Direction of relationship**	**Wald**	**Nagelkerke**	**AIC**
LTV	6.0763	<.0001	0.0791	1035.106
RatingNum	-1.4076	<.0001	0.0603	1054.736
Maturity	0.0553	<.0001	0.0431	1072.704
Age	-0.0248	0.0020	0.0095	1107.661
LastPayment	0.000097	0.0007	0.0082	1109.074
TimeFinance	0.0183	0.0019	0.0061	1111.288
CarAge	0.0463	0.0523	0.0035	1113.995
DealDate	0.0201	0.0799	0.0030	1114.496
Clock2	-0.3883	0.1924	0.0018	1115.727
Product2	-12.0302	0.9775	0.0018	1115.774
Gender	0.2774	0.1935	0.0017	1115.847
MonthCost	0.0575	0.3384	0.0010	1116.583
Price	-6.43E-6	0.5424	0.0004	1117.216
Product7	-0.4931	0.6274	0.0002	1117.392
City	-0.0772	0.7533	0.0001	1117.495
StartCost	0.000481	0.7634	0.0001	1117.504
CarPremium	-0.8845	0.8845	0.0000	1117.574
Product3	0.207	0.9021	0.0000	1117.580
DealMonth	-0.00214	0.9419	0.0000	1117.590

The third column of Table 21 (Wald test) tells us the significance of each variable when they were tested separately. According to 95% significance level (which is also significance level in the rest of our regressions), six out of our nineteen explanatory variables were statistically significant. In Table 21 the variables are listed in the order of their individual explanatory power measured with the Nagelkerke ratio (fourth column) which describes the coefficient of determination. The last column in Table 21 is for the AIC ratio which tells us the goodness-of-fit information for each regression.

When explaining defaults in single customer loans by using only one explanatory variable at a time we found that LTV (loan-to-value) seemed to have the strongest positive effect on default and also the best explaining power (0.0791). It means that when the amount of the loan rises, when compared to the value of the collateral, the risk of default also increases. The second best explanatory variable, according to the Nagelkerke ratio, seemed to be RatingNum which had a negative effect to the probability of default. Maturity, the third best individual explainer, had also a positive effect meaning that the increase of the loan's duration increases also the probability of default. LastPayment was also statistically significant but its effect was only slightly positive. Its Nagelkerke value was still the fourth best when compared with other variables. The last variable which was statistically significant was TimeFinance which seemed to have a positive impact to the probability of default. This means that when the time between signing the contract and financing it increases, it also increases the probability of default. CarAge is almost statistically significant with Wald value 0.0523. However its explaining power is only little bit more than a half of TimeFinance's corresponding value. All in all, six out of nineteen variables are statistically significant and three out of them seems to be dominant regressors when measured with explaining power.

According to the Pearson Correlation Coefficients the highest correlation in explanatory variables of private customer defaults was between: CarAge - Price -0.69303 and StartCost - CostMonth 0.58473. The relationship between CarAge and Price is very simple and easy to understand; The older car is usually also cheaper. The correlation between StartCost and MonthCost is probably because retailers decide which costs they will use in contract. Some retailers prefer high costs and determine both costs at high level.

Table 22: The results of the single estimations for private customer overdues. Explaining overdues by using one explanatory variable at a time.

PRIVATE.OVER – Single Estimations				
Variable	**Direction of relationship**	**Wald**	**Nagelkerke**	**AIC**
RatingNum	-0.8979	<.0001	0.0469	3446.815
LTV	2.3636	<.0001	0.0299	3496.372
Age	-0.0244	<.0001	0.0145	3540.630
CarPremium	0.7427	<.0001	0.0080	3559.482
Maturity	0.0155	<.0001	0.0078	3559.859
CarAge	0.0493	<.0001	0.0061	3564.886
TimeFinance	0.0150	<.0001	0.0043	3569.972
MonthCost	0.0768	0.0101	0.0026	3574.862
LastPayment	0.000047	0.0105	0.0020	3576.487
Product3	0.2357	0.0329	0.0016	3577.637
Product2	0.7295	0.0547	0.0011	3579.272
City	-0.1589	0.1872	0.0006	3580.565
Product7	0.6443	0.1765	0.0005	3580.804
Clock2	-0.1513	0.2478	0.0005	3580.978
Gender	-0.0875	0.3680	0.0003	3582.353
Price	-3.19E-6	0.5229	0.0001	3581.550
StartCost	-0.00044	0.5600	0.0001	3581.944
DealMonth	0.00594	0.6734	0.0001	3582.018
DealDate	0.000257	0.9625	0.0000	3582.178

Table 22 describes the results of single estimations for private customer overdues. When compared with Table 21 which describes results for private customer defaults we can state that RatingNum took the first place in explaining power ranking. Its effect to the probability of overdue was negative but its negative effect to probability of overdue was 36% smaller than its effect to probability of default. LTV was the second best regressor but its explaining power was weaker than in single estimation for private customer defaults. The Age variable had almost same kind of effect to the probability of overdue than to the probability of default as well as TimeFinance. CarPremium is interesting variable because it did not seem to have any explaining power in private customer default estimation but in overdue estimation it was statistically significant variable having the fourth best Nagelkerke ratio. It is also interesting that in the overdue estimation it got effect on different direction than in default estimation. Maturity was also statistically significant but its explaining power (0.0078) was weaker than in default estimation. The direction of CarAge was also positive in this estimation but its explaining power was also weaker than in default estimation. Product3 and MonthCost are variables which were not statistically significant in default estimation. In overdue estimation they seemed to have some role. A noteworthy detail was the Wald ratio for the Product3 variable in the default estimation and in the overdue estimation. In the default estimation it was the second worst (0.9419) but in the overdue estimation it was 0.0329 and thus statistically significant. LastPayment had even lighter positive effect on private customer overdues than defaults. Anyhow it was statistically significant variable also in this estimation. RatingNum, LTV and Age seemed to have the biggest individual effect on private customer overdues.

The Pearson Correlation Coefficients for the explanatory variables of private customer overdues were very similar than correlation coefficients for the explanatory variables of private customer defaults. That was obvious result because the datasets for both tests are very similar. The only difference between those two datasets is that in the private customer overdues dataset the observations which had default value are excluded. Therefore there are 105 observations less than in the private customer defaults dataset. The highest correlations were between: CarAge – Price -0.69442, StartCost – CostMonth 0.58898 and Product3 – Price -0.45647. The relationship between Product3 and Price is as easy to understand as the relationship between the age of the car and the Product3 variable: As we mentioned before, the definition of Product3 variable refers to a used cars which also refers to a cheaper price than in average car in the data.

Table 23: The results of single estimations for company customer defaults. Explaining defaults by using one explanatory variable at a time.

COMPANY.DEF – Single Estimations				
Variable	**Direction of relationship**	**Wald**	**Nagelkerke**	**AIC**
Maturity	0.1004	0.0002	0.1096	219.576
RatingNum	-0.8576	0.0019	0.0507	232.485
LTV	4.4556	0.0047	0.0494	232.767
LastPayment	0.000097	0.0232	0.0174	239.699
CarPremium	1.1702	0.0371	0.0158	240.057
CarAge	0.0924	0.0493	0.0157	240.059
Product3	0.7365	0.0910	0.0120	240.867
Clock2	0.8072	0.0907	0.0115	240.975
StartCost	0.00569	0.1512	0.0109	241.114
City	-0.6019	0.2737	0.0063	242.098
Product7	-12.0497	0.9863	0.0032	242.775
MonthCost	0.1105	0.4534	0.0031	242.786
Product2	-0.3772	0.4548	0.0028	242.859
DealMonth	0.0181	0.7555	0.0005	243.361
TimeFinance	0.00326	0.7830	0.0003	243.391
DealDate	-0.00396	0.8615	0.0001	243.427
Price	2.586E-6	0.0310	0.0001	243.427

Table 23 describes the results of the single estimations for company customer defaults. For company loans we got only seventeen explanatory variables because we did not have the age information for companies and obviously we did not have the gender information either. However, seven out of seventeen explanatory variables were statistically significant when explaining default situations only with one variable at a time.

Surprisingly, Maturity had the best explaining power (0.1096) which is also the best explaining power in this whole study for one variable. The direction of the Maturity was positive which is in line with the single estimations for private customers. The second best variable, with less than a half of Maturity's explaining power, was RatingNum which direction was also in line with estimations with private customers. The third best regressor was LTV and the fourth was LastPayment which had precisely the same direction than in private customer defaults estimation. CarPremium was the fifth best regressor with surprisingly high positive effect on probability of default. In this test it means that when the collateral of the loan is a premium car it seems to increase the probability of default. The CarAge variable had almost the same explaining power than CarPremium being the sixth best regressor and having more than a half bigger positive effect on the probability of default than in the private customers' loans. The Price variable is interesting regressor because it seemed not to have any explaining power, being the worst explanatory variable, but still statistically significant. However, its effect or direction was so marginal it had hardly any role in this matter. Maturity, RatingNum and LTV seemed to have the biggest individual effect on the company customers' defaults.

According to Pearson Correlation Coefficients the highest correlation in explanatory variables of company customer defaults was between: CarAge – Price -0.60428, StartCost – CostMonth 0.56560 and RatingNum – CarAge -0.46201. The only difference between previously presented correlation coefficients and correlation coefficients of company customer defaults dataset is the third highest correlation coefficient between RatingNum and CarAge. The negative relationship between RatingNum and CarAge is most likely because companies with weaker background information tend to by older cars which are usually also cheaper.

Table 24: The results of single estimations for company customer overdues. Explaining overdues by using one explanatory variable at a time.

COMPANY.OVER –Single Estimations				
Variable	**Direction of relationship**	**Wald**	**Nagelkerke**	**AIC**
RatingNum	-0.7340	<.0001	0.0620	671.909
LTV	2,4396	0.0001	0.0320	687.770
CarAge	0.0888	0.0004	0.0222	692.942
City	-0.5216	0.0473	0.0083	700.164
Maturity	0.0150	0.0440	0.0082	700.228
Price	-0.00001	0.0890	0.0060	701.371
Clock2	0.4820	0.0759	0.0056	701.561
DealMonth	-0.0439	0.1371	0.0043	702.242
CarPremium	0.5005	0.1835	0.0031	702.872
StartCost	0.0017	0.3404	0.0018	703.535
TimeFinance	0.0057	0.3110	0.0017	703.581
MonthCost	0.0544	0.3917	0.0015	703.673
Product3	0.2175	0.3833	0.0014	703.735
Product2	-0.1972	0.4102	0.0013	703.772
LastPayment	0.000025	0.4585	0.0010	703.958
Product7	0.3556	0.6426	0.0004	704.273
DealDate	0.00470	0.6816	0.0003	704.302

Table 24 describes the results of single estimations for company customer overdues. The Maturity variable which ruled the default estimations for company customer defaults was also statistically significant in overdue estimation but its explaining power was less than a half when compared with default estimation. RatingNum was the best single regressor having a little bit stronger explaining power and little smaller negative effect than in default estimation. The third best regressor was CarAge having more explaining power (0.0222) in explaining the probability of overdue than probability of default (0.0157). The City variable is an interesting regressor because it was statistically significant only in one test. That test was the single estimation explaining company customer overdues. The City variable had a negative effect on company loan overdues which was in line with three previous results. In this case it means that if company operates in some of the five biggest cities in Finland it decreases the probability of default. In company customer overdues three variables seemed to have the biggest individual effect and they were RatingNum, LTV and CarAge.

The Pearson Correlation Coefficients for the explanatory variables of company customer overdues were very similar than correlation coefficients for explanatory variables of company customer defaults. The highest correlations were between: CarAge – Price -0.60550, StartCost – CostMonth 0.56567 and RatingNum - CarAge -0.45472.

5.2 Backward selection

After single estimations our next target was to find out how independent variables explain payment problems together. We used logistic regression and backward elimination procedure to form our four final models. In this section we introduce our backward elimination steps for each model and also the final results after eliminations. In the end of this section is also a table where we sum up our goodness-of-fit ratios.

Table 25: The results of logistic regressions with backward selection procedure for private customer defaults

PRIVATE.DEF – Backward Elimination Procedure					
Step	**Eliminated variable**	**Variables included**	**Nagelkerke**	**AIC**	**Wald**
Start	-	19	0.1599	985.841	<.0001
1	Product2	18	0.1582	985.659	<.0001
2	CostMonth	17	0.1582	983.659	<.0001
3	DealMonth	16	0.1582	981.666	<.0001
4	Price	15	0.1581	979.754	<.0001
5	Product7	14	0.1579	977.925	<.0001
6	City	13	0.1575	976.417	<.0001
7	Age	12	0.1570	974.914	<.0001
8	CarPremium	11	0.1557	974.256	<.0001
9	StartCost	10	0.1546	973.467	<.0001
10	Clock2	9	0.1535	972.587	<.0001
11	DealDate	8	0.1509	973.377	<.0001
12	Gender	7	0.1480	974.475	<.0001
13	LastPayment	6	0.1449	975.728	<.0001

Table 25 presents steps in our backward selection procedure when constructing the best model to explain private customer defaults. As we can see the model is statistically significant in every step. In the beginning we had all nineteen explanatory variables included in the regression. Five out of nineteen variables survived to the final model which is presented in Table 26. When compared with the single estimations we can notice that LTV seemed to have the most meaningful effect on the probability of default. Its positive effect was little bit smaller than in single estimations (drop from 6.07 to 5.81). RatingNum seemed to have the second best effect on probability of default (-0.86) and it was not a surprise since it succeed also in single estimations. The third best regressor was Product3 which role was very interesting in this model. In single estimations it got hardly any explaining power (second worst variable) and it was definitely not statistically significant. However it managed to survive in the final PRIVATE.DEF model somehow. Another interesting thing was that the direction of the Product3 variable changed in the final model where it got negative effect on probability of default (-0.32). In single estimations it was positive (0.21). Another variable which was not statistically significant in individual level but still survived into the final model was CarAge. Its success in the backward selection was not a surprise because it was almost statistically significant in the individual level (Wald 0.0523). It is interesting that the positive effect of CarAge grew in the final model. It was 0.0463 in the single estimations and in the final model it was 0.1291.

There were two variables which were statistically significant in the single estimations and still, did not survive to the final model. They were Age and LastPayment. LastPayment was excluded last and its drop was predictable because of its very slight effect on probability of default in single estimations. Age instead did not perform as well as we thought. It was excluded in the step seven even though it was the fourth best variable in the single estimations.

Table 26: The final model explaining private customer defaults

PRIVATE.DEF – Final Model				
Analysis of Maximum Likelihood Estimates				
Parameter	**Estimate**	**Standard Error**	**Pr > ChiSq**	**Odds ratio estimate**
Intercept	-9.9305	1,3239	<.0001	-
TimeFinance	0.0180	0.00714	0.0119	1.018
RatingNum	-0,8575	0.2094	<.0001	0.424
LTV	5,8082	1.0422	<.0001	333.032
Maturity	0.0507	0.0111	<.0001	1.052
CarAge	0.1291	0.0337	0.0001	1.138
Product3	-0.3248	0.1201	0.0069	1.915

Table 27 presents ten steps in our procedure to construct the model to explain private customers' overdues. The final model is presented in Table 28. In this backward selection process, also as in previously presented, all models (or steps) were statistically significant. Nine variables out of nineteen survived into the final model. This final model had the biggest amount of significant explanatory variables in this study. As a contrast we can remind that in the logistic model of Altman et al. (2010) there were 13 significant variables included. When compared to single estimations the biggest surprises were the role of StartCost and Price. StartCost was not statistically significant in the single estimations and it was the seventeenth best regressor at individual level having hardly any explaining power. However it survived into the final model and also its negative effect on the probability of overdue increased. Price was the sixteenth best regressor at individual level and its effect was negative. In the final model the effect of Price on the probability of overdue was slightly positive (0.000029) and statistically significant.

Table 27: The results of logistic regression with backward selection procedure for private customer overdues.

PRIVATE.OVER – Backward Elimination Procedure					
Step	**Eliminated variable**	**Variables included**	**Nagelkerke**	**AIC**	**Wald**
Start	-	19	0.0813	3381.783	<.0001
1	DealDate	18	0.0813	3379.787	<.0001
2	DealMonth	17	0.0813	3377.839	<.0001
3	LastPayment	16	0.0813	3376.019	<.0001
4	Product7	15	0.0812	3374.251	<.0001
5	Product3	14	0.0810	3372.775	<.0001
6	Gender	13	0.0808	3371.295	<.0001
7	City	12	0.0805	3370.137	<.0001
8	Clock2	11	0.0802	3369.110	<.0001
9	Maturity	10	0.0798	3368.210	<.0001
10	Product2	9	0.0789	3369.135	<.0001

The best explanatory variables in the final model for the probability of private customer overdue estimation were LTV, RatingNum and CarPremium. LTV and RatingNum were also in the model for private customer default estimation but CarPremium was not. The effect of CarPremium variable was positive (0.46) so it seemed that if a private customer has a premium car as collateral of the loan contract it is more likely to cause serious overdue than other contracts. This is interesting because the CarPremium variable was excluded from the private customer default model at stage eight. So it seems that CarPremium can explain overdue but not default. Three variables which were statistically significant dropped out from the final model. They were Product3, LastPayment and Maturity. Maturity was maybe the biggest surprise because it was the fifth best regressor in the individual level.

Table 28: The final model explaining private customer overdues.

PRIVATE.OVER – Final Model				
Analysis of Maximum Likelihood Estimates				
Parameter	**Estimate**	**Standard Error**	**Pr > ChiSq**	**Odds ratio estimate**
Intercept	-2.3091	0.5331	<.0001	-
RatingNum	-0.6961	0.0924	<.0001	0.499
LTV	1,6829	0.3051	<.0001	5.381
CostMonth	0.0732	0.0372	0.0489	1.076
StartCost	-0.00190	0.000953	0.0462	0.998
Age	-0.0114	0.00391	0.0035	0.989
CarAge	0.0460	0.0180	0.0105	1.047
TimeFinance	0.0122	0.00424	0.0040	1.012
Price	0.000029	6.378E-6	<.0001	1.000
CarPremium	0.4607	0.1559	0.0031	0.631

In Table 29 we present steps for constructing a model for explaining company customer defaults. In Table 30 we can see that only two variables (out of seventeen) managed to survive to the final model. That is quite a low amount when compared to the private customers' models which got six and nine explanatory variables in the final models.

Table 29: The results of logistic regression with backward selection procedure for company customer defaults.

COMPANY.DEF – Backward Elimination Procedure					
Step	**Eliminated variable**	**Variables included**	**Nagelkerke**	**AIC**	**Wald**
Start	-	17	0.2170	227.682	0.0313
1	Product7	16	0.2137	226.423	0.0221
2	RatingNum	15	0.2137	224.425	0.0148
3	DealDate	14	0.2136	222.440	0.0097
4	FinanceTime	13	0.2133	220.504	0.0063
5	Price	12	0.2124	218.715	0.0039
6	DealMonth	11	0.2113	216.960	0.0024
7	MonthCost	10	0.2097	215.329	0.0013
8	CarPremium	9	0.2073	213.852	0.0010
9	Product2	8	0.2050	212.364	0.0005
10	City	7	0.1993	211.646	0.0004
11	Product3	6	0.1954	210.521	0.0003
12	StartCost	5	0.1859	210.653	0.0002
13	LTV	4	0.1761	210.841	<.0001
14	Clock2	3	0.1660	211.084	<.0001
15	LastPayment	2	0.1556	211.405	<.0001

However in this stage it is important to notice the role of the Maturity variable which was the best explanatory variable in the single estimations and it also survived to the final model. It is noticeable that Maturity had very weak role when explaining other effects but in company loans and defaults it had statistically significant positive effect (0.1223). So it seems that especially in the company loans the duration of the contract is very important variable. Another variable in the final model was CarAge which was the sixth best regressor in the single estimations but still, survived into the final model. Its positive effect increased from 0.0924 to 0.1959 when the single estimation and the final model were compared. In previously presented backward selection procedures LTV and RatingNum variables were the best ones. However both of them dropped from the final model even they were second and third best regressors at individual level. Furthermore, the RatingNum variable was excluded in very early stage (step2).

Table 30: The final model explaining company customer defaults.

COMPANY.DEF – Final Model				
Analysis of Maximum Likelihood Estimates				
Parameter	**Estimate**	**Standard Error**	**Pr > ChiSq**	**Odds ratio estimate**
Intercept	-10.7295	1,79380	<.0001	-
Maturity	0.1223	0.0302	<.0001	1.130
CarAge	0,1959	0.0585	0.0008	1.216

In Table 31 we present steps for constructing a model to explain company customer overdues. As we can see from Table 32, there are only two variables left in the model. The same amount as in the final model for company customer defaults.

Table 31: The results of logistic regression with backward selection procedure for company customer overdues.

COMPANY.OVER – Backward Elimination Procedure					
Step	**Eliminated variable**	**Variables included**	**Nagelkerke**	**AIC**	**Wald**
Start	-	17	0.0972	685.023	0.0003
1	Product7	16	0.0972	683.023	0.0002
2	Maturity	15	0.0971	681.046	<.0001
3	Product3	14	0.0969	679.180	<.0001
4	Price	13	0.0967	677.304	<.0001
5	MonthCost	12	0.0960	675.638	<.0001
6	CarPremium	11	0.0954	673.990	<.0001
7	DealDate	10	0.0947	672.381	<.0001
8	LastPayment	9	0.0933	671.094	<.0001
9	StartCost	8	0.0915	670.092	<.0001
10	Product2	7	0.0896	669.106	<.0001
11	CarAge	6	0.0883	667.830	<.0001
12	FinanceTime	5	0.0857	667.243	<.0001
13	City	4	0.0811	667.685	<.0001
14	Clock2	3	0.0769	667.969	<.0001
15	DealMonth	2	0.0713	668.939	<.0001

In Table 32 we can see that RatingNum and LTV are the best explanatory variables to explain company customer overdues. The role of those variables is not a surprise because of their success in the individual level; RatingNum was the best single regressor and LTV the second best. The effects of those variables were weaker in the final model but the directions were the same; negative for RatingNum and positive for LTV, as expected. The CarAge variable which survived to the final model in previously presented model for company customer default estimation was excluded in the step eleven. City which got surprisingly high explaining power in single estimations was excluded in the step thirteen which was later than we expected. The effect of the Maturity was surprisingly low. It was the fifth best regressor at individual level and got a huge role in the COMPANY.DEF model but in this process it dropped in the second step.

Table 32: The final model explaining company customer overdues.

COMPANY.OVER – Final Model				
Analysis of Maximum Likelihood Estimates				
Parameter	**Estimate**	**Standard Error**	**Pr > ChiSq**	**Odds ratio estimate**
Intercept	-1.8809	0.7166	0.0087	-
RatingNum	-0.6142	0.1426	<.0001	0.541
LTV	1,5105	0.6973	0.0303	4.529

Table 33 sums up the results of different tests which tell of the quality and performance of the models which are previously presented. The AIC value tells us the Kullback-Leibler distance between the model and the truth (Wagenmakers and Farrell 2004, pp. 192). Therefore the AIC value should be as small as possible. The difference percent in column three tells us how much the AIC value decreases when covariates are added to the model. As we can see, covariates increase the quality of the model much more when default is the one to explain. According to the Hosmer and Lemeshow test three models out of four are fitted with the data. In the Hosmer and Lemeshow test the null hypothesis is that the model fits the data. Therefore values more than 0.05 tells us that the model fits the data. As we can see from Table 33, the model 4 which explains the company customer overdues does not fit the data. That model was also the worst one measured with the AIC difference percentage. The best fit between model and data according to the Hosmer and Lemeshow test was with Model 3 which explains company customer defaults. The last column in Table 33 tells us the explanation power of each model. The model 3 seemed to be the best one also when measured with explanation power. Nagelkerke ratios also confirm previously stated note that our data has much more potential to explain defaults than overdues in both customer segments.

Table 33: The statistics of the quality of four default and overdue models.

	AIC: Intercept only	AIC: Intercept and Covariates	AIC: Difference %	Hosmer and Lemeshow: Pr > ChiSq	Nagelkerke
Model 1: PRIVATE.DEF	1115,568	975,728	12,54 %	0.0720	0.1449
Model 2: PRIVATE.OVER	3580,228	3369,135	5,90 %	0.1074	0.0789
Model 3: COMPANY.DEF	241,458	211,405	12,45 %	0.6458	0.1556
Model 4: COMPANY.OVER	702,47	668,939	4,77 %	0.0167	0.0713

5.3 Variable ranking and interpretation

In this chapter we rank our explanatory variables in order of their significance in this study. This ranking is our subjective viewpoint of the relevance of each variable in this study compared to previous studies and our expectations.

1. LTV (loan-to-value): As the most important explanatory variable we ranked LTV. We have discussed its role during this study several times. We added to our discussion for example the suggestion of the working group of Minister Antti Tanskanen (Ministry of Finance 2012) and the results of the research of Heitfield and Sabarwal (2004). Those viewpoints both approached the role of the LTV by using mortgage house loans as a baseline. In addition to the research of Heitfield and Sabarwal, LTV was also involved to the studies of Kocagil and Demir (2007) and Agarwall et al. (2008) which all were in line with our results concerning the role of LTV. According to our single estimations LTV was significant explainer in all models where it explained payment problems alone. LTV explained best the payment problems of private customers but it was a remarkable explainer in company level as well. In backward estimation LTV dropped only from the COMPANY.DEF model. Despite the COMPANY.DEF model, LTV was included in all of our three final models having strong positive relationship with payment problems in all of them.

These results are in line with previous studies introduced previously. The biggest correlation LTV had with maturity. That result is obvious. If the down payment is low, the LTV value is high and usually in those cases the payback time is also long. To sum up, we can say that according to previous studies LTV seems to be an important explainer in payment problems in all kind of loans where the collateral is included. However, that relationship seems to be unique in car loans. That is because the value of the collateral behaves, let us say, in deviant way. The market price of the car depend on so many attributes that it is very hard to estimate. The only thing which is sure is that the value will decrease when the time goes by. That is the most important difference with mortgage house loans where the value tends to increase more likely than decrease. The speed of the value drop-off is the unknown parameter and the down payment is the key for securing the position of the finance company. That is the reason for LTV being such an important variable in this study.

2.RatingNum: As the second best explanatory variable we ranked RatingNum which role was as big as we expected. We have to remember that it is formed to explain the probability of possible payment problems, especially default, and in that regard its explaining power is not a surprise. As we mentioned in Chapter 2.4 (Credit ratings) the rating should be a straightforward opinion of the creditworthiness of the customer described only with a few symbols. We must state that ratings made enormous job explaining defaults when comparing to other explainers. In single estimations RatingNum was statistically significant in all of our regressions where it explained payment problems alone. RatingNum was included to the three of our four final models and was dropped out only from the COMPANY.DEF model. In all of our regressions RatingNum had a negative relationship with the payment problems. The strongest negative relationship it recorded in the PRIVATE.DEF model.

In previous studies Koceda and Vojtec (2009) found credit rating as an insignificant variable when explaining default. Their definition of default reminds our definition of serious overdue and their data included only private customers so we can compare their results with our PRIVATE.OVER model. It is interesting that they found the credit rating variable insignificant while we found quite strong negative relationship with payment problems and credit rating in our PRIVATE.OVER model which was also statistically significant. The reason for this difference might be the way how the ratings are formed. In the study of Koceda and Vojtec the credit ratio was formed completely by comparing the wage of the customer to the possible expenditures and minimum wage levels. In our data the credit rating variable includes lot more information. The credit rating for private customers includes credit information from credit bureaus whether the customer has a default register mark or not. In company loans credit ratings include a rating value from a credit bureau which is something between AAA and C. The company rating includes information from financial statements, industry, persons in charge and for example facts of company’s payment behavior. In addition to information from credit bureau our RatingNum variable also includes information from the target company’s own customer register such as previous payment behavior and previous credit decisions. A credit score was also an explanatory variable in the study of Kocagil and Demir (2007) which both explained defaults especially in auto loans. The Credit score variable of Kocagil and Demir reminds more our corresponding variable because it represents the credit bureau’s opinion of the creditworthiness. Kocagil and Demir and Agarwal et al. both found credit score as a significant regressor with negative relationship with default. Those results are in line with ours.

3. Maturity: The durations of the contract is the third best explanatory variable in our study. In tests where Maturity explained payment problems alone it was a statistically significant explainer in all of our four regressions. In backward estimations it survived two out of the four final models. In the COMPANY.DEF model Maturity outperformed two tough variables LTV and RatingNum being only variables in final model with

CarAge. Maturity seemed to have a unique relationship with CarAge as an explainer of default in company loans. Both Maturity and CarAge recorded higher individual effect on default when they were tested together (when compared to individual estimations). We might say that the duration of the contract seems to strengthen the effect of CarAge causing default, and vice versa. In company loans which had no payment problems the average duration of the contract was 43.8 months and the average car age was 2.6 years. In the company loans which defaulted the corresponding values were 56.2 months and 4 years. In this light of evidence we suggest to spent more time and concentration for credit decisions when the obligor is a company asking a loan with 5 years durations and for four years old car. As an explainer of private customers' payment problems Maturity was successful in the PRIVATE.DEF model being one of six variables in the final model. From Figure 4 we can see the distribution of Maturity in company loans with default and non-default events. As we can see maturities are usually some amount of years even so it could be any amount of months. There is a clear accumulation of defaulted contracts in duration of 60 months.

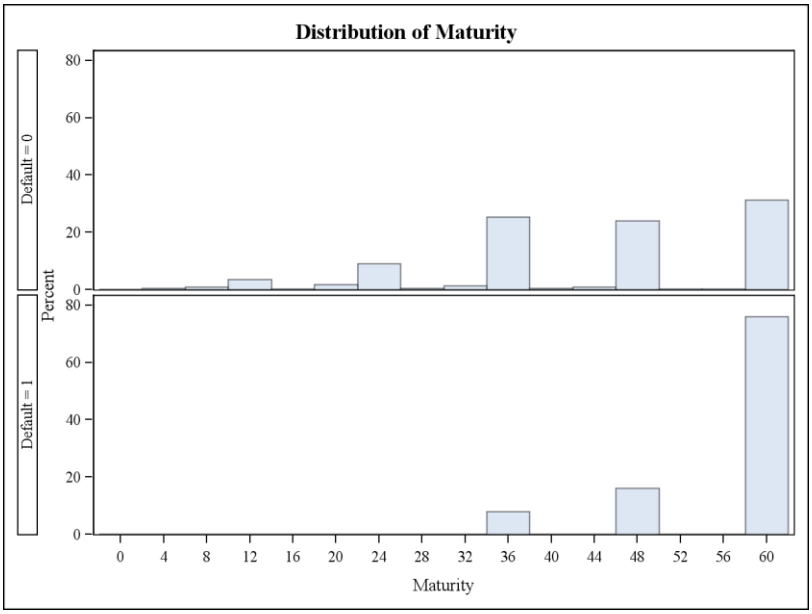


Figure 4: The distribution of Maturity in company loans with default and non-default values.

As a contrast we introduce the distribution of Maturity in company loans with overdue and non-overdue values in Figure 5. As we can see from Figure 5 the distribution of maturities is much more even in contracts with overdues than with defaults. Actually the distribution of overdue and non-overdue contracts seems not to have very clear difference.

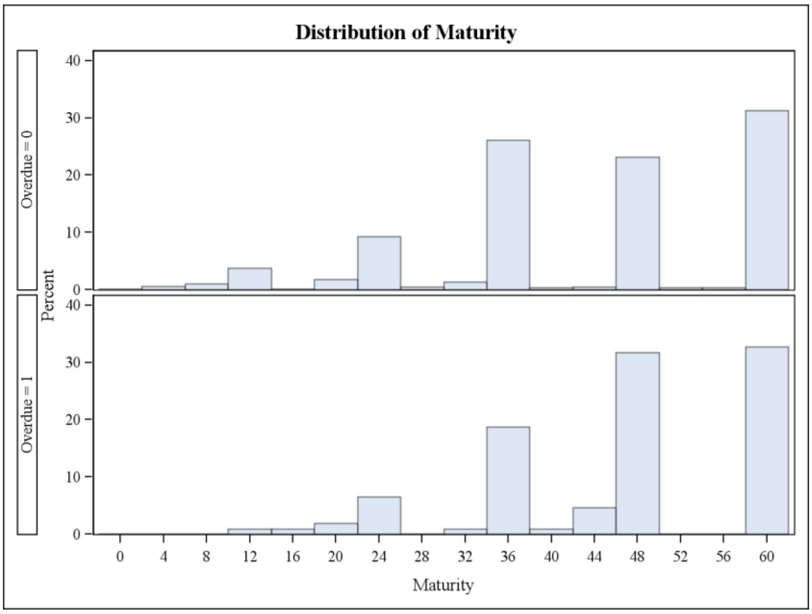


Figure 5: The distribution of Maturity in Company loans with overdue and non-overdue values.

One interesting finding is that Maturity was not included in the final models which explained overdues (PRIVATE.OVER and COMPANY.OVER). The Maturity variable also had significantly lower positive effect on overdues than on defaults, in single estimations. This is interesting result remembering our third research question of whether there are differences between defaulted contracts and contracts with serious overdue. In previous studies Kocagil and Demir (2007) found positive relationship with defaults and maturity with auto loan data. That result is in line with our results.

4. CarAge: As we stated in the interpretation of the Maturity variable the age of the car seemed to have a unique relationship with the maturity of the loan contract as an explainer of defaults in company loans. In addition the CarAge variable succeeds also as an explainer of private customers' payment problems. It was a significant explainer in all but one regression when it explained payment problems alone. In the backward elimination process it managed to survive three out of the four final models. In all of our regressions the CarAge variable had a positive effect on payment problems. The effect was strongest in the COMPANY.DEF model as we introduced earlier. Interestingly, among final models where CarAge was included, its role was lowest in the PRIVATE.OVER model where the Maturity variable was not included.

Agarwal et al. (2008) and Kocagil and Demir (2007) included the age of the car into their studies as a dummy variable new/old car. According to the results of Agarwal et. al., a new car as collateral decreases the probability of default 15 percent. As for Kocagil and Demir found that used car as collateral had a positive relationship with default situations. Those results are consistent with each other and also with our results.

5. TimeFinance: One of the most interesting variables in this study was TimeFinance. We are glad to be able to rank it as high as we do. It tells the time between signing the contract and financing it. There were no earlier evidences of this explanatory variable in the previous studies which we introduced. The reason is simple and familiar in this research frame: The lack of data and information. However, the TimeFinance variable performed well in our regressions. When it explained defaults alone it was statistically significant in both models which explained private customers' payment problems (PRIVATE.DEF and PRIVATE.OVER). In private customers' analysis TimeFinance also survived to the final models having positive effect on payment problems. In the payment problems of companies TimeFinance did not has much explaining power neither in single estimations nor in backward elimination. The average TimeFinance value for private customers' loans with no default value was 6.4 days. The corresponding value for loans which defaulted was 9.1. In company loans the range is smaller (8.4 and 9.2 days). According to these results we can state that especially in private customers' loans the time between signing and financing the contract is an important variable. As we discussed earlier, it should be advantage for all parties (finance company, car retailer and customer) to finance the contract as fast as possible. In other words there should be no reason for delay, unless there is something wrong with the contract or the customer.

6. Age: The age information we had only for our private customers. As an individual explainer age was a significant variable in both regressions (PRIVATE.DEF and PRIVATE.OVER). After the backward elimination process the Age variable was only included to the model explaining overdues. In all of four regressions age had negative relationship to payment problems. Because age survived to the PRIVATE.OVER model we state that the age seems to increase more the probability of overdue than default. However from the average numbers we can see that they are quite similar. From Figure 6 we can see that the distribution of age concentrates more to the left side in observations with default value. The distribution of age among non-default observations is much more even and does not concentrate to the left so much.

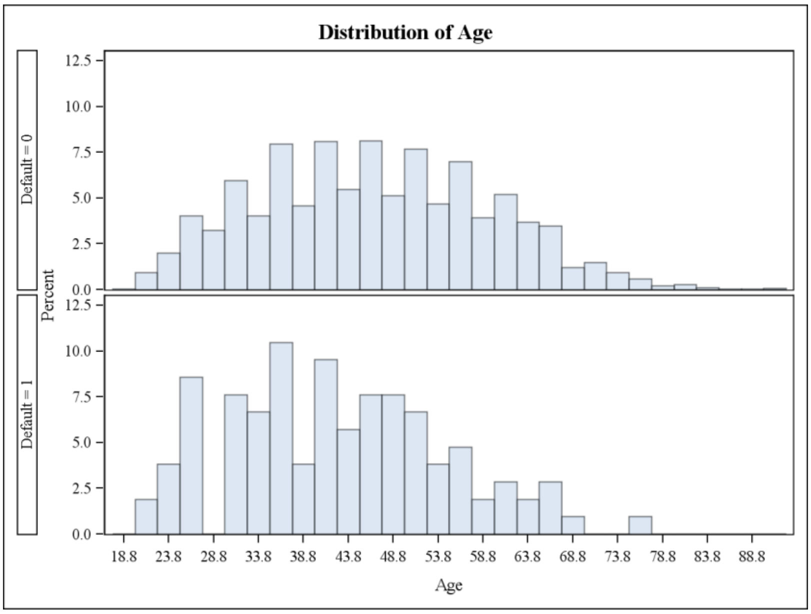


Figure 6: The distribution of age in contracts with default and no default.

To sum up we might say that the age of the customer decrease the probability of both types of payment problems. That is how we expected it to be. In the previous studies Agarwal et al. (2008) found similar results:

Younger borrowers had a higher probability of default than the older ones. In their study the median age of the borrowers was 40 years as for our median age for defaulted customers was 42 and for non-defaulted customers 46. In the study of Kocenda and Vojtec (2009) included age in their study with variable "Date of Birth" which had positive relationship with default. This means that the actual age of the customer had a negative relationship with default. Even though the relationship was quite week having the 18th best information value out of 22 variables.

7. CarPremium: Agarwal et al. (2008) found that the loans for most luxury automobiles have a higher probability of prepayment, while loans for most economy automobiles have a lower probability of default. This suggests that the variation in performance of premium car loans is high, so the risk of default is more than average but also the risk of prepayment is also high. Those evidences indicate that loans for luxury cars include more credit risk than loans for more economy cars. As we discussed in chapter 2.1 (The Definition of Risk) the variation increases risk because of the unknown change which exists in luxury car loans. If the interest rate level is the same for economy car loans and luxury car loans (as it is in our case), we must state that economy car loans are more profitable segment for the finance company.

When the CarPremium variable explained default and overdue situations alone it succeeds best by explaining defaults in company loans. In company loan defaults and private customer overdues CarPremium was statistically significant explanatory variable having positive effect on payment problems. Actually in single estimations it got a positive effect in three regressions out of four. In private customer defaults its effect was negative (-0.0572). CarPremium survived only to one final model and that was PRIVATE.OVER. In that model CarPremium had a positive effect (0.4607) so we might state that premium cars can explain only private customer overdues in statistically significant way. That result is also visible from Table 19 which shows the average values for the CarPremium variable where the PRIVATE.OVER gets the highest average value (0.13).

Our results suggest that when private customers take a car loan for a premium car they are more likely to overdue than default. In company loans that is other way around so companies with premium car loans are more likely to default than overdue.

8. LastPayment: The results concerning the LastPayment variable were very interesting. We did not found this variable from previous studies so it made this variable even more interesting. We thought that this variable could perform better in this study than it actually did. As a single estimator it performed well being statistically significant in three regressions out of four. In all our regressions LastPayment had positive effect on payment problems as we expected. What we did not expect was that the effect was so slight. Actually the effect was so weak, it did not survive to any of our final models. According to our average numbers we might state that if customer has payment problems with a loan it is more likely that the loan includes a bigger last payment than not. Even so, LastPayment variable cannot explain those payment difficulties.

The possibility for bigger last payment is a way to make monthly payment smaller. Alternative ways are to pay bigger down payment or make the duration of payment schedule longer. According to our results those two alternative methods are used before considering a bigger last payment. Only when the down payment is already at a maximum level and the maturity of the loans is also the longest possible and still the monthly payment is too high the customer or a car retailer requests a possibility of a bigger last payment from the finance company. In such case the LTV and Maturity variables already predict the higher probability of payment problems and a bigger last payment does not make that probability any higher, according to our results.

9. Product3: We had three explanatory variables in our data which tells the facts of collateral of the loans. From those variables Product3 performed best. As a single estimator it was not statistically significant in any of our four regressions. However it managed to get into one of our final models which was the PRIVATE.DEF model. Through that fact we might say that the Product3 variable has the biggest explanatory power when explaining private customer defaults. Strange to say but the direction of the effect of Product3 changed when it was included to the final PRIVATE.DEF model. In single estimation the effect was 0.207 but in final model it was -0.3248. It suggests that having a Product3 as collateral in a private customer loan decreases the probability of default. However according to the average values of the Product3 variable (Table 12) we can see that the average values are higher in the situation of payment problems so the average values suggest positive relationship between payment problems and the existence of Product3. Those results are inconsistent with each other.

10. City: In the previous studies Kocenda and Vojtek (2009) and Agarwal et al.(2008) used the region of the obligor as an explanatory variable. In the research of Kocenda and Vojtek they find out the home city of obligor by using postal codes as we did also in our study. Agarwal et al. got information of in which state the obligor lives. Both previous studies found region-information insignificant. Despite the fact that they did find differences between regions they still did not find them as a remarkable explanatory variables in regression analyses. Our results were in line with previous studies; We did find differences but those differences were not strong enough to survive as an explanatory variable to any of our final models. In our study we divided regions to big cities and little cities. When the City variable got value 1 the obligor lived in one of five biggest cities in Finland and when it got value 0 the obligor lived in some of the smaller cities. As we can see from our average values for our City dummy (Table 16) or from our single estimations the City variable and payment difficulties seems to have a negative relationship. That relationship is statistically insignificant but still it reflects that companies and private customers in smaller cities seem to have more payment problems than customers in the five biggest cities. That result, even though it is statistically insignificant, is quite surprising. The negative effect is the strongest in company customer defaults and second strongest in company customer overdues. This effect suggests that companies which operate in some of the five biggest cities of Finland have less payment problems than companies in smaller cities when auto loans are concerned. Figure 7 shows us the distribution of the City variable values in company customer defaults. That figure shows us the clear difference; the portion of contracts which have value one in the City dummy is much higher in non-defaulted contracts.

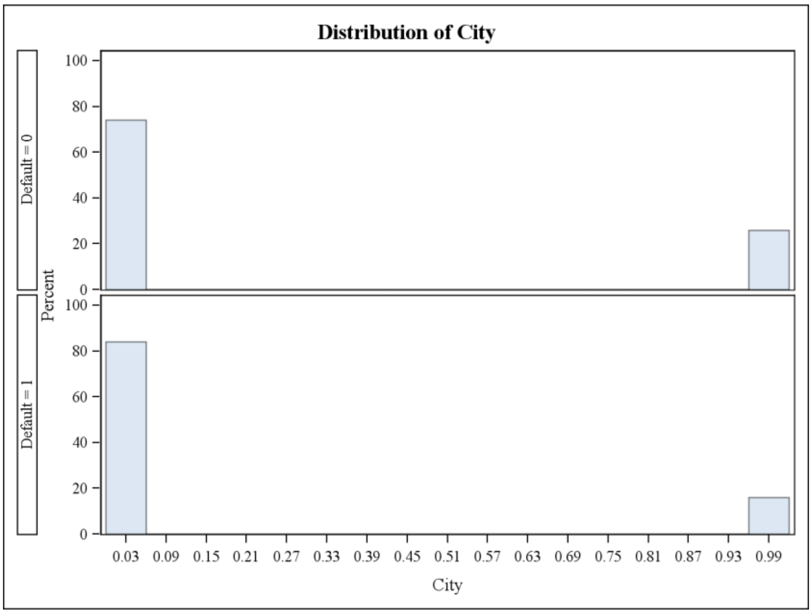


Figure 7: The distribution of the City variable in company customer contracts with default and no-default.

Interestingly the City variable had strongest correlations with CarAge and RatingNum variables. The relationship with CarAge and City is negative which suggests that private customers and companies in smaller cities prefer older cars. The relationship between RatingNum and City was positive. It means that private and company customers in smaller cities seem to get lower rating values.

<u>11. DealMonth:</u> As the 11th rating might tell, the DealMonth variable did not have much explaining power in any of our regressions. It was statistically insignificant in all of our tests. The effects were so small that we are not able to make any suggestions according to them. With this variable we tried to find out if the quality of the contracts varies during the year. Because we could not find any explaining power with our regression analyses we focused to investigate our distribution figures. The overdue distributions were quite constant in private and company customers but when we looked at the distribution figures of default situations we found something interesting which is possible to see from Figures 8 and 9 below.

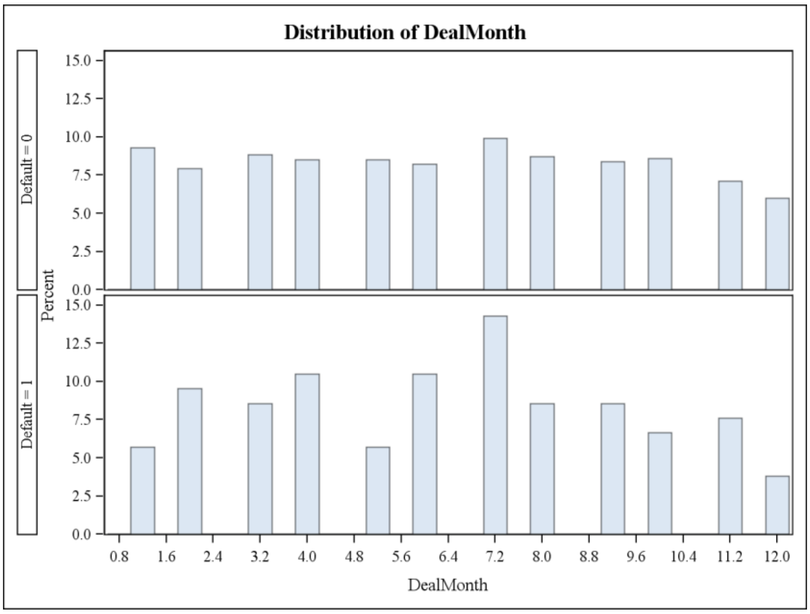


Figure 8: The distribution of the DealMonth variable in private customer contracts with default and no-default.

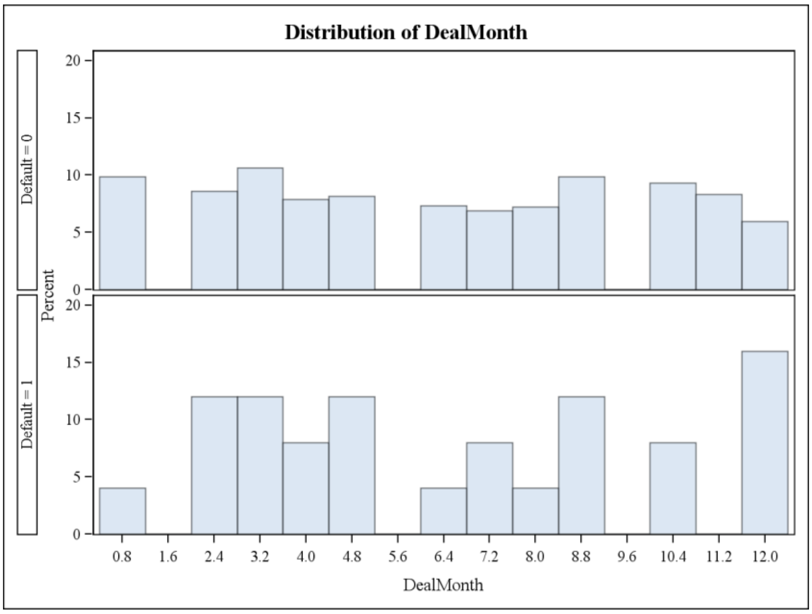


Figure 9: The distribution of the DealMonth variable in company customer contracts with default and no-default.

As we can see there is a default peak in July in Figure 8 and in December in Figure 9. Even those effects are statistically insignificant we might state that there is some kind of weakening in the quality of private customer contracts in the middle of the summer and in company customer contracts in the end of the year.

12: MonthCost: In this study we had two explanatory variables which indicated costs for customer. From those two variables MonthCost performed better as an explainer of payment problems. As a conclusion of these cost-related variables we can easily say that they have hardly any role as a payment problem explainer. In single estimations MonthCost was insignificant in all but one regression. When it explained private customer overdues alone it recorded statistically significant result (0.0768) and also survived to the final PRIVATE.OVER model. That was the only one of those final models which included MonthCost as an explanatory variable. Could this mean that private customers who accept relatively high monthly costs (which are extra costs and does not include interest) are more careless payers and are more likely to overdue than customers who demand and bargain low monthly costs? This has necessarily no connection with probability of default. Those customers might just be careless payers and that is why their payments delay. In all of our regressions MonthCost had a positive relationship with payment problems which is also visible in Table 14 which shows the average values of the MonthCost variable. The average values are systematically higher in the contracts with payment problems. The strongest positive relationship it recorded when it explained company customer defaults alone (0.1105). That was not statistically significant result and did not survive to the final model. In previous studies costs are included to regressions usually by taking only interest rate as an independent variable. In most cases interests are chose by the riskiness of the customer and that is why they include already some information of the creditworthiness of the customer. Because of that, previous results of interest rates explaining power are not comparable with our cost-related results. In out context the interest rate does not reflect any creditworthiness information.

13: Price: In previous studies the amount of a loan is a popular variable. However we wanted to find out what is the relationship between the price of the car and the possible payment behavior. Is it possible that high price cause payment problems because customers over estimate their payment capacity more easily than in cheaper cars or does high price reflect good incomes and thus good payment behavior? Our results are inconsistent with each other. In single estimations only statistically significant relationship we found when we made tests with company customers' loans. We found a very slight positive relationship between Price and default. In other single estimations the relationship was negative. The Price variable survived only to one final model. That model was PRIVATE.OVER where it got very marginal positive effect on payment problems. That positive relationship is surprising. If we look at Table 20 where we see the average prices we can see that the average price is higher in those loans which did not overdue. This means that in the final model PRIVATE.OVER some of the other eight variables have influence to the Price variable and makes its effect positive even it is negative when it explains overdue alone. As a conclusion we might say that the effect of price is very marginal in all of our tests. If we want to find some effect, the high price of the car might cause some delays in private customer loans.

14: Gender: The performance of the explanatory variable Gender was surprisingly low. Naturally this variable was included only to regressions which estimated payment problems of private customers, since it is quite hard to determine a gender for a company. According to the average values there is a clear difference; male customers are more likely to default but females are more likely to overdue. However, according to single estimations Gender is not a statistically significant explainer in any of our regressions and it is not included into our final models either. The Gender variable was the second last variable which dropped out of the PRIVATE.DEF final model so it was the 12th best explainer (out of 19 variables). In the backward elimination process for the PRIVATE.OVER model gender dropped 6th so in this light the Gender variable has more significance as an explainer of default than overdue. In single estimations the effect of the Gender variable was positive when it explained default and negative when it explained overdue situations.

Okumu et al. (2012) used gender as a main explanatory variable in their study which explained defaults with data from Kenya. Okumu et al. did not find statistically significance difference (with confidence limit of 95%) between survival curves of men and women. Kocenda and Vojtek (2009) ended up in to the same result with their regression analysis with the data from Czech Republic. In their study, the Gender variable was the 19th best explainer from 22 independent variables. Our results are in line with the results of previous studies.

15: StartCost: As we mentioned in the interpretations of the MonthCost variable, our two cost-related variables had hardly any role in our regressions. From those two variables StartCost was the worse one. In single estimations its effect was statistically insignificant in all four regressions. However it managed to survive to the final model PRIVATE.OVER having very slight negative effect on overdues of private customer contracts. Interestingly the effects were positive in all other regressions but in PRIVATE.OVER model it was negative. Even so, those effects were so marginal that we cannot make any suggestions based on them.

16: Clock2: The Clock2 variable was formed to describe if the quality of the credit decisions vary during the day. Naturally it was a unique variable because it described features of the credit decision process inside the target company and thus it is not comparable with other studies. From the average values in Table 10 we can see that in the private customer contracts it seems that the Clock2 variable has a negative relationship with overdue so credit decisions after office hours should not be any worse than the ones made during office hours. Actually they should be better. In company loans the effect is other way around and it seems that credit decisions made after office hours are definitely worse. Single estimations confirm the message of the average values. The effect of Clock2 is negative in regressions which explained payment problems in private customers' loans. In the company customers' loans the effect was positive. In all of our four single regression analyses the effect of the Clock2 variable was statistically insignificant. In the COMPANY.DEF and COMPANY.OVER models, the Clock2 variable would have been statistically significant with 90% significance level. This confirms our impression that the Clock2 variable has much more explaining power in estimation of payment problems of company customers. However the Clock2 variable did not survive to any of the four final models.

17: Product2: From our product-describing variables the Product2 variable was the second worst explainer of payment problems. According to the average values there seems to a positive relationship between overdues of private customers and Product2 variable. That suggests that private customers, that have a Product2 as the collateral in their loan, are more likely to overdue. In the company loans the relationship is negative and Product2 fore shadow smaller probability of payment problems. In all our four single regressions Product2 was statistically insignificant explanatory variable and did not survive to any of the final models. Actually it dropped out from final models in very early stage except from the PRIVATE.OVER model from where it dropped as the last variable. However the Product2 was one of the worst explanatory variables in this study.

18: Product7: With the Product7 variable we have almost same suggestions than with Product2. It did not survive to any of the final models and it was statistically insignificant in all our four single regressions. Actually it recorded the worst significance value in this whole study when it tried to explain defaults in company customers' loans. In default estimations the effect of Product7 was negative and in overdue estimations it was positive. However those effects were so marginal that we cannot make any suggestions based on them. We have to state the same what we stated with our Product2 variable; Product7 has hardly any explaining power on payment problems.

19: DealDate: The DealDate variable was the worst explainer in this study. It deserved this questionable honor because of its effects which were small, all insignificant and none of them survived to the final models. Its target was to describe if the quality of the contracts vary over the month. The average values reflect that the loans of private customers which defaulted are made little bit later than contracts which did not default. In company loans it was other way around. Actually all average values were so close to each other that it is impossible to make any suggestions based on them. The results suggest that the day of the month when the contract is made has nothing to do with payment problems in the future.

6 SUMMARY AND CONCLUSIONS

In this study we investigated if variables available at the time of the credit decision have an explaining power for default or serious overdue situations in secured instalment loans. Loans are taken by private or company customers for car purchasing purposes. We also investigated if there exists remarkable differences between loans which defaulted and loans which had serious delays but did not default. As a main research method we used logistic regression which we chose mainly because of its popularity in the previous studies. We wanted to have results which are easy to compare with earlier evidences.

We started our research by defining a few main concepts. At first we defined a word risk which leaded us to the credit risk and credit granting process. After those main concepts we delved into the details of the auto loan industry in Finland. To know how this topic has been researched before we got familiar with seven previous studies. Those studies included a classic of probability of default estimation made by Professor Edward I. Altman (1968) and other papers explaining defaults at common level but we also got up-to-date studies made especially to estimate auto loans defaults. After getting familiar with the theory background and previous studies we introduced our own quantitative analysis. At first we presented our data of 8931 observations including 1 to 60 month instalment loans for private and company customers. Our tests we started by finding out the directions of effects for each independent variable separately. In that stage we made 72 regressions. After that we wanted to know how they explain payment problems together so we made four different models by using the backward estimation procedure. Because the purpose of this study was to find out the role of each variable we decided to make a ranking of variables as the last part of our examination. The variable ranking is our subjective appreciation for the target company’s needs of how each variable affects on payment difficulties.

The answer for the first research question is yes; there are variables which can forecast the probability of payment problems. Not in a very extensive way but at least the directions of the effects were quite clear. We also found a few clear differences between contracts which defaulted and which got only serious payment delays. The main results of this study are summed up in Table 34.

Table 34: The most significant results of this study

<table>
<tr><th>Dependent variable</th><th colspan="2">Most siqnificant explainers</th><th>Remarkable differences</th></tr>
<tr><td rowspan="4">Private customer default</td><td>LTV 6.07</td><td>+</td><td rowspan="16">- LTV and rating value explains probability of default better than probability of overdue

- Premium car increases probability of overdue in private loans but probability of default in company loans

-Product2 increases probability of overdue in private loans but have no explaining power in other models

- If the credit decision is made after office hours the loan is more likely to default or overdue when company loans are concerned. In private loans the time of credit decision has no explaining power

- Men are more likely to default but women are more likely to overdue

- Older customers are less likely to cause defaults or payment delays than younger

- Maturity explains probability of defaults better than probability of overdue

- Domicile matters. Companies operating in big cities are less likely to cause defaults or payment delays. Same holds with private customers but the relationship is weaker.</td></tr>
<tr><td>RatingNum -1.41</td><td>-</td></tr>
<tr><td>Maturity 0.06</td><td>+</td></tr>
<tr><td>Age - 0.02</td><td>-</td></tr>
<tr><td rowspan="4">Private customer overdue</td><td>LTV 2.36</td><td>+</td></tr>
<tr><td>RatingNum -0.90</td><td>-</td></tr>
<tr><td>CarPremium 0.74</td><td>+</td></tr>
<tr><td>Product2 0.73</td><td>+</td></tr>
<tr><td rowspan="4">Company customer default</td><td>LTV 4.45</td><td>+</td></tr>
<tr><td>CarPremium 1.17</td><td>+</td></tr>
<tr><td>RatingNum -0.86</td><td>-</td></tr>
<tr><td>Clock2 0.81</td><td>+</td></tr>
<tr><td rowspan="4">Company customer overdue</td><td>LTV 2.44</td><td>+</td></tr>
<tr><td>RatingNum -0.73</td><td>-</td></tr>
<tr><td>City -0.52</td><td>-</td></tr>
<tr><td>Clock2 0.48</td><td>+</td></tr>
</table>

The best explanatory variable in the whole study was a continuous explanatory variable LTV (loan-to-value). Its strong positive relationship with payment problems was a result which is definitely in line with the previous studies. It also confirmed that the public discussion around the variable is reasonable when secured instalment loans are concerned. In short; a bigger down payment decreases the risk of default or serious payment delays. In company loans and individual loans the LTV explained better defaults than overdue situations. LTV survived to three out of the four final models and was only dropped out from the final model explaining company customer defaults. The second best explanatory variable was RatingNum which was formed as a continuous variable. The RatingNum variable implemented rating values from A to D with numerical values 1 - 4. Because the best rating value A got the number 4 in our dataset it was expected that there should be a strong negative relationship between the rating value and the probability of payment problems. The results confirmed these thoughts and RatingNum recorded negative effects in all of the regressions. The negative effect was slightly stronger in private customers’ contracts. The RatingNum variable also survived to three out of the four final models being dropped off only from the model which explained company customer defaults. The age of the customer was a variable which we got only for the private customers. Its unambiguous result was that the age and the payment problems have a negative relationship. The Gender variable we got also only for private customers. That variable did not survive to the final models but the single estimations revealed that men are more likely to default and women more likely to overdue. The maturity of the loan had a positive relationship with payment problems in all of the regressions. The maturity variable seemed be a better explainer of default situations than payment delays.

We got a few variables indicating what kind of a car is the collateral of the loan. As a whole those variables did not have much explaining power. A premium car was the best explainer of those collateral types. It survived to the final model which explains private customer payment delays having there a positive relationship with overdue situations. Interestingly it got a negative relationship with defaults. In company loans a premium car as collateral increased both the probability of default and the probability of overdue. In this sense it seems that private customers who buy a premium car might overestimate their payment capacity but they still manage to pay their loan back. In company loans the overestimation of the payment capacity leads more often to the default situation. The Product2 variable was the second best collateral type as an explainer. Its only reasonable explaining role was in the model explaining the private customer overdue situations where it got a positive relationship with payment delays.

The Clock2 variable in the dataset reflected whether the credit decision was made during normal business hours or during the last two hours when credit decisions are still available. If the credit decision was made after the normal business hours it decreased the risk of payment problems in private customer loans. In company loans it was other way around and the effect was stronger than in private loans. Even though Clock2 did not survive through the backward elimination process to any of the four final models it was very close to survive to the models explaining company customer payment problems. The age of the purchased car had a positive relationship with payment difficulties in all of the regressions. Its effect was not very strong but it still managed to survive to three of the four final models. The strongest effect the CarAge variable got in company loans. The City variable indicated the domicile of the customer. Its results were unambiguous. Customers living or operating in some of the five biggest cities in Finland are less likely to cause default or serious overdue situations. The relationship was not very strong so it did not survive to any of the final models. However the negative relationship existed in all of the regressions. One of the most interesting variables was the TimeFinance variable which reflected the time between signing the contract and financing it. Its relationship was positive with payment problems in all of the regressions. In company loans its effect was only marginal but in private loans it was significant. The month or date when a contract is made did not have any remarkable explaining power to the probability of payment problems. Also the price of the car, the amount of the last instalment or the costs included to the contract had no significant relationship with dependent variables.

As a managerial implication it is suggested that extra weight should be given for LTV and credit rating in the credit decisions. Those variables seem to outperform other variables more than we expected. For example maturity or car price which have had traditionally a great role in credit granting had hardly any role as an explainer in this study. Clock2 variable revealed that the target company should pay attention to the credit decisions for company customers made after normal business hours. Also when company customers are concerned the target company should pay attention to the contracts where the car is relatively old and the maturity of the contract is long. According to the final model for company customer defaults that combination seems to add the probability of default. The TimeFinance variable revealed that the target company should monitor how long it takes to get contracts inside the finance company after giving a positive credit decision.

Our suggestions for further research are simple. This same research should be repeated after a few years of time and the results should be compared with the results of this study. By doing continuous testing inside the target company it is possible to find out if the direction or the power of the effects change. Through that way it is possible to react to the change and modify the credit policy if necessary. As we stated in the introduction part this is possible only by looking behind the rating symbols. As a general suggestion for further research in the area of the PD estimations we suggest to investigate the relationship of the LTV and rating values more closely. In this research LTV seemed to outperform rating values. Does the down payment really have such a huge role in secured auto loans? How about in mortgage house loans? We have to remember the enormous amount of information which is included into the rating values and they still cannot beat the LTV as an explainer of the payment problems. If the role of the LTV variable is really that huge the suggestions concerning minimum down payment regulations starts to sound reasonable.

The purpose of this study was to reduce the risk of poor decisions concerning credit granting and supervision. As Mr. Buffett once said to reduce risk we have to know what we are doing. We think this study succeed in that field.

REFERENCES

Alhonsuo Sampo, Nisén Anne and Pellikka Tuula: *“Finanssitoiminnan käsikirja”*, Finanssi ja vakuutuskustannus Oy, Helsinki, 2009.

Allison Paul David: *”Logistic Regression Using the SAS System: Theory and Application”,* SAS Institute and Wiley, NC, USA, 2001.

Altman Edward I. and Rijken Herbert: “*Toward A Bottom-Up Approach to Assessing Sovereign Default Risk: An Update”*, <u>International Research Journal of Applied Finance</u>, Vol. 3, No. 2, 2012, pp. 118 – 136.

Altman Edward I., Herbert Rijken, Watt Matthew, Balan Dan, Forero Juan and Mina Jorge: *“The Z-Metrics™ Methodology For Estimating Company Credit Ratings And Default Risk Probabilities”,* RiskMetrics Group, Inc., 2010, Available [PDF-document]: http://pages.stern.nyu.edu/~ealtman/z-metrics.pdf, Referred: 9.11.2012.

Altman Edward. I.: “*Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy”,* <u>Journal of Finance</u>, Vol. 23, No. 4, 1968, pp. 589-609.

Bank for International Settlements: “*International Convergence of Capital Measurement and Capital Standards”*, 2004, Available [PDF-document] http://www.federalreserve.gov/boarddocs/press/bcreg/2004/20040626/attachment.pdf, Referred: 15.12.2012.

Bewick Viv, Cheek Liz and Ball Jonathan: *“Review -Statistics review 14: Logistic regression”,* <u>Critical Care</u>, No. 9, 2005, pp. 112-118.

Burnham Kenneth P. and Anderson David R.: *“Model Selection and Multi-Model Inference: A Practical Information-Theoretic Approach”*, *2nd Edition*, Springer Science+Business Media Inc., New York, Unite States of America, 2002.

Choudhry Moorad: *“An Introduction to Credit Derivatives”,* Elsevier Ltd. United Kingdom, 2004.

Fan Tsai-Hung and Cheng Kuang-Fu: *“Tests and variables selection on regression analysis for massive datasets”,* Data & Knowledge Engineering, Vol. 63, No. 3, 2007, pp. 811-819.

Finanssialan keskusliitto: “Kulutusluottoselvitys 2010”, 2010, Available [PDF-document]: http://www.fkl.fi/materiaalipankki/julkaisut/Julkaisut/Kulutusluottoselvitys%20syksy%202010.pdf, Referred: 11.10.2012.

Hull John C.: *”Risk Management and Financial Institutions” 2th Edition,* Pearson education Inc., Boston, United States of America, 2007.

Hand D.J. and Henley W.E.: *“Statistical Classffication Methods in Consumer Credit Scoring: a Review”,* Journal of the Royal Statistical Society: Series A (Statistics in Society), Vol. 160, No. 3, 1997, pp. 523-541.

Heitfield Erik and Tarun Sabarwal: *“What Drives Default and Prepayment on Subprime Auto Loans?”,* Journal of Real Estate Finance and Economics, Vol. 29, No. 4, 2004, pp. 457-477.

Howells Peter and Bain Keith: *“The Economics of Money, Banking and Finance – A European Text”, 3th Edition*, Pearson Education Limited, Harlow, United Kingdom, 1998.

Hyytinen Ari & Pajarinen Mika: *“Luottomarkkinoiden epätäydellisyydet ja pk-yritysten rahoitusympäristö Suomessa”,* Etlatieto Oy, Kauppa- ja teollisuusministeriö, Edita Publishing Oy, Helsinki, 2005.

Judicial system in Finland: *“Velkajärjestelyhakemuksen täyttöohjeet”*, 2012, Available [PDF-document]: http://www.oikeus.fi/uploads/9ghz0_1.pdf, Referred: 12.10.2012.

Jorion Philippe: *“Value at risk: The New Benchmark for Managing Financial Risk”, 2nd edition,* The McGraw-Hill Companies New York, United States, 2001.

Kuss Oliver: *“Global goodness-of-fit tests in logistic regression with sparse data”,* Statistics in Medicine, Vol. 21, No. 24, 2002, pp. 3789–3801.

Kocagil Ahmet E. and Demir Ebru*: “Empirical Examination of Drivers of Default Risk in Prime Auto Loans”*, FitchRatings, Quantitative Financial Research Special Report, 2007, Available [PDF-document]: http://www.securitization.net/pdf/Fitch/PrimeAuto_15Dec06.pdf, Referred: 30.9.2012.

Kocenda E. and Vojtek M.: *“Default Predictors and Credit Scoring Models for Retail Banking”,* CESifo Working Paper, No. 2862, 2009.

Lix L, Yogendran M, Burchill C, Metge C, McKeen N, Moore D. and Bond R.: *“Defining and Validating Chronic Diseases: An Administrative Data Approach”,* Manitoba Centre for Health Policy, Winnipeg, Canada, 2006, Available [PDF-document]: http://mchp-appserv.cpe.umanitoba.ca/reference/chronic.disease.pdf, Referred: 27.4.2013.

Maness Terry S. and Zietlow John T*.: “Short-Term Financial Management”, 2nd Edition*, Thomson Learning Inc., United States, 2002.

McCormack Gerard: *“Secured Credit under English and American Law”,* Cambridge University Press, Cambridge, United Kingdom, 2004.

Ministry of Finance: *“Finanssimarkkinoiden makrotaloudellisten vaikutusten sääntely ja valvonta”,* Työryhmän muistio, 2012, Available [PDF-document]: https://www.vm.fi/vm/fi/04_julkaisut_ja_asiakirjat/01_julkaisut/07_rahoitus markkinat/20121106Finans/Finanssimarkkinoiden.pdf, Referred: 28.4.2013.

Nagelkerke N.J.D: *"A Note on a General Definition of the Coefficient of Determination”,* Biometrika, Vol. 78, No. 3, 1991, pp. 691-692.

Okumu Arkan Wekesa, Mwalili Samuel and Mwita Peter*: ”Modelling Credit Risk for Personal Loans Using Product-Limit Estimator”,* International Journal of Financial Research, Vol. 3, No. 1, 2012, pp. 22-32.

Peura-Kapanen Liisa: *”Kuluttajien rahatalouden hallinta”*, Kauppa- ja teollisuusministeriö, Edita Publishing Oy, Helsinki, 2005.

Population Register Centre: *“Väestötietojärjestelmä – Kuntien asukasluvut suuruusjärjestyksessä 31.5.2012”*, 2012, Available [Web-page]: http://vrk.fi/default.aspx?docid=6707&site=3&id=0, Referred: 16.12.2012.

Rose, Peter S. and Hudgins, Sylvia C.: *“Bank management and financial services”, 8th edition*, International Edition, The McGraw-Hill Companies New York, Printed in Singapore, 2010.

Schönbucher Phillip J.: *“Credit Derivatives Pricing Models – Models, Pricing and Implementation”*, John Wiley and Sons Ltd, Great Britain, 2003.

Smith Marvin M.: *”What Determines Automobile Loan Defaults and Prepayment?”,* Community affairs publication Cascada, the Federal Reserve Bank of Philadelphia, No. 76 Winter, 2011, Available [PDF-document]: http://www.philadelphiafed.org/community-development/publications/cascade/76/cascade_no-76.pdf, Referred: 7.11.2012.

Statistics Finland*: "Granted loans in Finland 2002-2012"*, Official statistics of Finland, 2012, Available [PDF-document]: http://stat.fi/til/lkan/2012/02/lkan_2012_02_2012-09-14_fi.pdf, Referred: 8.10.2012.

Statistics Finland: *"Kaupan tilinpäätöstilasto, ennakko 2011"*, Official statistics of Finland, 2011, Available [PDF-document]: http://www.stat.fi/til/katipa/2011/katipa_2011_2012-09-20_fi.pdf, Referred: 7.11.2012.

Stephanou Constantinos and Mendoza Juan Carlos: *"Credit Risk Measurement Under Basel II:An Overview and Implementation Issues for Developing Countries",* Working Paper 3556, The World Bank, 2005, Available [PDF-document]: http://www-wds.worldbank.org/servlet/WDSContentServer/WDSP/IB/2006/01/10/000112742_20060110171251/Rendered/PDF/wps35560corrected.pdf, Referred: 11.10.2012.

Agarwal Sumit, Ambrose Brent W. and Chomsisengphet Souphala: *"Determinants of Automobile Loan Default and Prepaymen"*, Federal Reserve Bank of Chicago, Economic Perspectives, 2008 , Available [PDF-document]: http://www.chicagofed.org/digital_assets/publications/economic_perspectives/2008/ep_3qtr2008_part2_agarwal_etal.pdf, Referred: 29.4.2013.

Suomen Asiakastieto Oy: *"Statistics: New payment default entries years 2005-2011"*, 2012, Available [Web-page]: http://www.suomenasiakastieto.fi/asiakastieto/tilastot/ , Referred: 10.10.2012.

Suomen Asiakastieto Oy: *”Malliraportti: Rating Alfan uudet tiedot”,* 2010, Available [HTML-document]: http://www.asiakastieto.fi/pdf/selattavat/malliraportti_rating_alfa.html , Referred: 8.11.2012.

Thomas L.C., Oliver R.W. and Hand D.J.: *”A survey of the issues in consumer credit modelling research”,* Journal of the Operational Research Society, Vol. 56, No. 9, 2005, pp. 1006-1015.

Virolainen Kimmo: *“Macro stress testing with a macroeconomic credit risk model for Finland”*, Discussion papers, Bank of Finland, 2004, Available [Web-page]: http://ssrn.com/abstract=622682 , Referred: 28.4.2013.

Wagenmakers Eric-Jan and Farrell Simon*: “AIC model selection using Akaike weights”*, Psychonomic Bulletin & Review, Vol 11, No. 1, 2004, pp.192-196.

Wyss David: *“Credit Ratings”,* QFinance - The Ultimate Resource, Bloomsbury Information Ltd, London, United Kingdom, 2009.

Xu Lu and Zhang Wen-Jun: *“Comparison of different methods for variable selection”,* Analytica Chimica Acta, Vol. 446, No. 1-2, 2001, pp. 477–483.

Appendices

Appendix 1: Lending portfolio in Finland 2002-2012

Granted loans 2002-2012 for private and company customers (Statistic Finland 2012)

Year	Loans per M euros
2002	112 621
2003	121 603
2004	131 515
2005	146 005
2006	163 992
2007	179 485
2008	203 608
2009	206 406
2010	245 608
2011	276 527
2012/ Q2	301 662

Appendix 2: New payment default entries 2005 – 2011

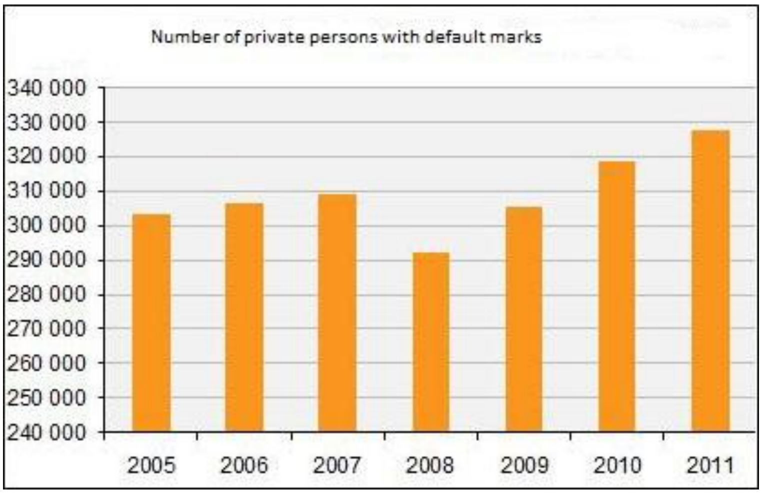
New payment default entries for individuals and companies 2005 – 2011 (Suomen Asiakastieto Oy 2012)

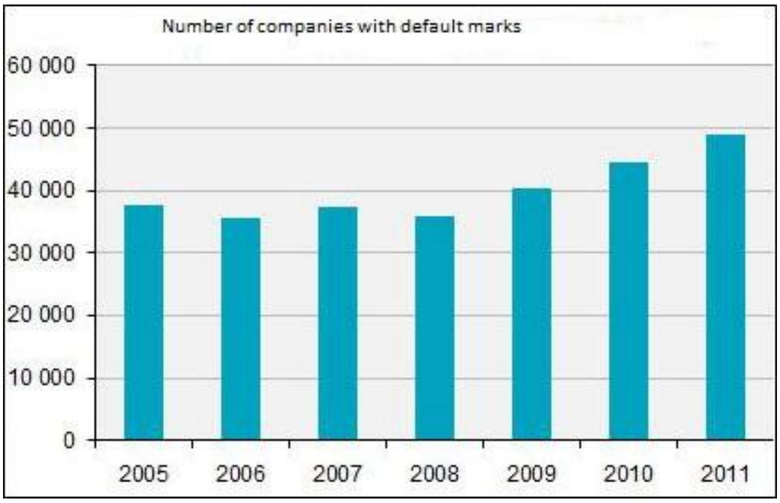
Year	For persons	For companies	All
2011	1 460 450	199 445	1 659 895
2010	1 153 409	162 674	1 314 073
2009	645 046	109 097	754 143
2008	546 713	85 404	632 117
2007	527 256	74 856	602 112
2006	429 606	69 373	498 979
2005	422 542	65 706	488 248

Appendix 3: Consumers and companies with default marks 2005 - 2011

Persons and companies who have a payment default entry 2005-2011 (Suomen Asiakastieto 2012)

Year	Persons	Companies	All
2011	327 491	49 119	376 610
2010	318 713	44 561	363 274
2009	305 439	40 505	345 944
2008	292 454	35 781	328 235
2007	309 296	37 429	346 725
2006	306 610	35 610	342 220
2005	303 200	37 738	340 938





Appendix 4: Pearson Correlation Coefficients

Correlation coefficients between variables in the PRIVATE.DEF model

Pearson Correlation Coefficients Prob > |r| under H0: Rho=0 Number of Observations								
	Age	**CarPremium**	**Product2**	**Product7**	**CarAge**	**DealDate**	**RatingNum**	**Price**
Age	1.00000 7812	-0.06526 <.0001 7812	0.03325 0.0033 7812	0.00159 0.8882 7812	-0.15249 <.0001 7812	-0.02160 0.0563 7812	0.24791 <.0001 7812	0.11597 <.0001 7812
CarPremium	-0.06526 <.0001 7812	1.00000 7813	-0.02557 0.0238 7813	-0.00807 0.4759 7813	0.19951 <.0001 7813	-0.00198 0.8612 7813	-0.07667 <.0001 7813	0.02409 0.0332 7813
Product2	0.03325 0.0033 7812	-0.02557 0.0238 7813	1.00000 7813	-0.00716 0.5270 7813	-0.09743 <.0001 7813	0.00964 0.3942 7813	0.02861 0.0114 7813	0.17450 <.0001 7813
Product7	0.00159 0.8882 7812	-0.00807 0.4759 7813	-0.00716 0.5270 7813	1.00000 7813	0.12462 <.0001 7813	-0.01536 0.1746 7813	-0.04180 0.0002 7813	-0.03628 0.0013 7813
CarAge	-0.15249 <.0001 7812	0.19951 <.0001 7813	-0.09743 <.0001 7813	0.12462 <.0001 7813	1.00000 7813	0.00222 0.8447 7813	-0.37433 <.0001 7813	-0.69303 <.0001 7813
DealDate	-0.02160 0.0563 7812	-0.00198 0.8612 7813	0.00964 0.3942 7813	-0.01536 0.1746 7813	0.00222 0.8447 7813	1.00000 7813	0.00221 0.8451 7813	-0.00515 0.6489 7813
RatingNum	0.24791 <.0001 7812	-0.07667 <.0001 7813	0.02861 0.0114 7813	-0.04180 0.0002 7813	-0.37433 <.0001 7813	0.00221 0.8451 7813	1.00000 7813	0.31822 <.0001 7813
Price	0.11597 <.0001 7812	0.02409 0.0332 7813	0.17450 <.0001 7813	-0.03628 0.0013 7813	-0.69303 <.0001 7813	-0.00515 0.6489 7813	0.31822 <.0001 7813	1.00000 7813

TimeFinance	0.00630 0.5779 7812	0.00047 0.9670 7813	0.04571 <.0001 7813	0.00360 0.7503 7813	-0.09051 <.0001 7813	0.00216 0.8483 7813	-0.00885 0.4341 7813	0.05637 <.0001 7813
Default	-0.03514 0.0019 7812	-0.00164 0.8845 7813	-0.01085 0.3374 7813	0.00555 0.6240 7813	0.02199 0.0519 7813	0.01987 0.0790 7813	-0.08367 <.0001 7813	-0.00689 0.5425 7813
DealMonth	-0.00393 0.7285 7812	0.02777 0.0141 7813	0.00298 0.7925 7813	0.01053 0.3522 7813	0.02039 0.0715 7813	-0.00386 0.7329 7813	-0.02319 0.0404 7813	-0.05182 <.0001 7813
CostMonth	-0.08916 <.0001 7812	0.07881 <.0001 7813	-0.04337 0.0001 7813	0.03398 0.0027 7813	0.24757 <.0001 7813	-0.02446 0.0306 7813	-0.10637 <.0001 7813	-0.15301 <.0001 7813
Clock2	-0.08839 <.0001 7812	-0.00571 0.6140 7813	-0.00555 0.6237 7813	-0.02174 0.0547 7813	0.00975 0.3887 7813	0.00448 0.6923 7813	0.00038 0.9733 7813	-0.02907 0.0102 7813
Maturity	-0.07673 <.0001 7812	0.00484 0.6689 7813	0.04860 <.0001 7813	-0.02112 0.0620 7813	-0.34746 <.0001 7813	-0.00897 0.4282 7813	-0.09065 <.0001 7813	0.38054 <.0001 7813
LTV	-0.16615 <.0001 7812	-0.01166 0.3028 7813	-0.01726 0.1272 7813	-0.00722 0.5234 7813	0.09359 <.0001 7813	0.01994 0.0779 7813	-0.30477 <.0001 7813	-0.17205 <.0001 7813
LastPayment	-0.01181 0.2965 7812	-0.01669 0.1402 7813	0.07731 <.0001 7813	-0.02677 0.0180 7813	-0.30197 <.0001 7813	0.01816 0.1085 7813	0.09347 <.0001 7813	0.42800 <.0001 7813
StartCost	-0.07076 <.0001 7812	0.06503 <.0001 7813	-0.03125 0.0057 7813	0.01554 0.1696 7813	0.15063 <.0001 7813	-0.03617 0.0014 7813	-0.06477 <.0001 7813	-0.01384 0.2212 7813

Product3	-0.16319 <.0001 7812	0.13871 <.0001 7813	-0.14855 <.0001 7813	-0.12292 <.0001 7813	0.44912 <.0001 7813	0.00352 0.7559 7813	-0.26540 <.0001 7813	-0.45527 <.0001 7813
City	-0.05160 <.0001 7812	-0.02650 0.0192 7813	-0.03134 0.0056 7813	-0.03179 0.0049 7813	-0.18753 <.0001 7813	0.01009 0.3727 7813	0.07404 <.0001 7813	0.07862 <.0001 7813
Gender	0.04844 <.0001 7812	0.06109 <.0001 7813	0.03891 0.0006 7813	0.02717 0.0163 7813	0.02908 0.0101 7813	-0.01057 0.3504 7813	0.06636 <.0001 7813	0.09633 <.0001 7813

Pearson Correlation Coefficients Prob > |r| under H0: Rho=0 Number of Observations							
	TimeFinance	**Default**	**DealMonth**	**CostMonth**	**Clock2**	**Maturity**	**LTV**
Age	0.00630 0.5779 7812	-0.03514 0.0019 7812	-0.00393 0.7285 7812	-0.08916 <.0001 7812	-0.08839 <.0001 7812	-0.07673 <.0001 7812	-0.16615 <.0001 7812
CarPremium	0.00047 0.9670 7813	-0.00164 0.8845 7813	0.02777 0.0141 7813	0.07881 <.0001 7813	-0.00571 0.6140 7813	0.00484 0.6689 7813	-0.01166 0.3028 7813
Product2	0.04571 <.0001 7813	-0.01085 0.3374 7813	0.00298 0.7925 7813	-0.04337 0.0001 7813	-0.00555 0.6237 7813	0.04860 <.0001 7813	-0.01726 0.1272 7813
Product7	0.00360 0.7503 7813	0.00555 0.6240 7813	0.01053 0.3522 7813	0.03398 0.0027 7813	-0.02174 0.0547 7813	-0.02112 0.0620 7813	-0.00722 0.5234 7813
CarAge	-0.09051 <.0001 7813	0.02199 0.0519 7813	0.02039 0.0715 7813	0.24757 <.0001 7813	0.00975 0.3887 7813	-0.34746 <.0001 7813	0.09359 <.0001 7813
DealDate	0.00216 0.8483 7813	0.01987 0.0790 7813	-0.00386 0.7329 7813	-0.02446 0.0306 7813	0.00448 0.6923 7813	-0.00897 0.4282 7813	0.01994 0.0779 7813

Pearson Correlation Coefficients
Prob > |r| under H0: Rho=0
Number of Observations

	TimeFinance	**Default**	**DealMonth**	**CostMonth**	**Clock2**	**Maturity**	**LTV**
RatingNum	-0.00885 0.4341 7813	-0.08367 <.0001 7813	-0.02319 0.0404 7813	-0.10637 <.0001 7813	0.00038 0.9733 7813	-0.09065 <.0001 7813	-0.30477 <.0001 7813
Price	0.05637 <.0001 7813	-0.00689 0.5425 7813	-0.05182 <.0001 7813	-0.15301 <.0001 7813	-0.02907 0.0102 7813	0.38054 <.0001 7813	-0.17205 <.0001 7813
TimeFinance	1.00000 7813	0.03733 0.0010 7813	0.01307 0.2479 7813	-0.00953 0.3996 7813	0.00038 0.9733 7813	0.00132 0.9074 7813	0.08214 <.0001 7813
Default	0.03733 0.0010 7813	1.00000 7813	-0.00082 0.9419 7813	0.01086 0.3370 7813	-0.01484 0.1897 7813	0.06840 <.0001 7813	0.08830 <.0001 7813
DealMonth	0.01307 0.2479 7813	-0.00082 0.9419 7813	1.00000 7813	0.02980 0.0084 7813	0.00619 0.5842 7813	-0.02801 0.0133 7813	0.01009 0.3725 7813
CostMonth	-0.00953 0.3996 7813	0.01086 0.3370 7813	0.02980 0.0084 7813	1.00000 7813	0.03840 0.0007 7813	-0.01942 0.0861 7813	0.04725 <.0001 7813
Clock2	0.00038 0.9733 7813	-0.01484 0.1897 7813	0.00619 0.5842 7813	0.03840 0.0007 7813	1.00000 7813	0.01284 0.2566 7813	-0.01026 0.3646 7813
Maturity	0.00132 0.9074 7813	0.06840 <.0001 7813	-0.02801 0.0133 7813	-0.01942 0.0861 7813	0.01284 0.2566 7813	1.00000 7813	0.41904 <.0001 7813
LTV	0.08214 <.0001 7813	0.08830 <.0001 7813	0.01009 0.3725 7813	0.04725 <.0001 7813	-0.01026 0.3646 7813	0.41904 <.0001 7813	1.00000 7813
LastPayment	0.05122 <.0001 7813	0.03895 0.0006 7813	-0.02317 0.0406 7813	-0.06095 <.0001 7813	-0.01791 0.1135 7813	0.27777 <.0001 7813	0.19851 <.0001 7813

Pearson Correlation Coefficients Prob > |r| under H0: Rho=0 Number of Observations							
	TimeFinance	**Default**	**DealMonth**	**CostMonth**	**Clock2**	**Maturity**	**LTV**
StartCost	-0.10389 <.0001 7813	0.00341 0.7634 7813	-0.01217 0.2821 7813	0.58473 <.0001 7813	0.08855 <.0001 7813	0.09544 <.0001 7813	-0.07461 <.0001 7813
Product3	-0.06420 <.0001 7813	0.00140 0.9016 7813	0.03905 0.0006 7813	0.24483 <.0001 7813	0.04820 <.0001 7813	-0.11173 <.0001 7813	0.13958 <.0001 7813
City	0.06092 <.0001 7813	-0.00356 0.7532 7813	-0.01019 0.3676 7813	-0.00614 0.5873 7813	0.04884 <.0001 7813	0.05378 <.0001 7813	0.05487 <.0001 7813
Gender	-0.01912 0.0911 7813	0.01476 0.1920 7813	0.01382 0.2218 7813	-0.00160 0.8872 7813	-0.03965 0.0005 7813	-0.03195 0.0047 7813	-0.05999 <.0001 7813

Pearson Correlation Coefficients Prob > |r| under H0: Rho=0 Number of Observations					
	LastPayment	**StartCost**	**Product3**	**City**	**Gender**
Age	-0.01181 0.2965 7812	-0.07076 <.0001 7812	-0.16319 <.0001 7812	-0.05160 <.0001 7812	0.04844 <.0001 7812
CarPremium	-0.01669 0.1402 7813	0.06503 <.0001 7813	0.13871 <.0001 7813	-0.02650 0.0192 7813	0.06109 <.0001 7813
Product2	0.07731 <.0001 7813	-0.03125 0.0057 7813	-0.14855 <.0001 7813	-0.03134 0.0056 7813	0.03891 0.0006 7813
Product7	-0.02677 0.0180 7813	0.01554 0.1696 7813	-0.12292 <.0001 7813	-0.03179 0.0049 7813	0.02717 0.0163 7813
CarAge	-0.30197 <.0001 7813	0.15063 <.0001 7813	0.44912 <.0001 7813	-0.18753 <.0001 7813	0.02908 0.0101 7813
DealDate	0.01816 0.1085 7813	-0.03617 0.0014 7813	0.00352 0.7559 7813	0.01009 0.3727 7813	-0.01057 0.3504 7813
RatingNum	0.09347 <.0001 7813	-0.06477 <.0001 7813	-0.26540 <.0001 7813	0.07404 <.0001 7813	0.06636 <.0001 7813

Pearson Correlation Coefficients Prob > |r| under H0: Rho=0 Number of Observations					
	LastPayment	**StartCost**	**Product3**	**City**	**Gender**
Price	0.42800 <.0001 7813	-0.01384 0.2212 7813	-0.45527 <.0001 7813	0.07862 <.0001 7813	0.09633 <.0001 7813
TimeFinance	0.05122 <.0001 7813	-0.10389 <.0001 7813	-0.06420 <.0001 7813	0.06092 <.0001 7813	-0.01912 0.0911 7813
Default	0.03895 0.0006 7813	0.00341 0.7634 7813	0.00140 0.9016 7813	-0.00356 0.7532 7813	0.01476 0.1920 7813
DealMonth	-0.02317 0.0406 7813	-0.01217 0.2821 7813	0.03905 0.0006 7813	-0.01019 0.3676 7813	0.01382 0.2218 7813
CostMonth	-0.06095 <.0001 7813	0.58473 <.0001 7813	0.24483 <.0001 7813	-0.00614 0.5873 7813	-0.00160 0.8872 7813
Clock2	-0.01791 0.1135 7813	0.08855 <.0001 7813	0.04820 <.0001 7813	0.04884 <.0001 7813	-0.03965 0.0005 7813
Maturity	0.27777 <.0001 7813	0.09544 <.0001 7813	-0.11173 <.0001 7813	0.05378 <.0001 7813	-0.03195 0.0047 7813
LTV	0.19851 <.0001 7813	-0.07461 <.0001 7813	0.13958 <.0001 7813	0.05487 <.0001 7813	-0.05999 <.0001 7813
LastPayment	1.00000 7813	-0.04604 <.0001 7813	-0.23952 <.0001 7813	0.05203 <.0001 7813	0.02414 0.0328 7813
StartCost	-0.04604 <.0001 7813	1.00000 7813	0.17910 <.0001 7813	0.06267 <.0001 7813	0.00814 0.4719 7813
Product3	-0.23952 <.0001 7813	0.17910 <.0001 7813	1.00000 7813	-0.04562 <.0001 7813	-0.05110 <.0001 7813
City	0.05203 <.0001 7813	0.06267 <.0001 7813	-0.04562 <.0001 7813	1.00000 7813	0.01106 0.3285 7813
Gender	0.02414 0.0328 7813	0.00814 0.4719 7813	-0.05110 <.0001 7813	0.01106 0.3285 7813	1.00000 7813

Correlation coefficients between variables in the PRIVATE.OVER model

Pearson Correlation Coefficients
Prob > |r| under H0: Rho=0
Number of Observations

	Age	Gender	City	CarPremium	DealDate	Product2	TimeFinance	LTV
Age	1.00000 7707	0.05019 <.0001 7707	-0.05135 <.0001 7707	-0.06459 <.0001 7707	-0.01777 0.1187 7707	0.03309 0.0037 7707	0.01144 0.3151 7707	-0.16547 <.0001 7707
Gender	0.05019 <.0001 7707	1.00000 7708	0.01165 0.3067 7708	0.06281 <.0001 7708	-0.01037 0.3624 7708	0.03932 0.0006 7708	-0.01751 0.1242 7708	-0.06178 <.0001 7708
City	-0.05135 <.0001 7707	0.01165 0.3067 7708	1.00000 7708	-0.02883 0.0114 7708	0.00961 0.3990 7708	-0.03158 0.0056 7708	0.05721 <.0001 7708	0.05392 <.0001 7708
CarPremium	-0.06459 <.0001 7707	0.06281 <.0001 7708	-0.02883 0.0114 7708	1.00000 7708	-0.00540 0.6353 7708	-0.02575 0.0238 7708	-0.00265 0.8161 7708	-0.01075 0.3455 7708
DealDate	-0.01777 0.1187 7707	-0.01037 0.3624 7708	0.00961 0.3990 7708	-0.00540 0.6353 7708	1.00000 7708	0.00993 0.3836 7708	0.00093 0.9350 7708	0.01869 0.1008 7708
Product2	0.03309 0.0037 7707	0.03932 0.0006 7708	-0.03158 0.0056 7708	-0.02575 0.0238 7708	0.00993 0.3836 7708	1.00000 7708	0.04659 <.0001 7708	-0.01640 0.1500 7708
TimeFinance	0.01144 0.3151 7707	-0.01751 0.1242 7708	0.05721 <.0001 7708	-0.00265 0.8161 7708	0.00093 0.9350 7708	0.04659 <.0001 7708	1.00000 7708	0.07730 <.0001 7708
LTV	-0.16547 <.0001 7707	-0.06178 <.0001 7708	0.05392 <.0001 7708	-0.01075 0.3455 7708	0.01869 0.1008 7708	-0.01640 0.1500 7708	0.07730 <.0001 7708	1.00000 7708

RatingNum	0.24800 <.0001 7707	0.06785 <.0001 7708	0.07443 <.0001 7708	-0.07814 <.0001 7708	0.00327 0.7739 7708	0.02804 0.0138 7708	-0.00548 0.6304 7708	-0.30172 <.0001 7708
Overdue	-0.07257 <.0001 7707	-0.01026 0.3679 7708	-0.01504 0.1868 7708	0.05993 <.0001 7708	0.00054 0.9625 7708	0.02235 0.0497 7708	0.04710 <.0001 7708	0.09963 <.0001 7708
StartCost	-0.07084 <.0001 7707	0.00623 0.5842 7708	0.06456 <.0001 7708	0.06576 <.0001 7708	-0.03485 0.0022 7708	-0.03147 0.0057 7708	-0.10030 <.0001 7708	-0.07340 <.0001 7708
Clock2	-0.08841 <.0001 7707	-0.03986 0.0005 7708	0.04902 <.0001 7708	-0.00464 0.6841 7708	0.00368 0.7464 7708	-0.00574 0.6142 7708	0.00076 0.9471 7708	-0.00939 0.4097 7708
CostMonth	-0.08678 <.0001 7707	-0.00252 0.8251 7708	-0.00701 0.5380 7708	0.07847 <.0001 7708	-0.02724 0.0168 7708	-0.04351 0.0001 7708	-0.01192 0.2956 7708	0.04751 <.0001 7708
Maturity	-0.07520 <.0001 7707	-0.03393 0.0029 7708	0.05399 <.0001 7708	0.00518 0.6491 7708	-0.01065 0.3499 7708	0.04960 <.0001 7708	-0.00060 0.9579 7708	0.41565 <.0001 7708
LastPayment	-0.01258 0.2694 7707	0.02467 0.0303 7708	0.05093 <.0001 7708	-0.01559 0.1712 7708	0.01853 0.1038 7708	0.07865 <.0001 7708	0.05169 <.0001 7708	0.19712 <.0001 7708
Price	0.11709 <.0001 7707	0.09602 <.0001 7708	0.07955 <.0001 7708	0.02300 0.0434 7708	-0.00524 0.6457 7708	0.17562 <.0001 7708	0.05864 <.0001 7708	-0.17477 <.0001 7708
CarAge	-0.15120 <.0001 7707	0.02929 0.0101 7708	-0.18837 <.0001 7708	0.20054 <.0001 7708	-0.00012 0.9918 7708	-0.09788 <.0001 7708	-0.09417 <.0001 7708	0.09534 <.0001 7708

DealMonth	-0.00379 0.7395 7707	0.01431 0.2091 7708	-0.01123 0.3242 7708	0.02612 0.0218 7708	-0.00426 0.7083 7708	0.00299 0.7933 7708	0.01310 0.2502 7708	0.01064 0.3505 7708
Product3	-0.16306 <.0001 7707	-0.05069 <.0001 7708	-0.04678 <.0001 7708	0.13837 <.0001 7708	0.00150 0.8955 7708	-0.14954 <.0001 7708	-0.07034 <.0001 7708	0.14101 <.0001 7708
Product7	0.00201 0.8602 7707	0.02648 0.0201 7708	-0.03150 0.0057 7708	-0.00775 0.4961 7708	-0.01678 0.1407 7708	-0.00718 0.5288 7708	0.00308 0.7870 7708	-0.00773 0.4976 7708

Pearson Correlation Coefficients
Prob > |r| under H0: Rho=0
Number of Observations

	RatingNum	Overdue	StartCost	Clock2	CostMonth	Maturity	LastPayment
Age	0.24800 <.0001 7707	-0.07257 <.0001 7707	-0.07084 <.0001 7707	-0.08841 <.0001 7707	-0.08678 <.0001 7707	-0.07520 <.0001 7707	-0.01258 0.2694 7707
Gender	0.06785 <.0001 7708	-0.01026 0.3679 7708	0.00623 0.5842 7708	-0.03986 0.0005 7708	-0.00252 0.8251 7708	-0.03393 0.0029 7708	0.02467 0.0303 7708
City	0.07443 <.0001 7708	-0.01504 0.1868 7708	0.06456 <.0001 7708	0.04902 <.0001 7708	-0.00701 0.5380 7708	0.05399 <.0001 7708	0.05093 <.0001 7708
CarPremium	-0.07814 <.0001 7708	0.05993 <.0001 7708	0.06576 <.0001 7708	-0.00464 0.6841 7708	0.07847 <.0001 7708	0.00518 0.6491 7708	-0.01559 0.1712 7708
DealDate	0.00327 0.7739 7708	0.00054 0.9625 7708	-0.03485 0.0022 7708	0.00368 0.7464 7708	-0.02724 0.0168 7708	-0.01065 0.3499 7708	0.01853 0.1038 7708
Product2	0.02804 0.0138 7708	0.02235 0.0497 7708	-0.03147 0.0057 7708	-0.00574 0.6142 7708	-0.04351 0.0001 7708	0.04960 <.0001 7708	0.07865 <.0001 7708

Pearson Correlation Coefficients Prob > |r| under H0: Rho=0 Number of Observations							
	RatingNum	**Overdue**	**StartCost**	**Clock2**	**CostMonth**	**Maturity**	**LastPayment**
TimeFinance	-0.00548 0.6304 7708	0.04710 <.0001 7708	-0.10030 <.0001 7708	0.00076 0.9471 7708	-0.01192 0.2956 7708	-0.00060 0.9579 7708	0.05169 <.0001 7708
LTV	-0.30172 <.0001 7708	0.09963 <.0001 7708	-0.07340 <.0001 7708	-0.00939 0.4097 7708	0.04751 <.0001 7708	0.41565 <.0001 7708	0.19712 <.0001 7708
RatingNum	1.00000 7708	-0.12808 <.0001 7708	-0.06491 <.0001 7708	-0.00120 0.9158 7708	-0.10615 <.0001 7708	-0.08497 <.0001 7708	0.09517 <.0001 7708
Overdue	-0.12808 <.0001 7708	1.00000 7708	-0.00664 0.5600 7708	-0.01318 0.2474 7708	0.02947 0.0097 7708	0.05267 <.0001 7708	0.02933 0.0100 7708
StartCost	-0.06491 <.0001 7708	-0.00664 0.5600 7708	1.00000 7708	0.08940 <.0001 7708	0.58898 <.0001 7708	0.09612 <.0001 7708	-0.04654 <.0001 7708
Clock2	-0.00120 0.9158 7708	-0.01318 0.2474 7708	0.08940 <.0001 7708	1.00000 7708	0.03777 0.0009 7708	0.01356 0.2338 7708	-0.01720 0.1311 7708
CostMonth	-0.10615 <.0001 7708	0.02947 0.0097 7708	0.58898 <.0001 7708	0.03777 0.0009 7708	1.00000 7708	-0.01951 0.0867 7708	-0.06016 <.0001 7708
Maturity	-0.08497 <.0001 7708	0.05267 <.0001 7708	0.09612 <.0001 7708	0.01356 0.2338 7708	-0.01951 0.0867 7708	1.00000 7708	0.27660 <.0001 7708
LastPayment	0.09517 <.0001 7708	0.02933 0.0100 7708	-0.04654 <.0001 7708	-0.01720 0.1311 7708	-0.06016 <.0001 7708	0.27660 <.0001 7708	1.00000 7708
Price	0.32071 <.0001 7708	-0.00728 0.5230 7708	-0.01371 0.2288 7708	-0.02918 0.0104 7708	-0.15287 <.0001 7708	0.38104 <.0001 7708	0.42647 <.0001 7708

Pearson Correlation Coefficients Prob > |r| under H0: Rho=0 Number of Observations							
	RatingNum	**Overdue**	**StartCost**	**Clock2**	**CostMonth**	**Maturity**	**LastPayment**
CarAge	-0.37581 <.0001 7708	0.04869 <.0001 7708	0.15224 <.0001 7708	0.01113 0.3285 7708	0.24686 <.0001 7708	-0.34800 <.0001 7708	-0.29939 <.0001 7708
DealMonth	-0.02379 0.0368 7708	0.00480 0.6735 7708	-0.01425 0.2109 7708	0.00700 0.5388 7708	0.02825 0.0131 7708	-0.02798 0.0140 7708	-0.02285 0.0449 7708
Product3	-0.26555 <.0001 7708	0.02436 0.0325 7708	0.18356 <.0001 7708	0.04835 <.0001 7708	0.24392 <.0001 7708	-0.11309 <.0001 7708	-0.23777 <.0001 7708
Product7	-0.04347 0.0001 7708	0.01566 0.1693 7708	0.01516 0.1831 7708	-0.02145 0.0596 7708	0.03341 0.0034 7708	-0.02252 0.0480 7708	-0.02644 0.0203 7708

Pearson Correlation Coefficients Prob > |r| under H0: Rho=0 Number of Observations					
	Price	**CarAge**	**DealMonth**	**Product3**	**Product7**
Age	0.11709 <.0001 7707	-0.15120 <.0001 7707	-0.00379 0.7395 7707	-0.16306 <.0001 7707	0.00201 0.8602 7707
Gender	0.09602 <.0001 7708	0.02929 0.0101 7708	0.01431 0.2091 7708	-0.05069 <.0001 7708	0.02648 0.0201 7708
City	0.07955 <.0001 7708	-0.18837 <.0001 7708	-0.01123 0.3242 7708	-0.04678 <.0001 7708	-0.03150 0.0057 7708
CarPremium	0.02300 0.0434 7708	0.20054 <.0001 7708	0.02612 0.0218 7708	0.13837 <.0001 7708	-0.00775 0.4961 7708
DealDate	-0.00524 0.6457 7708	-0.00012 0.9918 7708	-0.00426 0.7083 7708	0.00150 0.8955 7708	-0.01678 0.1407 7708
Product2	0.17562 <.0001 7708	-0.09788 <.0001 7708	0.00299 0.7933 7708	-0.14954 <.0001 7708	-0.00718 0.5288 7708
TimeFinance	0.05864 <.0001 7708	-0.09417 <.0001 7708	0.01310 0.2502 7708	-0.07034 <.0001 7708	0.00308 0.7870 7708
LTV	-0.17477 <.0001 7708	0.09534 <.0001 7708	0.01064 0.3505 7708	0.14101 <.0001 7708	-0.00773 0.4976 7708

Pearson Correlation Coefficients Prob > |r| under H0: Rho=0 Number of Observations					
	Price	**CarAge**	**DealMonth**	**Product3**	**Product7**
RatingNum	0.32071 <.0001 7708	-0.37581 <.0001 7708	-0.02379 0.0368 7708	-0.26555 <.0001 7708	-0.04347 0.0001 7708
Overdue	-0.00728 0.5230 7708	0.04869 <.0001 7708	0.00480 0.6735 7708	0.02436 0.0325 7708	0.01566 0.1693 7708
StartCost	-0.01371 0.2288 7708	0.15224 <.0001 7708	-0.01425 0.2109 7708	0.18356 <.0001 7708	0.01516 0.1831 7708
Clock2	-0.02918 0.0104 7708	0.01113 0.3285 7708	0.00700 0.5388 7708	0.04835 <.0001 7708	-0.02145 0.0596 7708
CostMonth	-0.15287 <.0001 7708	0.24686 <.0001 7708	0.02825 0.0131 7708	0.24392 <.0001 7708	0.03341 0.0034 7708
Maturity	0.38104 <.0001 7708	-0.34800 <.0001 7708	-0.02798 0.0140 7708	-0.11309 <.0001 7708	-0.02252 0.0480 7708
LastPayment	0.42647 <.0001 7708	-0.29939 <.0001 7708	-0.02285 0.0449 7708	-0.23777 <.0001 7708	-0.02644 0.0203 7708
Price	1.00000 7708	-0.69442 <.0001 7708	-0.05170 <.0001 7708	-0.45647 <.0001 7708	-0.04027 0.0004 7708
CarAge	-0.69442 <.0001 7708	1.00000 7708	0.02009 0.0778 7708	0.45182 <.0001 7708	0.12721 <.0001 7708
DealMonth	-0.05170 <.0001 7708	0.02009 0.0778 7708	1.00000 7708	0.03779 0.0009 7708	0.00830 0.4661 7708
Product3	-0.45647 <.0001 7708	0.45182 <.0001 7708	0.03779 0.0009 7708	1.00000 7708	-0.12238 <.0001 7708
Product7	-0.04027 0.0004 7708	0.12721 <.0001 7708	0.00830 0.4661 7708	-0.12238 <.0001 7708	1.00000 7708

Correlation coefficients between variables in the COMPANY.DEF model

Pearson Correlation Coefficients, N = 1118
Prob > |r| under H0: Rho=0

	City	CarPremium	CarAge	Product3	Price	LTV	Default	Product7	Clock2	StartCost
City	1.00000	0.00326 0.9133	-0.17631 <.0001	0.11324 0.0001	-0.02169 0.4687	0.10107 0.0007	-0.03321 0.2672	-0.06837 0.0222	-0.01278 0.6695	0.04884 0.1026
CarPremium	0.00326 0.9133	1.00000	0.14064 <.0001	0.17500 <.0001	0.08076 0.0069	0.04075 0.1733	0.06583 0.0277	0.10392 0.0005	-0.01379 0.6451	0.07193 0.0162
CarAge	-0.17631 <.0001	0.14064 <.0001	1.00000	0.19495 <.0001	-0.60428 <.0001	0.17811 <.0001	0.05962 0.0462	0.12390 <.0001	-0.00533 0.8588	0.09714 0.0011
Product3	0.11324 0.0001	0.17500 <.0001	0.19495 <.0001	1.00000	-0.15227 <.0001	0.12785 <.0001	0.05162 0.0845	-0.05592 0.0616	0.01825 0.5421	0.08437 0.0048
Price	-0.02169 0.4687	0.08076 0.0069	-0.60428 <.0001	-0.15227 <.0001	1.00000	-0.24734 <.0001	0.00527 0.8604	-0.03042 0.3095	0.02534 0.3974	0.03070 0.3051
LTV	0.10107 0.0007	0.04075 0.1733	0.17811 <.0001	0.12785 <.0001	-0.24734 <.0001	1.00000	0.08758 0.0034	0.01489 0.6189	-0.01880 0.5300	-0.07417 0.0131
Default	-0.03321 0.2672	0.06583 0.0277	0.05962 0.0462	0.05162 0.0845	0.00527 0.8604	0.08758 0.0034	1.00000	-0.01764 0.5558	0.05188 0.0829	0.04340 0.1470
Product7	-0.06837 0.0222	0.10392 0.0005	0.12390 <.0001	-0.05592 0.0616	-0.03042 0.3095	0.01489 0.6189	-0.01764 0.5558	1.00000	0.00253 0.9325	0.06743 0.0241
Clock2	-0.01278 0.6695	-0.01379 0.6451	-0.00533 0.8588	0.01825 0.5421	0.02534 0.3974	-0.01880 0.5300	0.05188 0.0829	0.00253 0.9325	1.00000	0.07164 0.0166

StartCost	0.04884 0.1026	0.07193 0.0162	0.09714 0.0011	0.08437 0.0048	0.03070 0.3051	-0.07417 0.0131	0.04340 0.1470	0.06743 0.0241	0.07164 0.0166	1.00000
DealDate	0.02260 0.4504	-0.00145 0.9614	-0.02362 0.4301	-0.00218 0.9420	0.00282 0.9249	0.07982 0.0076	-0.00522 0.8616	-0.03929 0.1892	0.01722 0.5652	-0.07097 0.0176
RatingNum	0.17185 <.0001	-0.12791 <.0001	-0.46201 <.0001	-0.11309 0.0002	0.27994 <.0001	-0.34031 <.0001	-0.09520 0.0014	-0.06073 0.0423	-0.05601 0.0612	-0.13012 <.0001
MonthCost	0.01380 0.6450	0.09067 0.0024	0.16256 <.0001	0.04472 0.1351	-0.01111 0.7105	0.05133 0.0863	0.02259 0.4504	0.08495 0.0045	0.02922 0.3290	0.56560 <.0001
LastPayment	0.08100 0.0067	0.06589 0.0276	-0.18187 <.0001	0.03860 0.1972	0.29818 <.0001	0.18001 <.0001	0.07191 0.0162	-0.03718 0.2142	-0.03651 0.2226	-0.05122 0.0869
Maturity	-0.04023 0.1789	0.03183 0.2877	-0.18488 <.0001	-0.04159 0.1647	0.23912 <.0001	0.37381 <.0001	0.12620 <.0001	-0.02288 0.4446	0.05331 0.0748	0.07594 0.0111
FinanceTime	-0.00777 0.7953	0.00241 0.9357	-0.07400 0.0133	-0.00539 0.8572	0.09710 0.0012	0.03334 0.2653	0.00826 0.7826	-0.00596 0.8422	-0.05673 0.0579	-0.02040 0.4955
DealMonth	0.02182 0.4662	-0.01128 0.7063	-0.00090 0.9761	0.05439 0.0691	-0.01307 0.6625	-0.01275 0.6702	0.00932 0.7556	-0.00667 0.8237	-0.02126 0.4775	0.02570 0.3906
Product2	-0.03238 0.2793	-0.13213 <.0001	-0.44030 <.0001	-0.28840 <.0001	0.37677 <.0001	-0.16364 <.0001	-0.02248 0.4527	-0.07014 0.0190	-0.02719 0.3638	-0.09762 0.0011

Pearson Correlation Coefficients, N = 1118 **Prob > |r| under H0: Rho=0**								
	DealDate	**RatingNum**	**MonthCost**	**LastPayment**	**Maturity**	**FinanceTime**	**DealMonth**	**Product2**
City	0.02260 0.4504	0.17185 <.0001	0.01380 0.6450	0.08100 0.0067	-0.04023 0.1789	-0.00777 0.7953	0.02182 0.4662	-0.03238 0.2793
CarPremium	-0.00145 0.9614	-0.12791 <.0001	0.09067 0.0024	0.06589 0.0276	0.03183 0.2877	0.00241 0.9357	-0.01128 0.7063	-0.13213 <.0001
CarAge	-0.02362 0.4301	-0.46201 <.0001	0.16256 <.0001	-0.18187 <.0001	-0.18488 <.0001	-0.07400 0.0133	-0.00090 0.9761	-0.44030 <.0001
Product3	-0.00218 0.9420	-0.11309 0.0002	0.04472 0.1351	0.03860 0.1972	-0.04159 0.1647	-0.00539 0.8572	0.05439 0.0691	-0.28840 <.0001
Price	0.00282 0.9249	0.27994 <.0001	-0.01111 0.7105	0.29818 <.0001	0.23912 <.0001	0.09710 0.0012	-0.01307 0.6625	0.37677 <.0001
LTV	0.07982 0.0076	-0.34031 <.0001	0.05133 0.0863	0.18001 <.0001	0.37381 <.0001	0.03334 0.2653	-0.01275 0.6702	-0.16364 <.0001
Default	-0.00522 0.8616	-0.09520 0.0014	0.02259 0.4504	0.07191 0.0162	0.12620 <.0001	0.00826 0.7826	0.00932 0.7556	-0.02248 0.4527
Product7	-0.03929 0.1892	-0.06073 0.0423	0.08495 0.0045	-0.03718 0.2142	-0.02288 0.4446	-0.00596 0.8422	-0.00667 0.8237	-0.07014 0.0190
Clock2	0.01722 0.5652	-0.05601 0.0612	0.02922 0.3290	-0.03651 0.2226	0.05331 0.0748	-0.05673 0.0579	-0.02126 0.4775	-0.02719 0.3638
StartCost	-0.07097 0.0176	-0.13012 <.0001	0.56560 <.0001	-0.05122 0.0869	0.07594 0.0111	-0.02040 0.4955	0.02570 0.3906	-0.09762 0.0011
DealDate	1.00000	0.06129 0.0405	-0.01676 0.5757	0.06949 0.0201	0.06339 0.0341	-0.03929 0.1892	0.00594 0.8427	-0.02414 0.4199
RatingNum	0.06129 0.0405	1.00000	-0.13442 <.0001	0.06106 0.0412	-0.26492 <.0001	0.03351 0.2629	-0.06091 0.0417	0.24432 <.0001

Pearson Correlation Coefficients, N = 1118 Prob > |r| under H0: Rho=0								
	DealDate	**RatingNum**	**MonthCost**	**LastPayment**	**Maturity**	**FinanceTime**	**DealMonth**	**Product2**
MonthCost	-0.01676 0.5757	-0.13442 <.0001	1.00000	-0.02607 0.3838	0.01996 0.5050	0.02130 0.4767	0.00668 0.8234	-0.06445 0.0312
LastPayment	0.06949 0.0201	0.06106 0.0412	-0.02607 0.3838	1.00000	0.26083 <.0001	0.07557 0.0115	-0.05884 0.0492	0.08298 0.0055
Maturity	0.06339 0.0341	-0.26492 <.0001	0.01996 0.5050	0.26083 <.0001	1.00000	-0.00090 0.9760	-0.05998 0.0450	0.11589 0.0001
FinanceTime	-0.03929 0.1892	0.03351 0.2629	0.02130 0.4767	0.07557 0.0115	-0.00090 0.9760	1.00000	0.04765 0.1113	0.09037 0.0025
DealMonth	0.00594 0.8427	-0.06091 0.0417	0.00668 0.8234	-0.05884 0.0492	-0.05998 0.0450	0.04765 0.1113	1.00000	-0.05175 0.0837
Product2	-0.02414 0.4199	0.24432 <.0001	-0.06445 0.0312	0.08298 0.0055	0.11589 0.0001	0.09037 0.0025	-0.05175 0.0837	1.00000

Correlation coefficients between variables in the COMPANY.OVER model

Pearson Correlation Coefficients, N = 1093 **Prob > |r| under H0: Rho=0**									
	City	**CarPremium**	**Product3**	**Product2**	**Product7**	**Start Cost**	**Ratin Num**	**Finance Time**	**DealMonth**
City	1.00000	0.00238 0.9373	0.11951 <.0001	-0.03940 0.1930	-0.06956 0.0215	0.04845 0.1094	0.17576 <.0001	-0.00745 0.8057	0.02253 0.4568
CarPremium	0.00238 0.9373	1.00000	0.17266 <.0001	-0.12878 <.0001	0.10835 0.0003	0.06380 0.0349	-0.11739 0.0001	-0.01418 0.6396	-0.01594 0.5986
Product 3	0.11951 <.0001	0.17266 <.0001	1.00000	-0.28661 <.0001	-0.05600 0.0642	0.08328 0.0059	-0.10273 0.0007	-0.00941 0.7560	0.04183 0.1670
Product 2	-0.03940 0.1930	-0.12878 <.0001	-0.28661 <.0001	1.00000	-0.07122 0.0185	-0.09676 0.0014	0.23448 <.0001	0.09615 0.0015	-0.05241 0.0833
Product 7	-0.06956 0.0215	0.10835 0.0003	-0.05600 0.0642	-0.07122 0.0185	1.00000	0.06870 0.0231	-0.06345 0.0360	-0.00584 0.8471	-0.00658 0.8279
StartCost	0.04845 0.1094	0.06380 0.0349	0.08328 0.0059	-0.09676 0.0014	0.06870 0.0231	1.00000	-0.12502 <.0001	-0.02217 0.4641	0.02484 0.4121
RatinNum	0.17576 <.0001	-0.11739 0.0001	-0.10273 0.0007	0.23448 <.0001	-0.06345 0.0360	-0.12502 <.0001	1.00000	0.03726 0.2183	-0.05727 0.0584
Finance Time	-0.00745 0.8057	-0.01418 0.6396	-0.00941 0.7560	0.09615 0.0015	-0.00584 0.8471	-0.02217 0.4641	0.03726 0.2183	1.00000	0.05027 0.0967
DealMonth	0.02253 0.4568	-0.01594 0.5986	0.04183 0.1670	-0.05241 0.0833	-0.00658 0.8279	0.02484 0.4121	-0.05727 0.0584	0.05027 0.0967	1.00000
Clock2	-0.00528 0.8616	-0.00647 0.8308	0.00843 0.7808	-0.01927 0.5246	0.00352 0.9074	0.07186 0.0175	-0.05219 0.0846	-0.05384 0.0752	-0.02255 0.4564

LastPayment	0.09044 0.0028	0.02103 0.4874	0.03243 0.2840	0.09350 0.0020	-0.03782 0.2115	-0.06948 0.0216	0.06887 0.0228	0.06307 0.0371	-0.06180 0.0411
MonthCost	0.01425 0.6379	0.09134 0.0025	0.04735 0.1177	-0.06361 0.0355	0.08563 0.0046	0.56567 <.0001	-0.13462 <.0001	0.02217 0.4641	0.00698 0.8178
Overdue	-0.06056 0.0453	0.04062 0.1797	0.02641 0.3831	-0.02495 0.4099	0.01407 0.6423	0.02887 0.3402	-0.16733 <.0001	0.03121 0.3025	-0.04506 0.1365
Maturity	-0.03896 0.1981	0.02359 0.4358	-0.04864 0.1080	0.12187 <.0001	-0.02089 0.4903	0.06909 0.0223	-0.25841 <.0001	-0.00149 0.9607	-0.06231 0.0394
LTV	0.10324 0.0006	0.03843 0.2042	0.12468 <.0001	-0.16581 <.0001	0.01653 0.5852	-0.07994 0.0082	-0.34086 <.0001	0.03341 0.2697	-0.01413 0.6409
Price	-0.02074 0.4934	0.06833 0.0239	-0.15014 <.0001	0.38139 <.0001	-0.03092 0.3071	0.03042 0.3150	0.27716 <.0001	0.09297 0.0021	-0.00727 0.8103
CarAge	-0.17544 <.0001	0.13686 <.0001	0.18941 <.0001	-0.43638 <.0001	0.12640 <.0001	0.09451 0.0018	-0.45472 <.0001	-0.07732 0.0106	-0.00081 0.9787
DealDate	0.02230 0.4614	-0.00051 0.9866	0.00145 0.9618	-0.02276 0.4523	-0.03989 0.1876	-0.07479 0.0134	0.06575 0.0297	-0.04077 0.1780	0.00816 0.7876

Pearson Correlation Coefficients, N = 1093
Prob > |r| under H0: Rho=0

	Clock2	LastPayment	MonthCost	Overdue	Maturity	LTV	Price	CarAge	DealDate
City	-0.00528 0.8616	0.09044 0.0028	0.01425 0.6379	-0.06056 0.0453	-0.03896 0.1981	0.10324 0.0006	-0.02074 0.4934	-0.17544 <.0001	0.02230 0.4614
CarPremium	-0.00647 0.8308	0.02103 0.4874	0.09134 0.0025	0.04062 0.1797	0.02359 0.4358	0.03843 0.2042	0.06833 0.0239	0.13686 <.0001	-0.00051 0.9866

Pearson Correlation Coefficients, N = 1093 Prob > |r| under H0: Rho=0									
	Clock2	**LastPayment**	**MonthCost**	**Overdue**	**Maturity**	**LTV**	**Price**	**CarAge**	**DealDate**
Product3	0.00843 0.7808	0.03243 0.2840	0.04735 0.1177	0.02641 0.3831	-0.04864 0.1080	0.12468 <.0001	-0.15014 <.0001	0.18941 <.0001	0.00145 0.9618
Product2	-0.01927 0.5246	0.09350 0.0020	-0.06361 0.0355	-0.02495 0.4099	0.12187 <.0001	-0.16581 <.0001	0.38139 <.0001	-0.43638 <.0001	-0.02276 0.4523
Product7	0.00352 0.9074	-0.03782 0.2115	0.08563 0.0046	0.01407 0.6423	-0.02089 0.4903	0.01653 0.5852	-0.03092 0.3071	0.12640 <.0001	-0.03989 0.1876
StartCost	0.07186 0.0175	-0.06948 0.0216	0.56567 <.0001	0.02887 0.3402	0.06909 0.0223	-0.07994 0.0082	0.03042 0.3150	0.09451 0.0018	-0.07479 0.0134
RatinNum	-0.05219 0.0846	0.06887 0.0228	-0.13462 <.0001	-0.16733 <.0001	-0.25841 <.0001	-0.34086 <.0001	0.27716 <.0001	-0.45472 <.0001	0.06575 0.0297
FinanceTime	-0.05384 0.0752	0.06307 0.0371	0.02217 0.4641	0.03121 0.3025	-0.00149 0.9607	0.03341 0.2697	0.09297 0.0021	-0.07732 0.0106	-0.04077 0.1780
DealMonth	-0.02255 0.4564	-0.06180 0.0411	0.00698 0.8178	-0.04506 0.1365	-0.06231 0.0394	-0.01413 0.6409	-0.00727 0.8103	-0.00081 0.9787	0.00816 0.7876
Clock2	1.00000	-0.03571 0.2381	0.02315 0.4445	0.05412 0.0737	0.04613 0.1275	-0.02554 0.3988	0.02693 0.3738	-0.01188 0.6949	0.01687 0.5773
LastPayment	-0.03571 0.2381	1.00000	-0.03409 0.2601	0.02245 0.4584	0.26307 <.0001	0.18209 <.0001	0.27466 <.0001	-0.18494 <.0001	0.06648 0.0280

Pearson Correlation Coefficients, N = 1093
Prob > |r| under H0: Rho=0

	Clock2	LastPayment	MonthCost	Overdue	Maturity	LTV	Price	CarAge	DealDate
MonthCost	0.02315 0.4445	-0.03409 0.2601	1.00000	0.02603 0.3900	0.01516 0.6165	0.04873 0.1074	-0.01337 0.6588	0.16254 <.0001	-0.01870 0.5369
Overdue	0.05412 0.0737	0.02245 0.4584	0.02603 0.3900	1.00000	0.06117 0.0432	0.11707 0.0001	-0.05114 0.0910	0.10904 0.0003	0.01241 0.6819
Maturity	0.04613 0.1275	0.26307 <.0001	0.01516 0.6165	0.06117 0.0432	1.00000	0.36843 <.0001	0.24167 <.0001	-0.19567 <.0001	0.06295 0.0374
LTV	-0.02554 0.3988	0.18209 <.0001	0.04873 0.1074	0.11707 0.0001	0.36843 <.0001	1.00000	-0.25553 <.0001	0.17886 <.0001	0.08113 0.0073
Price	0.02693 0.3738	0.27466 <.0001	-0.01337 0.6588	-0.05114 0.0910	0.24167 <.0001	-0.25553 <.0001	1.00000	-0.60550 <.0001	0.00361 0.9052
CarAge	-0.01188 0.6949	-0.18494 <.0001	0.16254 <.0001	0.10904 0.0003	-0.19567 <.0001	0.17886 <.0001	-0.60550 <.0001	1.00000	-0.02275 0.4524
DealDate	0.01687 0.5773	0.06648 0.0280	-0.01870 0.5369	0.01241 0.6819	0.06295 0.0374	0.08113 0.0073	0.00361 0.9052	-0.02275 0.4524	1.00000