Master's Thesis in Finance
Stockholm School of Economics

Consumption Credit Default Predictions[1]

VICTOR JACOBSSON and **SEBASTIAN SIEMIATKOWSKI**

19038@student.hhs.se **18721@student.hhs.se**

Abstract:

Consumption credit plays an increasingly important role in facilitating consumption and enables consumers to smooth consumption. Today, as much as 26% of all card transaction volume in Sweden is made with credit cards. In addition, many retailers themselves offer different types of credit alternatives. However, lending is associated with risks and it is therefore important to be able to correctly predict credit defaults. This paper investigates what factors are important to take into consideration when making credit default predictions by estimating a probit regression model using 170.000 approved consumption credits. While most traditional scoring methods mainly look at financial and demographic variables this paper shows that behavioural variables are at least as important when making default predictions.

Tutor: Associate Professor Andrei Simonov

Presentation: December 7, 2007

Time: 15.15

[1] Credits: Andrei Simonov, Jimmy Wales and Larry Sanger.

1 Introduction

From the beginning of time credit has been used to allow for smoothing of consumption. Lending, borrowing, instalments, payment after or before delivery, consumption credit, all kinds of forms have existed and have been a vital part to smooth transactions and enable economic growth. Today as well, consumer credit alternatives play an important role in the economy to facilitate consumption. In Sweden, where credit card penetration is much lower than in many other industrialized countries, invoicing and purchase by instalment plays an even more important role.[2] In Sweden about 12% of all card transactions are credit card transactions compared to 26% in Germany and 50% in the US. When looking at the total volume of money rather than the number of transactions the findings are even more convincing, only 21% of all card volume in Sweden is in the form of credit transactions, in Germany it is 27% and in the US 68%.[3] Credit allows consumers to smooth consumption in both the long and short term. In the short term consumers can purchase and pay after they receive their salary. In the longer term, younger people for example, might want to maintain consumption at a higher level than their current income allows, in the expectance of increasing income in the near future. Credit rather than prepayment is also often associated with the transfer of transaction risk from the buyer to the seller.

While the facilitation of consumption credits increases purchasing power and hence sales, it also includes risk taking, the risk of not getting paid in time, or not at all. Lenders, be they credit institutions or retailers, minimize risks by trying to predict defaults. Considering the vast amount of credit provided to Swedish consumers, thus enabling them to smooth consumption, it is of great importance for social welfare to improve the lenders' ability to predict defaults. Better default predictions mean that more people can be provided with credit at a lower cost. Many lenders use some type of scoring model to try to predict who will default on their loan. The most commonly used models are developed by external credit reporting agencies and based on primarily public data sources. However, many of the larger lenders have also developed internal credit scoring models.

The consequences of bad credit scoring routines or the lack of credit scoring models can prove devastating, not only to the individual firm but also to the society as a whole. One of

[2] http://www.ita.doc.gov/td/finance/publications/creditcards.pdf
[3] Bank for International Settlements (2005)

the important lessons learned in the wake of the subprime crisis in the second half of 2007 is that not only is it unreasonable issue mortgage with average equity levels of 0.71%. It is even more unreasonable to issue mortgages, no matter the circumstances, without proper documentation and investigation of a debtor's financial situation. In this example about 58% of the mortgages were issued with no or low documentation.[4]

Since most of the research on predicting defaults is made by credit reporting agencies and credit institutions as a part of their ongoing business the availability of analysis of credit defaults is limited. There is some public international research on the area and some Swedish research but it is primarily focused on evaluating banks' lending policy or looking at portfolio risk. In our thesis we will look closer at the determinants of default by estimating a probit regression model based on data from one of Sweden's largest consumer factoring companies. We will not only investigate how common, and publicly available, demographical variables such as income and age affect the probability of default. We will also investigate how behavioural factors, such as time of purchase, can change the probability of default. To our help we develop a framework for analysis in which we categorise the different variables by reason for increased risk. The categories are *direct financial ability*, *indirect financial ability* and *moral hazard*. We test 19 hypotheses as well as compare their relative economic significance. While we find that measures that have been based on publicly available financial and demographic factors still are important, private, behavioural data related to debtors' *indirect financial ability* and *moral hazard*-behaviour are even more important when trying to predict defaults. As a conclusion companies extending credit could benefit from developing specialised scoring models adapted to its particular business.

The thesis is organised as follows: Section 2 reviews the theoretical background and previous research, as well as provides a framework for the analysis, in section 3 the data itself and the work to create the dataset is described. Section 4 describes the hypotheses and the economic reasoning behind as well as the issues that are investigated and the approach taken. Section 5 describes the econometric model and the methodology used while section 6 discusses the empirical findings. Section 7 finishes off with the conclusion and some suggestions for further research.

---

[4] http://money.cnn.com/2007/10/15/markets/junk_mortgages.fortune/index.htm?postversion=2007101609

6. Jacobsson & Siemiatkowski – Consumption Credit Default Predictions

Unfortanetly due to integrity issues we were unable to publish the underlying data accompanying the thesis.

2 Theoretical Framework and Previous Research

In this section, we walk through the theoretical background of credit risk modelling, its purpose and some basics in the use of the models. We do a review of the different types of methods used in credit risk modelling. Finally we also outline the framework we have chosen to structure our data and the types of risk that framework is associated with.

2.1 Credit Risk Management

Credit has always been a vital part of commercial transactions, and important for a well functioning economy. People have become more and more dependent on credit and credit is used not only to finance large personal investments such as house purchases but also to finance other kinds of investment and even consumption. For example credit card penetration which can be seen as a good indicator of our dependence on credit, increased by 43% from 1998 to 2005 in Sweden.[5] However, things have changed since the days when credit was personal, like the one between the local bank and a well known client. Nowadays, lending has become more anonymous and the debtor is rarely known to the party that takes the credit risk. This development has been enabled by the standardisation of transactions and different methods have been developed to control the risk involved. When one extends a loan, the lender has to have some way of estimating the risk of default and account for this risk. The method used when estimating the risk of default for personal loans is called credit scoring, and the importance of credit scoring has increased with the development of different securitisation-techniques. Securitisation has not only led to an even further increase in the distance between the debtor and the lender, but credit scoring is also used in the pricing of the security.

2.2 Credit Scoring

2.2.1 General Purpose

Before the rise of statistical methods to assess credit applications, applicants were assessed based on the lender's previous experience of the debtor and/or the perceived credit worthiness of the applicant. In this process the lender had to rely on the judgement skill of the credit

[5] Economist Intelligence Unit (2007)

application reviewer whose perceptions often were based on accepted myths concerning good and bad debtor characteristics rather than proved relationships.[6] As with any system based on prejudice rather than statistical observations this model has proven to give unsatisfactory results and the effects of using substandard scoring methods can be severe. In the well known US subprime mortgage crisis, it has been reported that many of the underlying mortgages were issued without any or with limited documentation and credit scoring procedures. The method used when screening the applicants was not conducted in an appropriate way, and in many cases not even a basic check of the information supplied by the lenders was performed.[7]

The lack of well developed credit scoring methods can, as shown in the US case, cause substantial losses to the lender. It is therefore important that there is a formalised credit scoring process that is carried out in a scientific and objective way. By doing this the error produced by human factors and wrongfully accepted truths can be eliminated. Moreover, the technological improvements have made the collection and analysis of data easier and cheaper than it used to be.[8]

2.2.2 Regulatory requirements: Basel II

Yet another reason for the application of credit scoring methods is the central role it has come to play in the Basel Accords. The Basel Accords dictate laws and regulations aimed at stabilising the international banking system. It rests on three pillars; Minimum Capital Requirements, Supervisory Review Process and Market Discipline.[9] In the calculation of Minimum Capital Requirements credit risk is an important factor and the better ability one has to estimate credit risk the lower capital requirements are needed. This in turn implies a lower cost of capital and higher profitability for the firm hence an increased return to its owners.[10] Moreover, in the increasingly interconnected financial world the ability to predict defaults accurately is of great importance to the stability of the banking system and thus to the society as a whole. In order to estimate credit risk, lenders are allowed to use default prediction models based on historical data.[11] However, most lenders rely on ratings provided by credit rating agencies and credit reporting agencies as Standard & Poor's, Moody's or

---

[6] Henley and Hand (1997)
[7] http://money.cnn.com/2007/10/15/markets/junk_mortgages.fortune/index.htm?postversion=2007101609
[8] Henley and Hand (1997)
[9] Bank for International Settlements (2004)
[10] Ibid
[11] Ibid

Swedish alternatives such as Upplysningscentralen. These ratings are based primarily on financial and demographical data such as age, income and gender as well as public records of how well a debtor handles her or his financial situation. The Basel accord has received heavy criticism for letting companies rely solely on external credit reporting agency ratings since it may result in cyclically lagging capital requirements.[12]

2.2.3 Application and development of scoring models

Credit scoring models are used today everywhere where credit is extended. Apart from the obvious users such as (e.g.) credit card companies, banks and other financial institutions, credit scoring methods are also used by various retailers, such as mail order companies, internet retailers and other companies extending consumer credit. Basically all companies that provide services or products which are delivered, consumed, or used before payment is made, need to be able to assess the credit worthiness of their customers. Most of the credit scoring research is made by private organisations. Therefore, in both Sweden and the rest of the world a large industry has been built around credit information and scoring of individuals.

Table 1:

List of Swedish and International credit reporting agencies (alphabetical order)

Swedish	International
Business Check	Dun & Bradstreet
Creditsafe	Experian
Dun & Bradstreet (Soliditet)	Equifax
Upplysningscentralen	TransUnion

Source: Upplysningscentralen, Dun & Bradstreet, Creditsafe, Business Check, Experian, Equifax, TransUnion

Apart from the external rating information, many lenders use different types of complementary internal scoring models to increase the accuracy of their credit default prediction.[13] In Sweden, information registered with the Enforcement Authority has a strong impact in credit scoring models developed by credit reporting agencies. If a lender does not receive payment on a bill or a loan which is due, he will first try to collect his debt by hiring a private debt collection agency. If the collection agency fails he may ask the Enforcement Authority to enforce his claim. The Enforcement Authority will send the debtor a letter requiring the debtor to either pay or contest the claim. If the debtor does not contest or pay the claim within a little more than a week, the Enforcement Authority will deliver a verdict which

[12] Altman and Saunders (2001)
[13] Brunner et al (2000)

requires the debtor to pay. At the same time the verdict will become registered by the credit reporting agencies. If the debtor settles the claim the information will be stored for three years before it disappears from the credit reporting agencies' databases. If the debtor does not settle the claim the information will remain in the official database until settled. The Enforcement Authority may use various methods, including seizures, to collect the debt.[14]

2.3 Credit Scoring Methods

2.3.1 General

A general method that is used to create score cards is to first classify historical debtors as good, bad or indeterminate. After dropping indeterminate debtors one looks for characteristics that indicate the propensity to pay and try to estimate their relative importance. Characteristics that are used in credit scoring can be divided into two types; financial and demographic, that describe person characteristics, and behavioural, that say something about the applicant's behaviour.

Table 2:

Type of characteristics	
Financial / Demographic	Behavioural
Sex	Number of late payments
Age	Purpose of loan
Occupation	Exceeded credit limit
Annual income	Prior month's purchase record
Running water	Amount of loan

Common problems that arise when estimating score models is population drift, reject inference and sample selection bias. Population drift is the tendency that population change over time as the environment in which the population is active changes. Reject inference is one of the problems that arise when you try to create new credit risk models based on accepted applicants only. Since the applications are based on previously accepted applications you cannot really tell what has happened to the applicants that are rejected. Sample selection bias is another problem that arises when you construct new models based on an unbiased training set.[15]

[14] The Enforcement Authority: www.kronofogden.se
[15] Henley and Hand (1997)

2.3.2 Review of credit scoring methods in use

Altman (1981) and Henley and Hand (1997) provide good introductions to the field of credit scoring methods.[16] The first credit scoring methods and the most widely used are discriminant analysis and linear regression.[17] They have the advantage of being fairly straightforward to use and are often included in statistical software programs. During the last 30 years a broad variety of scoring methods have been developed and in the later part of this period the technological evolution of computers and computational capacity has enabled the use of expert systems, neural networks and non-parametric methods such as the nearest neighbourhood method as well as time varying models taking the time factor into account. Below we will present the various types of methods applied to credit scoring of consumer loans.[18]

2.3.2.1 Discriminant Analysis

With discriminant analysis one investigates which variables discriminate between two or more naturally occurring groups. In our case the two naturally occurring groups are good and bad debtors where bad debtors are defined as debtors that default on their loans. Durand (1941) was the first to use discriminant analysis to create a scoring system that made predictions on good and bad debtors.[19] His studies are still regarded as one of the most comprehensive, best, and statistically correct applications of discriminant analysis.[20] Criticism of the method has been expressed and discussed by e.g. Eisenbeis (1977, 1978) and Rosenberg and Gleit (1994), the main issue has been that a critical assumption in the model requires the members of the evaluated groups to be multivariate normally distributed.[21] However, Reichert et al (1983) empirically showed that the assumption of normal distribution is not a critical limitation.

---

[16] Altman et al (1981) and Henley and Hand (1997)
[17] Altman et al (1981)
[18] Henley and Hand (1997)
[19] Durand (1941)
[20] Altman et al (1981)
[21] Eisenbeis (1977)
Eisenbeis (1978)

2.3.2.2 Regression

Regression analysis examines the relation of the dependent variable to some independent (explanatory) variables. According to Lachenbruch (1978), a regression model using dummy variables produces a function which is parallel to the discriminant analysis function. Ewert (1969) presented a model for evaluating risks associated with granting of trade credit which correctly classified 82% of the accounts. He also recognised the cost of misclassification but it was not included in the model. Fitzpatrick (1976), Lucas (1992) and Henley (1995) have also made studies describing the use of logistical regression.[22]

2.3.2.3 Logistic Regression (Logit and Probit)

Logistic regression is theoretically a more appropriate statistical tool than linear regression analysis.[23] Many of the conceptual and computational issues inherent in linear regression models are dealt with, e.g. the problem with negative possibility or possibility larger than one. One of the first applications of logistic regression to credit scoring was made by Wiginton (1980) who concluded that it was far better than discriminant analysis.[24] Srinivasan and Kim (1987) and Leonard (1993) have also applied logistic regression on credit scoring. The study was, however, made on commercial loans.[25]

2.3.2.4 Mathematical Programming Methods

Mathematical programming, or optimisation, is the study of problems in which one seeks to minimise or maximise a function by choosing the values of real or integer variables from an allowed set.[26] A typical task could be to minimise the number of incorrectly classified loan applicants. Researchers e.g. Hand (1981), Showers and Chakrin (1981) and Kolesar and Showers (1985) describe various mathematical programming methods used to maximise the proportion of correctly classified applicants, e.g. by using integer/linear programming.[27]

---

[22] Fitzpatrick (1976)
Lucas (1992)
Henley (1995)
[23] Henley and Hand (1997)
[24] Wiginton (1980)
[25] Srinivasan and Kim (1987)
Leonard (1993)
[26] Mordecai (2003)
[27] Hand (1981)
Showers and Chakrin (1981)
Kolesar and Showers (1985)

2.3.2.5 Recursive Partitioning

Recursive partitioning creates a decision tree that strives to correctly classify members of the population based on a dichotomous dependent variable. It was originally developed for use in life sciences and Breiman et al (1984) are one of its most important references.[28] However, there have also been examples of the method used in credit scoring by for example Mehta (1968) who developed a partitioning method to minimise cost and Boye et al (1992) who compared the method to discriminant analysis.[29]

2.3.2.6 Expert Systems

An expert system can be compared to the online help files readily available for software programme users. By asking questions one is guided to the correct answer, in the case of credit scoring to determine good and bad credits. One advantage of this method is that it is easy to explain why an applicant was rejected. There is however not much written in this field but Zocco (1985) and Davis (1987) provide some insights.[30]

2.3.2.7 Neural Networks

Henley and Hand (1997) describes neural networks as:

"A statistical model involving linear combinations of nested sequences of non-linear transformations of linear combinations of variables"

The application of this methodology seem to be somewhat rare but Rosenberg and Gleit (1994) described applications of neural networks to credit decisions and Davis et al (1992) compared them to alternative methods.[31] The mixed performance of the method has made lenders sceptic about switching from functioning and well established credit scoring methods.[32]

---

[28] Breiman et al (1984)
[29] Mehta (1968)
Boyle et al (1992)
[30] Zocco (1985)
Davis (1987)
[31] Rosenberg and Gleit (1994)
Davis et al (1992)
[32] Vellido et al (1999)

2.3.2.8 Smoothing Nonparametric Methods

The most common non-parametric method is the nearest-neighbourhood method which classifies applicants depending on what group they resemble most. Chatterjee and Barcun (1970) studied personal loan applications using this method and Henley and Hand (1996) studied data from a large mail order company.[33] One of the advantages is that the data is easy to update, thereby avoiding the problem with population drift. A problem with the method is the computational demand in storing the data, and the classification of applicants using a huge set of variables.[34]

2.3.2.9 Time Varying Models

Credit scoring models generally tries to classify good and bad debtors. However, depending on legislation and other imposed characteristics, this goal may not by default be the best to aim for by a profit maximising organisation. Depending on the nature of the lending in some cases where the total debt of a debtor becomes smaller and the interest rates increase, as with for example credit card debt, to minimise the number of bad lenders can be subordinated to the goal of forecasting debtors that will prepay their loans. In the end credit risk is the risk of financial losses and therefore should be weighed against the risk of for example prepayment. A financial loss on a prepayment typically occurs when the lender has paid a commission to the retailer from whom the claim originated. When the debtor prepays the lender has not had the time to earn even the cost of the commission. Also lenders can be good or bad depending on the circumstances and how they change over time, e.g. the importance of a payment remark can decrease as the frequency rises. Bierman and Hausman (1970), Dirickx and Wakeman (1976) and Srinivasan and Kim (1987b) all use profit based approaches to distinguish good lenders from bad.[35] Roszbach (2003), from the Swedish Riksbanken, use statistical data from Swedish banks and among other things recognises not only the risk of default but also the prepayment risk.[36]

---

[33] Chatterjee and Barcun (1970)
[34] Henley and Hand (1996)
[35] Bierman and Hausman (1970)
Dirickx and Wkeman (1976)
[36] Roszbach (2003)

2.3.3 Our regression

Probit and logit models are the most frequently used generalized linear models with binary dependent variables and are attractive to use in modelling problems where the dependent variable can take on only two values, e.g. default or non-default.[37] A probit regression model is similar to the logit regression model and they essentially give similar results.[38] Choosing between the two is basically a matter of taste and after a discussion with our tutor we have chosen to use a probit regression model in our analysis.

2.4 Framework for analysis

To better understand the underlying drivers for why a debtor may default on his/her loans we have created a framework for analysis in which we have divided the variables and our hypotheses into three different classes. The three classes are; characteristics that are directly indicative of a person's financial ability, characteristics that are indicative of their indirect financial ability but where there is not necessarily a clear, intuitive relationship between the dependent and independent variable, and characteristics that are related to a debtor's behaviour and the concept of moral hazard. The first class consist of mainly demographical factors that tell us something about the person's financial reality. For example it might be unreasonable to believe that a person with an annual income of 100,000 SEK will be able to repay 80,000 SEK within a year. The next two categories also include many variables that are behavioural in addition to the demographic variables that are normally used to predict defaults. Some researchers, for example Orgler (1971) have found that behavioural factors are generally more statistically and economically significant predictors of default than the demographical factors. The second category consists of characteristics indicative of a debtor's indirect financial ability, hence how well a person can make judgements of, manage and/or cares about her/his financial situation. A person that has been overdue on debt previously may be less financially able to make financial judgements and young people might be called credit inexperienced, these types of individuals will thus be more likely to default. Finally, there are factors that might indicate moral hazard; people assuming debt they never have the intention of paying. When a person finds himself in a situation where he is unable to pay off his debt, such a person might become self destructive and take on more debt to cover for old debt due, or simply because the marginal loss of one more crown in debt seems to be of no real value to

[37] Altman et al (1981)
[38] Chambers and Cox (1967)

someone that will default on a larger sum of money. In our third group we will form hypotheses on variables that we believe are indicative of this type of behaviour. We will structure our hypotheses according to those categories and this will hopefully make the paper more interesting to read.

3 Data

In this section we present the origin of our data, how it has been collected and how it has been used in the thesis. In addition we give a descriptive overview of the collected data and the various variables.

3.1 Origin

3.1.1 General

This paper is based on approved invoice credits given on purchases made in approximately 900 online stores and service providers in 2006. The data consists of more than 170,000 observations. The source of the data is a proprietary dataset from one of Sweden's largest consumer factoring companies and the dataset was originally created for other purposes internal to the company. It does, however, contain most of the information needed for our study. Purchases/credits amount to between 100 and 10,000 SEK, with a mean of about 600 SEK. Payment due date (duration of credit) is normally 15 to 30 days from the delivery date.

3.1.2 Credit process

The starting point in the credit process is when the consumer enters the checkout procedure and chooses invoice as a payment option. After filling in the invoice information (e.g. name, address etc) the information is submitted to the factoring company. The factoring company evaluates the consumer and approves the credit instantly with the help of a basic scoring model which denies credit to consumers with registered payment remarks. The consumer, now debtor, receives a confirmation that the purchase has been completed and delivery is normally made 0 – 3 days after the order date. If the debtor does not pay a reminder is sent, followed by a debt collection notice[39]. Examples of data collected at the time of purchase are (e.g.) date and time of purchase, store identification number and address as registered by the Swedish Population Registry. We merge this data with a list of store identification numbers that we match with the category of goods it mainly carries.

[39] Sw: "Inkassobrev"

3.1.3 Complementary data

The data provided by the factoring company was then merged with a complementary, proprietary dataset provided by one of the leading credit reporting agencies in Sweden. The credit reporting agency collects private and public data from numerous sources including the Swedish Tax Authorities and the Swedish Enforcement Authority. The dataset contains individual financial and demographic characteristics such as property ownership, marital status, etc.

4 Hypotheses

In this section we formulate and explain our hypotheses. The variables used to test whether our hypotheses should be rejected or accepted are also presented.

The hypotheses are divided into the three groups outlined in the theory section. The first group consists of demographical factors that tell us something about the debtor's direct financial ability to repay a loan. The second group of hypotheses tests a debtor's indirect financial ability, i.e. it consists of variables indicative of a person's ability and/or willingness to make judgements and manage her/his finances. The third group of hypotheses is related to the problem with moral hazard in lending. We hope to contribute by showing that while financial and demographical factors still are important there is much to learn from an applicant's behaviour at the time of application.

4.1 Direct financial ability

The reader might recognise all variables connected to the hypotheses in the category direct financial ability from the Theory-section since they are all variables used in traditional credit scoring models. We would therefore expect them all to be statistically significant.

H1: High income is negatively correlated with probability of default

All else equal a higher income increases a debtor's ability to repay a loan. It is therefore reasonable to assume that a high income would lead to lower default levels.

H2: A high debt burden is positively correlated with probability of default

Adding more debt to an already high debt level should increase the probability of default.

H3: Personal wealth decreases the probability of default

Wealthy people, debtors with a registered wealth of more than 1.5 million SEK,[40] will be more likely to pay off their debt all else equal.

[40] If wealth does not exceed 1.5 million SEK it is not registered by the authorities: http://www.skatteverket.se/funktioner/svarpavanligafragor/privatovrigt/privatformogenhetsskattfaq/20050415vilkareglergallerforformogenhetsbeskattning.5.18e1b10334ebe8bc80001119186.html

H4: Marriage is negatively correlated with probability of default

Marriage is a proof of partnership and if one party fails to meet her/his payments, it is plausible that she/he may rely on help from her/his partner. This should reduce the risk of default. Moreover, all else equal, there are economies of scale in living together which should result in a larger disposable income.

Table 3

Hypotheses: Direct financial ability

#	Hypothesis	Variable(s)
H1	High income is negatively correlated with probability of default	*INCOME2*
H2	A high debt burden is positively correlated with probability of default	*DEFICIT_CAPITAL2*
H3	Personal wealth decreases the probability of default	*TAXED_PROPERTY2*
H4	Marriage is negatively correlated with probability of default	*MARRIED*

4.2 Indirect financial ability

Many of the variables used to test the hypotheses below are, similar to the hypotheses under direct financial ability, well known from earlier studies on credit scoring. We would thus expect them to be statistically significant. There are, however, some hypotheses that we have not seen in the literature before (H7 through H12) which we have added to see whether they have statistical significance and economic relevance.

H5: Age is relevant in determining the probability of default

We test in what way age can be used to predict the probability of default. For example one might expect a higher probability of default among younger people since they are less likely to have defaulted before and hence not screened out in the basic credit approval process. Moreover, they might be less able to make sound calculations on what kind of expenses they can handle. Hence, experience of credit, which generally increases with age, might decrease the risk of default. Finally, older people retiring from full employment might have problems to get accustomed with a lower standard of living which might lead to higher default ratios.

H6: Men are more likely to default than women

Conventional wisdom, and to some extent previous research, says men are less risk averse than women and hence should form a riskier sub group.[41]

H7: People from the countryside are less likely to default

[41] c.f. Fehr Duda (2006) and Charness and Gneezy (2007)

Life on the countryside and in smaller societies is less anonymous than city life. For example, people on the countryside get their mail delivered by a rural mailman who also provides bank services and there is often a personal contact between the rural mailman and the inhabitants of smaller communities. The fact that people are less anonymous implies an increased insight into their financial situation. As an effect it is plausible that this would in turn imply an even greater fear of debt collectors and letters from the Enforcement Authority in the countryside than in the city, as such things might easily become public knowledge. We will therefore investigate whether they are less likely to default than others. We will test our hypothesis by transforming each debtor's zip code to a dummy variable. The dummy will reflect if the debtor receives mail from a regular mailman or a rural mailman, delivering mail on the countryside[42].

H8: People's willingness and/or ability to pay varies between regions

Although perhaps less plausible we find it interesting to investigate whether there are regional differences in the willingness or ability to pay and, hence, if the probability of default varies depending on what region people live in. Some regions, for example, could be affected by macroeconomic changes that have an impact on default rates, another explanation could be cultural differences between regions.

H9: People's probability of default should differ depending on where they were born

It seems plausible that behaviour in managing loans and other types of credit in some way may be an inherited behaviour connected to the values given by parents, friends and the society where one grows up. Moreover, the effects of a payment remark are serious in Sweden but that might not be apparent to someone brought up abroad. Hence, we would like to investigate if the place where you are born might have an impact on your credit worthiness.

H10: City size has an impact on the probability of default

H10 is connected to hypothesis *H7: People from the countryside are less likely to default*. We would like to investigate if the probability of default increases with city size and anonymity.

H11: People living on a care of-address are more likely to default

[42] Sw: Lantbrevbärare

Our theory is that people that are registered on a care of-address have a less stable life situation, and possibly a weaker financial situation and therefore are more likely to default.

H12: People's probability of default should not depend on in which month they were born
We see no reason why probability of default in any way should depend on in what month they were born. But to rule out the contrary we would like to perform a test.

H13: Payment history is relevant when estimating the probability of default
Past paid debt should be negatively correlated with probability of default, late payments could be an indication of both negligence and low credit worthiness but severely late payments, i.e. payments that are substantially overdue, should be strongly correlated with the probability of default. Payments on time, on the other hand, ought to indicate well run personal finances and should have a decreasing effect on the probability of default.

Table 4

Hypotheses: Indirect financial ability

#	Hypothesis	Variable(s)
H5	Age is relevant in determining the probability of default	*AGE; AGE2*
H6	Men are more likely to default than women	*GENDER*
H7	People from the countryside are less likely to default	*COUNTRYMAIL*
H8	People's willingness and/or ability to pay varies between regions	*MAILLAN*
H9	People's probability of default should differ depending on where they were born	*LANCODE*
H10	City size has an impact on probability of default	*INHABITANTS2*
H11	People living on a care of-address are more likely to default	*CO*
H12	People's probability of default should not depend on in which month they were born	*MONTH*
H13	Payment history is relevant when estimating the probability of default	*PREVIOUSUNPAID2; PREVIOUSPAID2; PREVIOUSUNPAIDR2; PREVIOUSPAIDR2; PREVIOUSUNPAIDD2; PREVIOUSPAIDD2*

4.3 Moral hazard

None of the variables below are included in traditional scoring models developed by credit reporting agencies. The main reason is that the information is not available to them. There might be internal rating models that take factors like these into consideration but we did not find any research on this area.

H14: People that submit voluntary information are less likely to default

The reasoning behind this hypothesis is that people that provide extra information voluntarily are more likely to have good intentions with their purchase and thus will be more likely to pay their debts.

H15: Probability of default should differ depending on type of store

Depending on what the credit is used for, i.e. what is to be purchased, the probability of default should differ. Some stores tend to have goods that are more attractive on the second hand market and would thus be more attractive for people taking the "big bath". The big bath is when someone knows they will default on their loans and try to maximize their credit. The big bath phenomenon is related to the field of behavioural economics and Kilborn (2005) provides some insights into the theories of time inconsistency etc.[43] However, the case might also be that customer segments vary across industries and some segments attract less solid customers. In that case this hypothesis might also be included in indirect or direct financial ability, above. Also, probability of default should be negatively correlated with store size since larger stores attract the general public whereas smaller stores are more likely to have a higher proportion of "bad apples" that might be looking for stores with less developed routines in handling problematic customers.

H16: People that try to maximise their credit have a higher probability of default

Sometimes people who are denied a credit at a specific level try to obtain smaller credits. Such behaviour indicates that the person is not interested in a particular product but rather in the credit itself, this can be because the person is more or less aware that they will default and hence feel that they have nothing to lose by obtaining one more credit.[44] Individuals with previously failed purchase attempts are thus more likely to default on their credit if it is approved.

H17: Loan size increases probability of default

A large loan is financially more demanding than a small one, hence larger loans should increase the probability of default, however this effect, one may argue, is of marginal importance when in the debt range of 100 – 4,000 SEK. More important then, is the loan size when viewed from a moral hazard perspective. As previously described people with no

[43] Kilborn (2005)
[44] Niklas Adalberth, Kreditor

intentions of paying their dues may tend to maximize their credit, and this will be reflected in larger mean sums of debt in the default population than in the paying population.

H18: People's email-addresses tell us something about the probability of default

Our theory is that a debtor's email-address is a good indicator of how well organised lives they live and thereby a proxy for how well they may handle their personal financial situation. For example, people that have an email address connected to a broadband supplier in general live more organised lives and are more likely to pay their bills than debtors with an anonymous email address, e.g. a hotmail address. Even when comparing to the reference of supplying no email address at all anonymous email addresses such as hotmail may be used in moral hazard situations, to be able to confirm, order and retrieve information that is often being sent by email. Moreover, many stores demand an email address to accept a purchase. People with bad intentions will avoid their work e-mail address or other e-mail addresses that are more closely connected to their identity.

H19: People ordering at awkward times of the day are more likely to default

Our theory is that people ordering at night are more likely to live a less stable life and are thus more likely to default. This combined with the more anonymous feeling of the night and the fact that more people are intoxicated at night, something that might result in poor decisions and over spending should lead to an increased default risk. As a comparison Felson and Poulsen (2003) has written about how crime is distributed over the course of the day and one can clearly see that crime rates rise during the night.[45]

Table 5

Hypotheses: Moral hazard

#	Hypothesis	Variable(s)
H14	People that submit voluntary information are less likely to default	*SUBM_PHONE TYPE;*
H15	Probability of default should differ depending on type of store	*AVERAGESALES2*
H16	People that try to maximise their credit have a higher probability of default	*FAILEDBUYS*
H17	Loan size increases probability of default	*SUM2*
H18	People's email-addresses tell us something about the probability of default	*DOMAIN_NAME*
H19	People ordering at awkward times of the day are more likely to default	*ORDERTIME*

[45] Felson and Poulsen (2003)

5 Methodology

Here we describe how we went about making the data usable. We will also describe the econometric model used.

5.1 Methodology

5.1.1 Econometric Model

The probit function is the inverse cumulative distribution function associated with the standard normal distribution. *Y*n is the dependent variable that takes on only two values:

$y_n =$ 1

0

We want to model the probability of default, the probability that the consumer does not pay.

P_n = The probability that the *n*th person does not pay, $0 < P_n < 1$

P_n is affected by some independent variables. An example of an independent variable is for example a person's income, denominated X_n. The probability of default expressed as a function of income

Equation 1

$$P_n = E(y_n \mid X_n) = F(\alpha + \beta X_n)$$

where

Equation 2

$F(\alpha + \beta X_n) = \int_{-\infty}^{\alpha+\beta X_n} f(z)dz$ is the cumulative standard normal distribution function

and

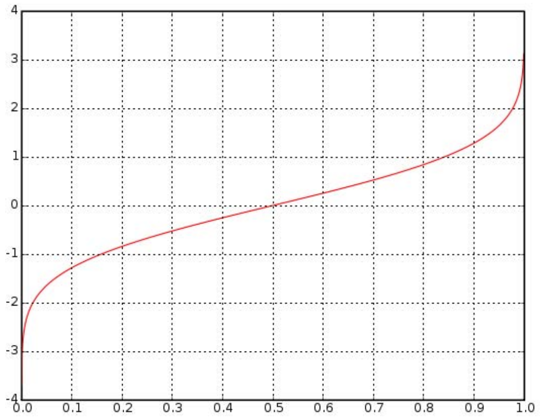
Equation 3

$$f(z) = [1/(2\pi)]^{1/2} \exp(-z^2/2)$$

is the normal density function. Default is determined by a *probit probability model*.

We use the probit probability model to estimate the significance and importance of different variables in the credit decision process.

Figure 1*: Plot of probit function* – *P_n* take values from 0 to 1 on the X-axis and *X_n* take values from *$(-1-\alpha)/\beta$* to *$(1-\alpha)/\beta$* on the Y-axis.



Goodness of fit and inferential statistics is based on the log likelihood and chi-square test statistics.

5.1.2 Regressions

We estimate three regression models: First we look at the demographical data or data derived from demographics in isolation and investigate how they can be used to predict defaults. Second, we run demographic variables together with behavioural variables. Our third regression includes all variables.

Table 6

Regressions

Regression	Type of variables	Comment
1	Demographic	
2	Behavioural and Demographic	Cluster on (*ID*)
3	All variables	Cluster on (*ID*)

Another way to structure the regressions would have been to first run a regression with variables pertaining to *Direct financial ability*, then run a regression on *Indirect financial ability* and finish off with *Moral hazard*. However, we chose to run these three regressions to see what kind of variables were most important to be able to predict default. The model that gets the highest *Pseudo R^2* indicates what model best predicts defaults. We would like to investigate what kind of information is the most valuable; the financial and demographic information available from more or less public databases or the behavioural information that can be extracted from the interaction with the debtor.

The following regression model was estimated when carrying out the third regression, the regression including all variables:

Equation 4

$$DEFAULT = \beta_0 + \beta_1(FAILEDBUYS) + \beta_2(STORE\ CATEGORY) + \beta_3(AVERAGESALES2) + \beta_4(SUM2) + \beta_5(TIMELASTCREDITCHECK) + \beta_6(PREVIOUSUNPAID2) + \beta_7(PREVIOUSPAID2) + \beta_8(PREVIOUSUNPAIDR2) + \beta_9(PREVIOUSPAIDR2) + \beta_{10}(PREVIOUSUNPAIDD2) + \beta_{11}(PREVIOUSPAIDD2) + \beta_{12}(TIMEOFPURCHASE^*) + \beta_{13}(DEFICITCAPITAL2) + \beta_{14}(TAXED\_PROPERTY2) + \beta_{15}(COUNTRYMAILL^*) + \beta_{16}(LANBORN^*) + \beta_{17}(GENDER^*) + \beta_{18}(LIVING\ LAN^*) + \beta_{19}(SUBM\_PHONE^*) + \beta_{20}(MARRIED^*) + \beta_{21}(EMAILDOMAIN^*) + \beta_{22}(CO^*) + \beta_{23}(INHABITANTS2) + \beta_{24}(INCOME2) + (\beta_{25} * \beta_{26}(AGE)) + \beta_{27}(BIRTH\ MONTH^*)$$

Where (*STORE CATEGORY**), (*TIMEOFPURCHASE**), (*LANBORN**), (*LIVING LAN**), (*EMAILDOMAIN**) and (*BIRTH MONTH**) in fact are lists of categories but where an applicant only can belong to one category at a time. E.g. for (*LIVING LAN**) a number of dummy variables are created and set in relation to the most frequently occurring characteristic which in this case is that a person lives in Stockholm. If an applicant is not from (*STOCKHOLM**) but from another region, e.g. (*BLEKINGE**), the estimated risk of default changes with the value for that particular (sub) category. For a full, detailed list of all (sub) variables please find the list in Appendix A1-1.

5.1.3 Definition of default

If the debt is not paid before the recovery notice is due, the loan is normally considered defaulted. However, it is relatively common that the debtor repays her/his debt after the recovery notice is due e.g. when receiving a claim from the Enforcement Authority. We have therefore chosen to define defaults, bad debtors, as credits issued in 2006 and unpaid by the 1 of April 2007. When we run our first regression on demographic variables there are cases with individuals that can be considered both good and bad debtors. That is, they have both paid and unpaid debt. We have chosen to treat those individuals as bad debtors, i.e. defaults, since they have incurred credit losses. When we run the second and third regression we were more interested in the behavioural aspects and we did not want to drop observations where one person had made two different purchases. But since the same person might appear on several occasions we chose to cluster on *ID* to offset the effect that one defaulter appearing several times might give rise to a bias.

5.1.4 Natural logarithm of stochastic variables

The list and description of variables created to test our hypotheses can be found in Appendix A1-1. Worth mentioning is that we in some cases will use the logarithm of numeric values to use as variables to offset the effects of extreme values, e.g. extremely wealthy individuals. This was the case for example with *CITY SIZE, INCOME, WEALTH* and *SUM*. Debtor's age was derived from the social security number and is a numeric variable. We used both *AGE* and the square of age, *AGE2* in our regression to better estimate the effect age has on the probability of default. For *INCOME* it turned out there were some individuals who had negative incomes. After a discussion with our tutor we decided to compensate for this; when using the natural logarithm we set all negative incomes to -9.21034 which is the negative natural logarithm of 10,000.[46]

5.1.5 Deriving demographic data

By using a zip code table from Posten[47] we were able to see what type of mail address the debtors had. The main distinction we want to make is to distinguish whether their mail was delivered by a rural postman or not. To compare the probability of default between regions we assume that people in general live at the address that they have registered with the Tax Authorities. We then group by regions and set the largest region, Stockholm (*STOCKHOLM**), as the default region.

[46] -ln(10,000)
[47] Swedish Post Office

5.1.6 Treatment of missing variables

Dealing with missing variables can be hard but since the amount of individuals with missing values was very low compared to our sample size we decided to drop all observation with missing values instead of applying any of the many techniques used to deal with this problem.

5.1.7 Multicollinearity

To solidify our results we will check for multicollinearity by analyzing the independent variables' intercorrelations. Multicollinearity is the correlation between independent variables. If there is perfect multicollinearity the explanatory power on the dependent variable (default) cannot be isolated and in that case it is not possible to estimate all of the coefficients in the model.[48]

[48] Brooks (2007)

6 Empirical Findings

In this section we will describe our empirical results and discuss our findings. First, we will walk you through and briefly describe our results and the differences between our three regressions. Second, we will use the framework described in the Theory-section and discuss our findings in depth. Third, we will discuss some of the limitations we have experienced.

6.1 Regressions

Our first regression consists of demographic variables. The second regression will mainly include behavioural variables whereas our third regression includes all our variables. For each regression we will first discuss the relative statistical significance of the regression itself before we continue to describe our results.

6.1.1 First regression

6.1.1.1 Regression statistical significance

In our first regression we focus on the demographic variables of the debtors. First we want to investigate whether there is a chance that all regression coefficients are simultaneously equal to zero. This is indicated by the *Prob > Chi²*-measure. If that was the case our regression would not be statistically significant. Since the *Prob > Chi²*-measure is close to zero on the 5% level we can see that the regression coefficients are not zero at the same time. It tells us that there is an extremely low chance of getting a chi-square statistic of 1369.89 or more if there would be no effect of the independent variables. We also have the *Pseudo* R^2-statistic at 0.08145. In Ordinary Least Squares-regression (OLS) this is a good measure of the explanatory power in the estimated model. However, since this is not a typical OLS R^2-statistic, but the McFadden's *Pseudo* R^2, it does not really tell us that much on an absolute basis but we will be able to use it to compare this model with the other two regressions since the statistic is calculated on the same data and predicts the same outcome.

6.1.1.2 Statistical and economic significance of variables

Among the most significant variables were *INCOME2*, *AGE* and *DEFICITCAPITAL2* with a Z-statistic of -21.6, -8.7 and 8.26 respectively. Both income and age decrease the probability of default as they increase, all else equal, whereas a high debt burden increases the probability of default. This is in line with our hypotheses and H1, H2, H4 are accepted on a 95% confidence interval level. Personal wealth (*TAXED_PROPERTY2*) is on the other hand not as significant as we had thought and was rejected on the 95%-level. On the 90%-level it would have been accepted and in line with our hypothesis, that it reduces the probability of default. We arrive at the same conclusion for the variable *MARRIAGE* which has a negative impact on the probability of default but is not significant in our regression.

The regional differences, such as where people live or where they were born is significant in some cases, hence partly in line with our hypothesis. We will discuss these regional differences further in section 6.2.

City size (*INHABITANTS2*) does, however, not have a significant impact on the probability of default. *COUNTRYMAIL** and *SUBMITTED_PHONE** on the other hand are variables that both are significant and have the expected impact, reducing the risk of default. We found it interesting to see that people living in the countryside are better at paying their bills than the rest of us.

We continue by reporting the marginal effects displayed in Appendix A2-1b. Here one can clearly see that *INCOME2*, *AGE2* and *DEFICITCAPITAL2* while being highly significant their marginal effect is lower. Deficit of capital (*DEFICITCAPITAL2)*, large capital costs relative to capital income, increases the default risk by 21% simply when going from 304 SEK to 23.819 SEK. Income (*INCOME2)*, which has the strongest statistical significance, also has the largest economic significance of those three variables. The marginal effect of income, moving one standard deviation, decreases the default ratio by close to 29%. However, since a movement with one standard deviation represents going from 70.000 to more than 2.000.000 in annual income it may be of limited practical use. Finally the marginal effect of age is decreasing the default ratio by about 28 % when moving from a 34 to 46 year old.

Interestingly, we can see that many of the dummies have quite large impact on the default ratio. Having a Gmail e-mail address (*GMAIL**) rather than not having specified an e-mail address decreases the risk of default by 42% while an MSN e-mail address (*MSN**) increases the probability of default by 47%. If we would have to speculate into why this is the case we would guess, based on our own prejudice, that Gmail itself attracts users of higher education and hence better financial abilities, MSN or Hotmail-users, on the other hand, might just be looking for a free, anonymous e-mail address.

Living in Skåne (*SKÅNE**) compared to Stockholm (*STOCKHOLM**) surprisingly increases the default ratio by 40% while Västerbotten (*VÄSTERBOTTEN**) residents have a 42% lower probability of default. More controversial is the immigrant/foreigner dummy (*IMMIGRANT/FOREIGNER**) extracted from the social security number which increases the probability of default by 45%.

Finally, surprisingly the month of birth gave a high increasing effect on default, something we did not expect. For example probability of default increases by as much as 31% for debtors born in February instead of March[49].

6.1.2 Second regression

6.1.2.1 Regression statistical significance

In our second regression we wanted to investigate how the statistical significance changed when one estimates a model based mainly on transaction specific variables, i.e. variables collected at the time of purchase. We did however include some basic demographic variables: *AGE, AGE2*, *INCOME2* and *GENDER**. In the second regression the *Pseudo* R^2-statistic changed dramatically, from 0.0814 to 0.2135. This is in line with findings by Orgler (1971) mentioned above in section 2, who recognised that the behavioural characteristics were more statistically significant predictors of default than the demographic factors.[50]

6.1.2.2 Statistical and economic significance of variables

We can see that the demographic factors did not change dramatically but their marginal effect decreased somewhat, e.g. the marginal effect on income (*INCOME2*) went from 29% in the first regression down to approximately 19%.

One of the most statistically significant type of variables as well as the ones that have the highest marginal effect are the hour of the day (*ORDERTIME**) when the purchase was made.

[49] March is the month when most debtors were born.
[50] Orgler (1971)

We found this especially interesting since, to our knowledge, no such study has been made previously. People who have made orders between midnight and 4 a.m. or between 6-7 a.m. were much more likely to default than people ordering between 9-10 p.m., the time of day when most purchases are made and hence the base case. The probability of default increases by as much as 133% and 108% when purchases are made between 2 and 3 or 3 and 4 respectively. Essentially all purchases made at awkward times of the day had a statistically, as well as economically, significant effect. For some reason lunch hours also had a significant negative effect on probability of default.

Looking at the goods purchased a couple of them had a significant impact on the risk of default, most of them, apart from the category *GADGETS** decreased the risk of default as compared to category *OTHER**. Looking at marginal effects *CARS** were most dominant, decreasing the risk of default by 28%.

When looking at how the debtor has managed previous debt with the factoring company, we needed to take into account that the proportion of the sample that had more than one transaction was approximately 25%. The variables contain the number of paid or unpaid invoices at the time of purchase, and since most debtors had only made one purchase, moving one standard deviation represented a move from 0.04 to 0.247. This is quite pointless considering that a purchase only can take a discrete value. We have therefore corrected for this by, instead of moving one standard deviation, calculated the change in default ratio when going from 0 to 1 paid or unpaid invoice (*PREVIOUSPAID2, PREVIOUSUNPAID2*), reminder (*PREVIOUSPAIDR2, PREVIOUSUNPAIDR2*) and debt collection (*PREVIOUSUNPAIDD2, PREVIOUSPAIDD2*). After this adjustment we can clearly see that e.g. a previously paid invoice (*PREVIOUSPAID2)* reduces the risk of default by as much as 66% (moving 1,6 standard deviations) while an unpaid invoice (*PREVIOUSUNPAID2*) at the time of purchase increases the risk by 70 % (moving 2,8 standard deviations). These results are more applicable than the 60 standard deviations event required to move from 0 to 1 unpaid debt collection (*PREVIOUSUNPAIDD2)*. Even if it is both statistically and economically significant its practical use is very limited due to its infrequency.[51] The paid debt collection

[51] According to the factoring company debtors with unpaid debts in debt collection should by design be blocked from taking on more debt. This explains the low frequency of purchases made when unpaid debt collection claims exist.

(*PREVIOUSPAIDD2*) however gave a little bit more sound results with an 82% increase in risk moving 6.2 standard deviations.

We have a couple of other variables that also were of discrete character and where we corrected the number of standard deviations to get an indication of their marginal effects' importance on a viable change in the variable. The number of failed purchase attempts (*FAILEDBUYS*) for example stated that the increase in default was 9% when having two previously failed transactions before the approved one (representing a 0.81 standard deviation move). Continuing we have store size (*AVERAGESALES*), sum of debt (*SUM*) and time since last credit report (*TIMELASTCREDITCHECK*), which all were significant (-6.44, 16.9 and 8.77 respectively). Moving one standard deviation, from 336 SEK to 846 SEK in debt sum, accounted for an increase of 26% in the probability of default and moving from 8 to 44 days for time since last credit report increased the risk by 9% which gives a feeling for the importance of new data when making credit decisions. Finally one can clearly see that large retailers have lower credit losses. A purchase made with an online retailer that made 20 sales per day instead of 6 (representing a move of one standard deviation) decreased the probability of default by 12%.

6.1.3 Third regression

6.1.3.1 Regression statistical significance

The third regression which incorporates all our variables of interest was the one with the highest explanatory power when looking at the McFadden's *Pseudo* R^2 which reached 0.2351. Interestingly adding the rest of the demographic variables increased the *Pseudo* R^2 by only 10%, which once again reflects the relative importance of behavioural factors compared to demographic factors.

6.1.3.2 Statistical and economic significance of variables

Seven variables lost so much in significance that they were rejected on the 5%-level and with the exception for *GADGETS** they were pre-dominantly variables that were of demographic nature. The excluded variables in the final model were *GÖTEBORG AND BOHUS LÄN**, *SKARABORG LÄN**, *GMAIL**, *TELIA**, and people using another e-mail domain (*OTHER**) as well the statistical significance of people born in *JANUARY** and people purchasing

*GADGETS**.[52] Some new variables become significant when compared to the first and second regression. These are *DECEMBER**, *SEPTEMBER**, *FASHION** and *TAXED_PROPERTY*2, hence three demographic and one behavioural variable. *DECEMBER** has the highest marginal effect, increasing the probability of default by roughly 16%. We will continue to discuss the results from our third regression in depth in the remainder of this chapter.

6.2 Discussion

Below we will use the framework described in the Theory-section and discuss our results and subsequently accept or reject our hypotheses. In the framework we have divided the variables into three groups; direct financial ability, indirect financial ability and moral hazard. Since the third regression was the one with the highest statistical significance and largest explanatory power we will use that output to decide whether to accept or reject our hypotheses.

6.2.1 Direct financial ability

Looking at the direct financial ability hypotheses at the set 5% significance level, three out of four hypotheses were accepted.

Table 7

Hypotheses: Direct financial ability

#	Hypothesis	Variable(s)	Decision
H1	High income is negatively correlated with probability of default	*INCOME2*	Accepted
H2	A high debt burden is positively correlated with probability of default	*DEFICIT_CAPITAL2*	Accepted
H3	Personal wealth decreases the probability of default	*TAXED_PROPERTY2*	Accepted
H4	Marriage is negatively correlated with probability of default	*MARRIED*	Rejected

However, all of the accepted hypotheses have a relatively small marginal effect on the default ratio. An increase by one standard deviation decreases probability of default by 15%. However, due to the large spread in income in the population a one standard deviation move represents going from 65,000 to more than 2,000,000. This means that the income variable (*INCOME2*) is of little practical use in the prediction of defaults. The proportion of debtors with a deficit of capital is relatively small as can be seen in Appendix A3-2a. The mean taxed property is as low as 0.14 (meaning most debtors do not have registered property) but moving one standard deviation takes us to only 3.53 and leads to a reduction in probability of default

[52] Please find Appendix A2-3b for the report on regression 3.

with 7.36%.[53] What we can say is that having property does decrease the probability of default but the marginal effect is unclear.

Table 8

Hypotheses: Direct financial ability

#	Variable(s)	Change in prob. of default	Exp(Mean (μ))	Exp(Mean + 1 std dev. (μ + δ²))
H1	*INCOME2*	*-15.11%*	65,609	2,150,843
H2	*DEFICIT_CAPITAL2*	*6.11%*	304	23,819
H3	*TAXED_PROPERTY2*	*-7.36%*	0.14	3.53

6.2.2 Financial ability

As mentioned in section 4 above, some of our hypotheses are of more experimental nature with limited economic theory. We are rather looking for possible connections than testing a hypothesis. This is also reflected in the lower degree of accepted hypothesis in this category compared to financial reality.

Table 9

Hypotheses: Indirect financial ability

#	Hypothesis	Variable(s)	Decision
H5	Age is relevant in determining the probability of default	*AGE; AGE2*	Accepted
H6	Men are more likely to default than women	*GENDER*	Rejected
H7	People from the countryside are less likely to default	*COUNTRYMAIL*	Accepted
H8	People's willingness and/or ability to pay varies between regions	*MAILLAN*	Accepted
H9	People's probability of default should differ depending on where they were born	*LANCODE*	Accepted
H10	City size has an impact on probability of default	*INHABITANTS2*	Rejected
H11	People living on a care of-address are more likely to default	*CO*	Rejected
H12	People's probability of default should not depend on in which month they were born	*MONTH*	Rejected
H13	Payment history is relevant when estimating the probability of default	*PREVIOUSUNPAID2; PREVIOUSPAID2; PREVIOUSUNPAIDR2; PREVIOUSPAIDR2; PREVIOUSUNPAIDD2; PREVIOUSPAIDD2*	Accepted

As can be seen from **Table 9**, four out of totally nine hypotheses (H6 and H10 through H12) can be rejected. We find no evidence for the claim that men are more likely to default than women, nor that people living on a care of address (*CO**) are more likely to default. We did however, contrary to our hypothesis, find some evidence that the month of birth (*MONTH**) had an impact on the default ratio when compared to the reference month March. We will

[53] A few very wealthy individuals that have not defaulted distort the data and that is the reason for why a standard deviation move represents such a small change in absolute terms.

leave to the astrologists to figure out why. Last but not least we could find no evidence for our claim that the feeling of anonymity that exist in larger cities, bring about higher default ratios.

H5 was accepted, age (*AGE2*) reduces probability of default by 17 % when moving one standard deviation, going from 35 to 45 years. Looking at the dummies we found significant statistical evidence that being an immigrant increases the probability of default by 25%. Our speculation is that it might depend on the ability, of people with a foreign background, to fully understand the severity of a registered payment remark with the Swedish Enforcement Authority. We can also see that people who reside in some regions had a significantly changed probability of default compared to people living in Stockholm. People from densely populated areas in the south as e.g. Västra Götaland (*VÄSTRA GÖTALAND**), Skåne (*SKÅNE**) and nearby Västmanland län (*VÄSTMANLAND**) were significantly worse debtors (25, 24 and 24% higher probability of default respectively) than people from less populated areas such as Västerbotten (*VÄSTERBOTTEN**). This was somewhat surprising and the only common factor we could find for the high default ratio-regions was that they, together with Stockholm, attracted and accepted more foreigners than the rest of Sweden.[54] The causality is, however, only our speculation and we have not made any statistical tests. The fact that being from a less populated region like Västerbotten decreases probability of default with as much as 22% does add some evidence to our H7 which also was accepted at the 5% significance level. It showed that people that get their mail delivered from a rural mailman are approximately 11% less probable to default on their loans.

Looking at previous payment behaviour we noticed that while previous unpaid debt (*PREVIOUSUNPAIDD2*) can be said to be of little importance due to the very few observations, previous paid debt (*PREVIOUSPAIDD2*), previous paid invoices (*PREVIOUSPAID2*), and previous unpaid invoices (*PREVIOUSUNPAID2*) did have a large impact on default levels. When changing from 0 to 1, all affect the default ratio with more than 50%. It is also interesting to see that going from previous paid invoices through previous paid reminders over to previous paid debt, the sign changes and the latter has a negative impact on the default ratio. In line with our hypothesis one can see that it is unclear whether a previous paid reminder (PREVIOUSPAIDR2) is a sign of low creditworthiness or sloppiness.

[54] Statistics Sweden, www.scb.se

6.2.3 Moral hazard

Finally we look at the moral hazard hypotheses which we find the most interesting considering the lack of previous research. Of these hypotheses all had at least one or more significant variables to report. Voluntarily submitting a phone number (*SUBMITTED_PHONE**) decreased the default ratio by closely 12%, indicating that people with no intention to pay will submit a minimum of information. The reason might be to avoid contact in order to prolong the period from the moment when one takes on debt until it is registered with the Enforcement Authority. A similar reasoning might explain the e-mail domains where the reference group is providing no email at all. Domains such as *YAHOO**, *MSN** and *HOTMAIL** (Hotmail alone account for almost half the supplied e-mail addresses) are all increasing the risk of default. *MSN** increase the probability of default with as much as 31% and the others, 29% and 23% are not far behind.

Table 10

Hypotheses: Moral hazard

#	Hypothesis	Variable(s)	Decision
H14	People that submit voluntary information are less likely to default	*SUBM_PHONE*	Accepted
H15	Probability of default should differ depending on type of store	*TYPE; AVERAGESALES2*	Accepted
H16	People that try to maximise their credit have a higher probability of default	*FAILEDBUYS*	Accepted
H17	Loan size increases probability of default	*SUM2*	Accepted
H18	People's email-addresses tell us something about the probability of default	*DOMAIN_NAME*	Accepted
H19	People ordering at awkward times of the day are more likely to default	*ORDERTIME*	Accepted

Moral hazard also seems to be reflected in the size of the store and the type of goods purchased. Some stores have indicated that a listing on Google Ad-words with the words “pay by invoice” increased sales some, but increased defaults even more.[55] This is an indication of the rent-seeking mentality that exists. It seems logical that bad debtors will expect smaller stores (*AVERAGESALES2*) to be less experienced with handling non-payers. Our regression showed that the marginal effect of moving one standard deviation from a store with 6 sales per day (a small store) to 20 sales per day decreases the risk by 10%, hence accepting the hypothesis. The type (*TYPE**) of goods one buys also have significant impact on the default ratio. All listed decrease the probability of default compared to the *OTHER** category. And all - *CARS**, *LEISURE**, *FITNESS**, *HOME** and *FASHION** - mainly carry goods that are often

[55] Niklas Adalberth, Kreditor Europe AB

individual in sizes, materials etc or in other ways have characteristics which make them particularly hard to sell on a second hand market.

A very reliable variable is failed purchase attempts (*FAILEDBUYS2)*. While failed purchase attempts could depend on a number of things such as the wrong submitted address or an exceeded credit limit, they all indicate that the applicant rather tries to obtain credit than expresses an interest in a specific product. Some failed attempts were of the type where the customer initially had tried to purchase for 6,000 SEK, then moved down to 5,000 SEK and so on and so forth until credit was granted, hence strongly suggesting an interest in credit rather than in the product.

Finally, the most statistically significant and important variable is what time purchase was made (*ORDERTIME*). Purchases made in the middle of the night often doubled default risks. One interpretation is that people that never have the intention of paying prefer engaging in this behaviour during the night when hidden away rather than during working hours. However, even lunch seems to provide such an opportunity. A significant increases in risk can be observed between 11 a.m. and 14 p.m.

6.3 Limitations

Below we will go through limitations we have encountered in our research; the problem with sample selection bias, the inability to evaluate the estimated models, lack of relevant information and the need for specialised models.

6.3.1 Sample selection bias

In Sweden, a registered payment remark is one of the strongest indicators of low credit worthiness. It is therefore customary, and in line with what in Swedish law is called "god kreditgivningssed" – good faith in lending – to deny credit to applicants with a registered payment remark. As described in the Data-section above this has also been the case with our source of data. However, this will give rise to the problem of sample selection bias in our data. The ideal data source would have included the full information on denied credit applications. Data that was not available to us. A way to improve the thesis would have been to find a data source which included this data. It would however give rise to yet another problem; reject inference. When one denies credit there is no absolute way to determine the outcome if it had been accepted. This is a general problem when one wants to evaluate default prediction models.

6.3.2 Evaluation of the model

It would have been interesting to evaluate our model. We have chosen not to, due to lack of new data to run the model on.

6.3.3 Lack of information on profitability

Lending small sums of money can be a profitable business even when loans are extended to debtors with low credit worthiness. Debtors with low credit worthiness are to a larger extent late with their payments and since this leads to fees that are high, relative to the assumed risk, the debtors can be very profitable. Developing a model that only focuses on risk, and not reward, might therefore lead to sub optimisation. This is a general problem in credit scoring models. One way to deal with this problem is to calculate the expected loss and set it in relation to expected income before credit losses. To maximise profits credits should be granted when expected income before credit losses exceeds expected credit losses. However, we did not have sufficient information to calculate the expected income for each applicant. Another factor that was hard to account for was the actual credit losses given default. Loss given default varies depending on a number of different factors including various debtor characteristics and total debt sum owed.

6.3.4 Different models for different applicants

We have developed a general model that is optimal for the average person in our sample. Hence, the model does not work so well for applicants that differs from the average. This is of course not the best solution. To successfully estimate credit risks we believe there is a need for models that are specialised to predict defaults in different sub groups. For example, high income generally implies that the probability of default is lowered. However, young people are generally paid less than older people. Does that mean that a low income is equally bad for credit worthiness for a person in her twenties than a person in her fifties? Our guess is that it is not. To build better credit risk models one need to identify sub groups in which the importance of determinants of default differs. One example of such a sub group might be students where low income in itself does not imply low credit worthiness. We have chosen to limit our work in this thesis to a generalised model since the main objective was not to develop an optimal model but rather to investigate variables of importance in credit granting decisions.

7 Conclusion

Our regressions have shown that the explanatory power of the data collected at the time of purchase is better at explaining defaults than purely demographical data. The explanatory power in the regression with behavioural parameters was 2.6 times as strong as the regression that only included demographic parameters. The *Pseudo* R^2 was 0.2135 compared to 0.0814 for the financial and demographic variables. Our findings are in line with the previous research on the area.

With the rise of new technologies, new ways of applying for and extending credit have been developed. This also has implications for the development of credit scoring methods. We have looked at some parameters relating to moral hazard that clearly can have a great value when predicting default. To improve credit scoring methods one needs to incorporate behavioural information indicating moral hazard and/or poor financial ability. It seems that as the application process for credits moves from the back-office of a bank directly to the internet this suddenly extends the behavioural information available. We have shown that this information can successfully be used to improve default predictions. Time of application is one of the most effective predictors of default. As the capacity and technology of today's databases and software expand, even details such as how you move the cursor over your bank's internet site, could prove to be valuable information. Hence, letting the applicant fill in their loan applications might have much more to it than saving time for customer service, and might in the future turn out to be essential for predicting default.

References

Books and periodicals

Altman, E. I., Avery, R. B., Eisenbeis, R. A. and Sinkey, J. F., *Application of classification techniques in business, banking and finance,* Connecticut: JAI Press Inc, 1981.

Altman, E. I. and Saunders, A., *An analysis and critique of the BIS Proposal on capital adequacy and ratings,* Journal of Banking and Finance, 25 (1): 25-46, 2001.

Bank for International Settlements, *Statistics on payment and settlement systems in selected countries - Figures for 2003,* CPSS Publications No 66, 2005.

Bank for International Settlements, *International convergence of capital measurement and capital standards – a revised framework,* Basel Committee on Banking Supervision, 2004.

Bierman, Jr, H. and Hausman, W.H., *The credit granting decision,* Management Science, 16, 519-532, 1970.

Brooks, C., *Introductory Econometrics for Finance*, Cambridge: Cambridge University Press, 2007.

Boyle, M., Crook,. J. N., Hamilton, R. and Thomas, L. C., *Methods for credit scoring applied to slow payers.* Oxford: Clarendon, 1992.

Breiman, L., Friedman, J.H., Olshen, R. A., and Stone, C. J., *Classification and Regression Trees,* Belmont: Wadsworth, 1984.

Brunner, A., Krahnen, J. P. and Weber, M., *Information Production in Credit Relationships: On the Role of Internal Ratings in Commercial Banking*, Working Paper No 2000/10 Center for Financial Studies, 2000.

Chambers E. A. and Cox, D. R., *Discrimination between alternative binary response models,* Biometrika 54, 573-578, 1967.

Charness, G. and Gneezy, U., *Strong Evidence for Gender Differences in Investment,* Working Paper available at http://ssrn.com/abstract=648735, September 18, 2007.

Chatterjee, S. and Barcun, S., *A nonparametric approach to credit scoring to credit screening*, Journal of the American Statistical Association., 65, 150-154, 1970.

Davis, D. B., *Artificial intelligence goes to work*, High Technology, Apr., 16-17, 1987.

Davis., R. H., Edelman, D. B. and Gammerman, A. J., *Machine-learning algorithms for credit card applications*. IMA Journal of Math. Appl. Bus. Industry, 4, 81-87, 1992.

Dirickx, Y. M. I. and Wakeman, L., *An extension of the Bierman-Hausman model for credit granting*, Management Science, 22, 1229-1237, 1976.

Durand, D., *Risk Elements in Consumer Instalment Financing.* New York: National Bureau of Economic Research, 1941.

Economist Intelligence Unit, *European Marketing Data and Statistics*, Euromonitor International, 2007.

Eisenbeis, R. A., *Pitfalls in the application of discriminant analysis in business, finance, and economics,* Journal of Finance, 32, 875-900, 1977.

Eisenbeis, R. A., *Problems in applying discriminant analysis in credit scoring models,* Journal of Banking and Finance. 2, 205-219, 1978.

Fehr-Duda, H., Gennaro, M. and Schubert, R., *Gender, Financial Risk, and Probability Weights,* Theory and Decision, Vol. 60, Nr 2-3, pp. 283-313 (31), Springer, 2006.

Felson, M. and Poulsen, E., *Simple indicators of crime by time of day,* International Journal of Forecasting 19, 595–601, 2003.

Fitzpatrick, D. B., *An analysis of bank credit card profit,* Journal of Bank Research, 7, 199-205, 1976.

Hand, D. J., *Discrimination and classification.* Chichester: Wiley, 1981.

Henley, W. E. and Hand, D. J., *Statistical Classification Methods in Consumer Credit Scoring: a Review*, Journal of the Royal Statistical Society A, 1997.

Henley, W. E. and Hand, D. J., *A k-nearest neighbourhood classifier for assessing consumer credit risk*, Statistician, 45, 77-95, 1996.

Henley, W. E., *Statistical aspects of credit scoring.* PhD Thesis. The Open University, Milton Keynes, 1995.

Kilborn, J. J., *Behavioral Economics, Overindebtedness & Comparative Consumer Bankruptcy: Searching for Causes and Evaluating Solutions,* Bankruptcy Developments Journal, Vol. 22, p. 13, 2005.

Kolesar, P. and Showers, J. L., *A robust credit screening model using categorical data*, Management Science, 31, 123-133, 1985.

Leonard, K. J., *Empricial Bayes analysis of the commercial loan evaluation process*, Statist. Probab. Lett. 18, 289-296, 1993.

Lucas, A., *Updating scorecards; removing the mystique*, Credit Scoring and Credit Control, pp 180-197. Oxford: Clarendon, 1992.

Mehta, D., *The formulation of credit policy models.* Management Science, 15, 30-50, 1968.

Mordecai, A., *Nonlinear Programming: Analysis and Methods,* Dover Publishing, 2003.

Orgler, Y. E., *Evaluation of bank consumer loans with credit scoring models.* J. Bank. Res. spring, 31-37, 1971.

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

Roszbach, K., *Bank Lending Policy, Credit Scoring and the Survival of Loans*, Sveriges Riksbank Working Paper Series, 2003.

Showers, J. L. and Chakrin, L. M., *Reducing uncollectible revenue from residential telephone customers*, Interfaces, 11, 21-31, 1981.

Srinivasan, V. and Kim, Y. H., *Credit granting: a comparative analysis of classification procedures,* Journal of Finance, 42, 665-683, 1987.

Vellido, A., Lisboa, P. J. G. and Vaughan, J., *Neural networks in business: a survey of applications (1992-1998)*, Expert Systems with Applications, Volume 17, Number 1, July 1999, pp. 51-70(20), 1999.

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

Zocco, D. P., *A framework for expert systems in bank loan management*. Journal of Commercial Bank Lending, 67, 47-54, 1985.

Interviews

Niklas Adalberth, Chief Operating Officer, Kreditor Europe AB

Internet

Business Check
http://www.businesscheck.se

Creditsafe
http://www.creditsafe.se

CNN
http://money.cnn.com/2007/10/15/markets/junk_mortgages.fortune/index.htm?postversion=2007101609, 2007-10-25

Department of Commerce; International Trade Administration
http://www.ita.doc.gov/td/finance/publications/creditcards.pdf, 2007-09-25

Dun & Bradstreet
http://www.dnb.com

Equifax
http://www.equifax.com

Experian
http://www.experian.com

Swedish Enforcement Authority
http://www.kronofogden.se

Statistics Sweden
http://www.scb.se, 2007-10-01

Swedish Tax Authorities
http://www.skatteverket.se/funktioner/svarpavanligafragor/privatovrigt/privatformogenhetsskattfaq/20050415vilkareglergallerforformogenhetsbeskattning.5.18e1b10334ebe8bc8000119186.html, 2007-11-20

Trans Union
http://www.transunion.com

Upplysningscentralen
http://www.uc.se

Databases

Proprietary database from a Swedish factoring company

Proprietary database from a Swedish credit reporting agency

Appendix A: Variables

A1 Description of variables

Appendix A1-1: Working dataset

Variable	Description
DEFAULT	customer defaulted on debt, dependent variable
FAILEDBUYS	number of failed purchase attempts
TYPE	Shop Classification, see Appendix A1:2 TYPE
AVERAGESALES2	logarithm of store size measure
SUM2	logarithm of sum of credit
TIMELASTCHECK	date when last credit check was made
PREVIOUSUNPAID2	logarithm of number of previously unpaid credits
PREVIOUSPAID2	logarithm of number of previously paid credits
PREVIOUSUNPAIDR2	logarithm of number of unpaid credits in reminder state
PREVIOUSPAIDR2	logarithm of number of paid credits in reminder state
PREVIOUSUNPAIDD2	logarithm of number of unpaid credits in recovery state
PREVIOUSPAIDD2	logarithm of number of paid credits in recovery state
ORDERTIME	time of day credit was approved, see Appendix A1:3 ORDERTIME
DEFICIT_CAPITAL2	debtor's deficit of property, 0 if no deficit
TAXED_PROPERTY2	logarithm of debtor's taxed property exceeding 1 500 000 kr, 0 if no property
COUNTRYMAIL	dummy variable, rural = 1, all other = 0
MAILLAN	debtor's region (derived from zipcode), see Appendix A1:4 MAILLAN
GENDER	sex, male = 1; female = 0
LANCODE	region where debtor was born, see Appendix A1:5 LANCODE
SUBM_PHONE	phone number was submitted, yes = 1; no = 0
MARRIED	debtor is married, yes = 1; no = 0
DOMAIN_NAME	known e-mail providers, see Appendix A1:6 DOMAIN_NAME
CO	debtor has a care of-address, yes = 1; no = 0
MAILADDRESS	debtor does not live at registered address, yes = 1; no = 0
INHABITANTS2	logarithm of size of debtors city
INCOME2	logarithm of debtor's total income
AGE	debtor's age
AGE2	logarithm of debtor's age
MONTH	debtor's month of birth, see Appendix A1:7 Month

Appendix A1-2: TYPE: Shop Classification

Variable	Description
*T BABY**	Baby Clothes and Accessories
*T CARS**	Car & Motorcycle Accessories
*T CHILDREN**	Toys and Gadgets for Kids
*T COMPUTERS**	Computer and Computer Accessories
*T COSMETICS**	Cosmetics and Beauty Products
*T ENTERTAINMENT**	Entertainment
*T EROTIC**	Erotic
*T FASHION**	Fashion
*T FITNESS**	Exercise and Fitness Accessories and Services
*T GADGETS**	Gadgets
*T GIFTS**	Gifts
*T HEALTH**	Health
*T HOME**	Furniture, home interior etc
*T LEISURE**	Leisure
*T PETS**	Pet Food and Accessories
*T OTHER**	Other - Base case

Appendix A1-3: ORDERTIME: Time of day when purchase was made

Variable	Description
*00 - 01**	Purchase made between 12 a.m. and 1 a.m.
*01 - 02**	Purchase made between 1 a.m. and 2 a.m.
*02 - 03**	Purchase made between 2 a.m. and 3 a.m.
*03 - 04**	Purchase made between 3 a.m. and 4 a.m.
*04 - 05**	Purchase made between 4 a.m. and 5 a.m.
*05 - 06**	Purchase made between 5 a.m. and 6 a.m.
*06 - 07**	Purchase made between 6 a.m. and 7 a.m.
*07 - 08**	Purchase made between 7 a.m. and 8 a.m.
*08 - 09**	Purchase made between 8 a.m. and 9 a.m.
*09 - 10**	Purchase made between 9 a.m. and 10 a.m.
*10 - 11**	Purchase made between 10 a.m. and 11 a.m.
*11 - 12**	Purchase made between 11 a.m. and 12 p.m.
*12 - 13**	Purchase made between 12 p.m. and 1 p.m.
*13 - 14**	Purchase made between 1 p.m. and 2 p.m.
*14 - 15**	Purchase made between 2 p.m. and 3 p.m.
*15 - 16**	Purchase made between 3 p.m. and 4 p.m.
*16 - 17**	Purchase made between 4 p.m. and 5 p.m.
*17 - 18**	Purchase made between 5 p.m. and 6 p.m.
*18 - 19**	Purchase made between 6 p.m. and 7 p.m.
*19 - 20**	Purchase made between 7 p.m. and 8 p.m.
*20 - 21**	Purchase made between 8 p.m. and 9 p.m.
*21 - 22**	Purchase made between 9 p.m. and 10 p.m. - Base case
*22 - 23**	Purchase made between 10 p.m. and 11 p.m.
*23 - 00**	Purchase made between 11 p.m. and 12 p.m.

Appendix A1-4: MAILLAN: Region where debtor lives

Variable	Comment
*BLEKINGE**	Blekinge län
*DALARNA**	Dalarnas län
*GOTLAND**	Gotlands län
*GÄVLEBORG**	Gävleborgs län
*HALLAND**	Hallands län
*JÄMTLAND**	Jämtlands län
*JÖNKÖPING**	Jönköpings län
*KALMAR**	Kalmar län
*KRONOBERG**	Kronobergs län
*NORRBOTTEN**	Norbottens län
*SKÅNE**	Skåne län
*STOCKHOLM**	Stockholms län - Base case
*SÖDERMANLAND**	Södermanlands län
*UPPSALA**	Uppsala län
*VÄRMLAND**	Värmlands län
*VÄSTERBOTTEN**	Västerbottens län
*VÄSTERNORRLAND**	Västernorrlands län
*VÄSTMANLAND**	Västmanlands län
*VÄSTRA GÖTALAND**	Västra Götalands län
*ÖREBRO**	Örebro län
*ÖSTERGÖTLAND**	Östergötlands län

Appendix A1-5: LANCODE: Region where debtor was born

Variable	Comment
*BLEKINGE LÄN**	Blekinge län
*GOTLAND LÄN**	Gotlands län
*GÄVLEBORG LÄN**	Gävleborgs län
*GÖTEBORG AND BOHUS LÄN**	Göteborg och Bohus län, now a part of Västra Götalands län
*HALLAND LÄN**	Hallands län
*IMMIGRANT/FOREIGNER**	Born abroad or foreign nationals born in Sweden
*IMMIGRANTAFTER1990**	Born abroad or foreign nationals born in Sweden
*IMMIGRANTORADOPTED**	Born abroad or adopted to Sweden before 1992
*JÄMTLAND LÄN**	Jämtlands län
*JÖNKÖPING LÄN**	Jönköpings län
*KALMAR LÄN**	Kalmar län
*KOPPARBERG LÄN**	Kopparbergs län, now Dalarnas län
*KRISTIANSTAD LÄN**	Kristianstads län, now a part of Skåne län
*KRONOBERG LÄN**	Kronobergs län
*MALMÖHUS LÄN**	Malmöhus län, now a part of Skåne län
*NORRBOTTEN LÄN**	Norbottens län
*SKARABORG LÄN**	Skaraborgs län, now a part of Västra Götaland
*STOCKHOLMS LÄN**	Stockholms län - Base case
*SÖDERMANLAND LÄN**	Södermanlands län
*UPPSALA LÄN**	Uppsala län
*VÄRMLAND LÄN**	Värmlands län
*VÄSTERBOTTEN LÄN**	Västerbottens län
*VÄSTERNORRLAND LÄN**	Västernorrlands län
*VÄSTMANLAND LÄN**	Västmanlands län
*ÄLVSBORG LÄN**	Älvsborgs län, now a part of Västra Götalands län
*ÖREBRO LÄN**	Örebro län
*ÖSTERGÖTLAND LÄN**	Östergötlands län

Appendix A1-6: DOMAIN_NAME: List of known e-mail providers

Variable	Comment
*BREDBANDSBOLAGET**	Broadband Infrastructure provider
*COMHEM**	Broadband Infrastructure provider
*GLOCALNET**	Broadband Infrastructure provider
*GMAIL**	E-mail provider
*HOME**	Broadband Infrastructure provider
*HOTMAIL**	E-mail provider
*MSN**	E-mail provider
*OTHER**	All other providers
*SPRAY**	E-mail provider
*STUDENT**	University associated e-mail address
*SWIPNET**	Broadband Infrastructure provider
*TELE2**	Broadband Infrastructure provider
*TELIA**	Broadband Infrastructure provider
*YAHOO**	E-mail provider
*NO SUBMITTED MAILADDRESS**	Base case since most debtors did not submit e-mail address

Appendix A1-7: Month: Debtor's month of birth

Variable	Comment
*JANUARY**	
*FEBRUARY**	
*MARCH**	Base case
*APRIL**	
*MAY**	
*JUNE**	
*JULY**	
*AUGUST**	
*SEPTEMBER**	
*OCTOBER**	
*NOVEMBER**	
*DECEMBER**	

A2 Regressions

Appendix A2-1a: Probit regression: Financial and demographic variables

Log likelihood	-7731,452	Number of obs	126518
		LR chi2(83)	1369,89
		Prob > chi2	0
		Pseudo R2	0,0814

Variable	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
DEFICITCAPITAL2	0,025038	0,003033	8,26	0,0000	0,01909	0,03098
TAXED_PROPERTY2	-0,037618	0,020281	-1,85	0,0640	-0,07737	0,00213
COUNTRYMAILL*	-0,118227	0,038112	-3,10	0,0020	-0,19292	-0,04353
BLEKINGE*	0,109778	0,100434	1,09	0,2740	-0,08707	0,30663
DALARNA*	0,043816	0,084861	0,52	0,6060	-0,12251	0,21014
GOTLAND*	0,047212	0,164883	0,29	0,7750	-0,27595	0,37038
GÄVLEBORG*	0,092322	0,080674	1,14	0,2520	-0,06580	0,25044
HALLAND*	0,009495	0,087273	0,11	0,9130	-0,16156	0,18055
JÄMTLAND*	0,086022	0,106526	0,81	0,4190	-0,12276	0,29481
JÖNKÖPING*	0,059879	0,075107	0,80	0,4250	-0,08733	0,20709
KALMAR*	0,072672	0,088289	0,82	0,4100	-0,10037	0,24572
KRONOBERG*	-0,110593	0,102867	-1,08	0,2820	-0,31221	0,09102
NORRBOTTEN*	-0,015927	0,084121	-0,19	0,8500	-0,18080	0,14895
SKÅNE*	0,179048	0,049706	3,60	0,0000	0,08163	0,27647
SÖDERMANLAND*	0,008420	0,077883	0,11	0,9140	-0,14423	0,16107
UPPSALA*	-0,085987	0,075817	-1,13	0,2570	-0,23458	0,06261
VÄRMLAND*	-0,001312	0,088988	-0,01	0,9880	-0,17573	0,17310
VÄSTERBOTTEN*	-0,319774	0,099253	-3,22	0,0010	-0,51431	-0,12524
VÄSTERNORRLAND*	0,042233	0,082901	0,51	0,6100	-0,12025	0,20472
VÄSTMANLAND*	0,153381	0,076663	2,00	0,0450	0,00312	0,30364
VÄSTRA GÖTALAND*	0,177432	0,046538	3,81	0,0000	0,08622	0,26865
ÖREBRO*	0,037174	0,078984	0,47	0,6380	-0,11763	0,19198
ÖSTERGÖTLAND*	-0,000313	0,070862	0,00	0,9960	-0,13920	0,13857
GENDER*	0,010819	0,020960	0,52	0,6060	-0,03026	0,05190
BLEKINGE LÄN*	-0,072647	0,101573	-0,72	0,4740	-0,27173	0,12643
GOTLAND LÄN*	0,040729	0,152031	0,27	0,7890	-0,25725	0,33870
GÄVLEBORG LÄN*	-0,009768	0,078825	-0,12	0,9010	-0,16426	0,14473
GÖTEBORG AND BOHUS LÄN*	-0,108840	0,055549	-1,96	0,0500	-0,21771	0,00003
HALLAND LÄN*	-0,063449	0,094757	-0,67	0,5030	-0,24917	0,12227
IMMIGRANT/FOREIGNER*	0,191281	0,071853	2,66	0,0080	0,05045	0,33211
IMMIGRANTAFTER1990*	0,013754	0,111824	0,12	0,9020	-0,20542	0,23292
IMMIGRANTORADOPTED*	0,185382	0,148883	1,25	0,2130	-0,10642	0,47719
JÄMTLAND LÄN*	0,016231	0,100165	0,16	0,8710	-0,18009	0,21255
JÖNKÖPING LÄN*	-0,079872	0,076682	-1,04	0,2980	-0,23016	0,07042
KALMAR LÄN*	-0,059311	0,083492	-0,71	0,4770	-0,22295	0,10433
KOPPARBERG LÄN*	-0,107598	0,083491	-1,29	0,1970	-0,27124	0,05604
KRISTIANSTAD LÄN*	-0,047321	0,067635	-0,70	0,4840	-0,17988	0,08524
KRONOBERG LÄN*	-0,001199	0,092430	-0,01	0,9900	-0,18236	0,17996
MALMÖHUS LÄN*	-0,067365	0,055578	-1,21	0,2250	-0,17630	0,04157
NORRBOTTEN LÄN*	-0,006720	0,076937	-0,09	0,9300	-0,15751	0,14407
SKARABORG LÄN*	-0,168965	0,074360	-2,27	0,0230	-0,31471	-0,02322
SÖDERMANLAND LÄN*	-0,033781	0,076671	-0,44	0,6600	-0,18405	0,11649
UPPSALA LÄN*	-0,009369	0,078672	-0,12	0,9050	-0,16356	0,14483
VÄRMLAND LÄN*	0,048141	0,085133	0,57	0,5720	-0,11872	0,21500
VÄSTERBOTTEN LÄN*	0,153096	0,085463	1,79	0,0730	-0,01441	0,32060
VÄSTERNORRLAND LÄN*	0,003179	0,081032	0,04	0,9690	-0,15564	0,16200

VÄSTMANLAND LÄN*	-0,098276	0,079872	-1,23	0,2190	-0,25482	0,05827
ÄLVSBORG LÄN*	-0,105932	0,063765	-1,66	0,0970	-0,23091	0,01905
ÖREBRO LÄN*	0,025009	0,078834	0,32	0,7510	-0,12950	0,17952
ÖSTERGÖTLAND LÄN*	-0,001282	0,068958	-0,02	0,9850	-0,13644	0,13387
SUBM_PHONE*	-0,124697	0,023111	-5,40	0,0000	-0,16999	-0,07940
MARRIED*	-0,045730	0,028263	-1,62	0,1060	-0,10112	0,00966
BREDBANDSBOLAGET*	-0,164900	0,096960	-1,70	0,0890	-0,35494	0,02514
COMHEM*	-0,044439	0,114743	-0,39	0,6990	-0,26933	0,18045
GLOCALNET*	0,016182	0,150162	0,11	0,9140	-0,27813	0,31049
GMAIL*	-0,319206	0,106251	-3,00	0,0030	-0,52745	-0,11096
HOME*	0,086438	0,117923	0,73	0,4640	-0,14469	0,31756
HOTMAIL*	0,138821	0,025675	5,41	0,0000	0,08850	0,18914
MSN*	0,197153	0,097082	2,03	0,0420	0,00688	0,38743
OTHER*	-0,095202	0,040906	-2,33	0,0200	-0,17538	-0,01503
SPRAY*	-0,082351	0,070088	-1,17	0,2400	-0,21972	0,05502
STUDENT*	-0,218768	0,191486	-1,14	0,2530	-0,59407	0,15654
SWIPNET*	-0,378023	0,234649	-1,61	0,1070	-0,83793	0,08188
TELE2*	-0,209373	0,199654	-1,05	0,2940	-0,60069	0,18194
TELIA*	-0,119194	0,057267	-2,08	0,0370	-0,23143	-0,00695
YAHOO*	0,156469	0,068175	2,30	0,0220	0,02285	0,29009
CO*	-0,163326	0,114435	-1,43	0,1540	-0,38762	0,06096
INHABITANTS2	-0,008141	0,009257	-0,88	0,3790	-0,02628	0,01000
INCOME2	-0,044438	0,002058	-21,60	0,0000	-0,04847	-0,04040
AGE	-0,048925	0,005621	-8,70	0,0000	-0,05994	-0,03791
AGE2	0,000467	0,000070	6,66	0,0000	0,00033	0,00060
JANUARY*	0,103370	0,050256	2,06	0,0400	0,00487	0,20187
FEBRUARY*	0,141609	0,050028	2,83	0,0050	0,04356	0,23966
APRIL*	0,023757	0,050885	0,47	0,6410	-0,07598	0,12349
MAY*	0,065445	0,049753	1,32	0,1880	-0,03207	0,16296
JUNE*	0,127253	0,049697	2,56	0,0100	0,02985	0,22466
JULY*	0,081945	0,050327	1,63	0,1030	-0,01669	0,18058
AUGUST*	0,103824	0,049939	2,08	0,0380	0,00594	0,20170
SEPTEMBER*	0,087865	0,050309	1,75	0,0810	-0,01074	0,18647
OCTOBER*	0,083243	0,051022	1,63	0,1030	-0,01676	0,18324
NOVEMBER*	0,031506	0,053542	0,59	0,5560	-0,07343	0,13645
DECEMBER*	0,096184	0,051778	1,86	0,0630	-0,00530	0,19767
_CONS	-0,884910	0,146624	-6,04	0,0000	-1,17229	-0,59753

Appendix A2-1b: Probit regression: Marginal effects on financial and demographic variables

Log likelihood	-77 310	Number of obs	126 518
Mean default	1,2346%	LR chi2(83)	1 369,89
		Prob> chi2	0,0000
		Pseudo R2	0,0814

Variable	dF/dx	Std. Err.	z	P>|z|	x-bar	dF/dX*std in % pts.	Change in default ratio
DEFICITCAPITAL2	0,00059	0,00007	8,26	0,0000	5,67735	0,26%	20,68%
TAXED_PROPERTY2	-0,00088	0,00047	-1,85	0,0640	0,14307	-0,12%	-9,83%
COUNTRYMAILL*	-0,00248	0,00071	-3,10	0,0020	0,10957	-0,25%	-20,10%
BLEKINGE*	0,00291	0,00301	1,09	0,2740	0,01864	0,29%	23,61%
DALARNA*	0,00108	0,00219	0,52	0,6060	0,03288	0,11%	8,72%
GOTLAND*	0,00117	0,00431	0,29	0,7750	0,00628	0,12%	9,46%
GÄVLEBORG*	0,00239	0,00231	1,14	0,2520	0,03373	0,24%	19,39%
HALLAND*	0,00022	0,00209	0,11	0,9130	0,02819	0,02%	1,82%
JÄMTLAND*	0,00222	0,00303	0,81	0,4190	0,01512	0,22%	18,01%
JÖNKÖPING*	0,00150	0,00200	0,80	0,4250	0,03487	0,15%	12,13%
KALMAR*	0,00185	0,00243	0,82	0,4100	0,02536	0,18%	14,95%
KRONOBERG*	-0,00228	0,00186	-1,08	0,2820	0,02022	-0,23%	-18,49%
NORRBOTTEN*	-0,00037	0,00190	-0,19	0,8500	0,03863	-0,04%	-2,97%
SKÅNE*	0,00494	0,00160	3,60	0,0000	0,12126	0,49%	40,01%
SÖDERMANLAND*	0,00020	0,00186	0,11	0,9140	0,03087	0,02%	1,61%
UPPSALA*	-0,00183	0,00146	-1,13	0,2570	0,03600	-0,18%	-14,82%
VÄRMLAND*	-0,00003	0,00208	-0,01	0,9880	0,03132	0,00%	-0,25%
VÄSTERBOTTEN*	-0,00530	0,00110	-3,22	0,0010	0,03268	-0,53%	-42,94%
VÄSTERNORRLAND*	0,00104	0,00213	0,51	0,6100	0,03294	0,10%	8,39%
VÄSTMANLAND*	0,00427	0,00250	2,00	0,0450	0,03104	0,43%	34,55%
VÄSTRA GÖTALAND*	0,00482	0,00145	3,81	0,0000	0,15531	0,48%	39,04%
ÖREBRO*	0,00091	0,00201	0,47	0,6380	0,03250	0,09%	7,34%
ÖSTERGÖTLAND*	-0,00001	0,00166	0,00	0,9960	0,04456	0,00%	-0,06%
GENDER*	0,00025	0,00049	0,52	0,6060	0,44157	0,03%	2,05%
BLEKINGE LÄN*	-0,00156	0,00201	-0,72	0,4740	0,01953	-0,16%	-12,67%
GOTLAND LÄN*	0,00100	0,00391	0,27	0,7890	0,00748	0,10%	8,09%
GÄVLEBORG LÄN*	-0,00023	0,00180	-0,12	0,9010	0,03645	-0,02%	-1,83%
GÖTEBORG AND BOHUS LÄN*	-0,00228	0,00104	-1,96	0,0500	0,07245	-0,23%	-18,48%
HALLAND LÄN*	-0,00138	0,00191	-0,67	0,5030	0,02093	-0,14%	-11,18%
IMMIGRANT/FOREIGNER*	0,00558	0,00256	2,66	0,0080	0,02097	0,56%	45,18%
IMMIGRANTAFTER1990*	0,00033	0,00270	0,12	0,9020	0,00560	0,03%	2,65%
IMMIGRANTORADOPTED*	0,00541	0,00532	1,25	0,2130	0,00205	0,54%	43,84%
JÄMTLAND LÄN*	0,00039	0,00243	0,16	0,8710	0,01703	0,04%	3,13%
JÖNKÖPING LÄN*	-0,00171	0,00150	-1,04	0,2980	0,03462	-0,17%	-13,86%
KALMAR LÄN*	-0,00130	0,00171	-0,71	0,4770	0,02986	-0,13%	-10,52%
KOPPARBERG LÄN*	-0,00224	0,00153	-1,29	0,1970	0,03464	-0,22%	-18,11%
KRISTIANSTAD LÄN*	-0,00105	0,00142	-0,70	0,4840	0,03280	-0,11%	-8,51%
KRONOBERG LÄN*	-0,00003	0,00216	-0,01	0,9900	0,02144	0,00%	-0,23%
MALMÖHUS LÄN*	-0,00147	0,00113	-1,21	0,2250	0,07558	-0,15%	-11,93%
NORRBOTTEN LÄN*	-0,00016	0,00177	-0,09	0,9300	0,04667	-0,02%	-1,26%
SKARABORG LÄN*	-0,00328	0,00118	-2,27	0,0230	0,02973	-0,33%	-26,57%
SÖDERMANLAND LÄN*	-0,00076	0,00166	-0,44	0,6600	0,03063	-0,08%	-6,16%
UPPSALA LÄN*	-0,00022	0,00180	-0,12	0,9050	0,02806	-0,02%	-1,76%
VÄRMLAND LÄN*	0,00119	0,00222	0,57	0,5720	0,03145	0,12%	9,63%
VÄSTERBOTTEN LÄN*	0,00425	0,00279	1,79	0,0730	0,03253	0,43%	34,45%
VÄSTERNORRLAND LÄN*	0,00007	0,00191	0,04	0,9690	0,03480	0,01%	0,60%
VÄSTMANLAND LÄN*	-0,00206	0,00150	-1,23	0,2190	0,03225	-0,21%	-16,70%

ÄLVSBORG LÄN*	-0,00221	0,00118	-1,66	0,0970	0,04302	-0,22%	-17,90%
ÖREBRO LÄN*	0,00060	0,00195	0,32	0,7510	0,03125	0,06%	4,87%
ÖSTERGÖTLAND LÄN*	-0,00003	0,00161	-0,02	0,9850	0,04590	0,00%	-0,24%
SUBM_PHONE*	-0,00303	0,00058	-5,40	0,0000	0,61715	-0,30%	-24,53%
MARRIED*	-0,00105	0,00063	-1,62	0,1060	0,30304	-0,10%	-8,48%
BREDBANDSBOLAGET*	-0,00321	0,00154	-1,70	0,0890	0,02201	-0,32%	-25,96%
COMHEM*	-0,00099	0,00242	-0,39	0,6990	0,01205	-0,10%	-8,00%
GLOCALNET*	0,00039	0,00365	0,11	0,9140	0,00653	0,04%	3,12%
GMAIL*	-0,00523	0,00115	-3,00	0,0030	0,01740	-0,52%	-42,36%
HOME*	0,00224	0,00336	0,73	0,4640	0,00865	0,22%	18,13%
HOTMAIL*	0,00351	0,00070	5,41	0,0000	0,26990	0,35%	28,47%
MSN*	0,00582	0,00354	2,03	0,0420	0,00784	0,58%	47,16%
OTHER*	-0,00205	0,00081	-2,33	0,0200	0,13960	-0,21%	-16,64%
SPRAY*	-0,00176	0,00136	-1,17	0,2400	0,03280	-0,18%	-14,24%
STUDENT*	-0,00397	0,00261	-1,14	0,2530	0,00413	-0,40%	-32,15%
SWIPNET*	-0,00576	0,00211	-1,61	0,1070	0,00712	-0,58%	-46,67%
TELE2*	-0,00384	0,00279	-1,05	0,2940	0,00616	-0,38%	-31,13%
TELIA*	-0,00247	0,00104	-2,08	0,0370	0,07158	-0,25%	-20,03%
YAHOO*	0,00438	0,00225	2,30	0,0220	0,02066	0,44%	35,50%
CO*	-0,00316	0,00181	-1,43	0,1540	0,00950	-0,32%	-25,63%
INHABITANTS2	-0,00019	0,00022	-0,88	0,3790	10,9261	-0,02%	-1,98%
INCOME2	-0,00104	0,00005	-21,60	0,0000	11,1572	-0,35%	-28,48%
AGE	-0,00114	0,00013	-8,70	0,0000	34,2730	-1,37%	-111,1%
AGE2	0,00001	0,00000	6,66	0,0000	1318,20	1,03%	83,54%
JANUARY*	0,00268	0,00144	2,06	0,0400	0,08171	0,27%	21,73%
FEBRUARY*	0,00382	0,00155	2,83	0,0050	0,07951	0,38%	30,97%
APRIL*	0,00057	0,00125	0,47	0,6410	0,09395	0,06%	4,61%
MAY*	0,00163	0,00132	1,32	0,1880	0,09154	0,16%	13,22%
JUNE*	0,00338	0,00149	2,56	0,0100	0,08372	0,34%	27,39%
JULY*	0,00208	0,00138	1,63	0,1030	0,08285	0,21%	16,85%
AUGUST*	0,00270	0,00143	2,08	0,0380	0,08238	0,27%	21,83%
SEPTEMBER*	0,00224	0,00140	1,75	0,0810	0,08176	0,22%	18,18%
OCTOBER*	0,00212	0,00141	1,63	0,1030	0,08008	0,21%	17,15%
NOVEMBER*	0,00076	0,00133	0,59	0,5560	0,07374	0,08%	6,16%
DECEMBER*	0,00248	0,00147	1,86	0,0630	0,07500	0,25%	20,10%
obs. P |	0,0123461						
pred. P |	0,0086141	(at x-bar)					

(*) dF/dx is for discrete change of dummy variable from 0 to 1
z and P>|z| correspond to the test of the underlying coefficient being 0

Appendix A2-1c: Adjusted and non-adjusted marginal effects on significant non-dummy financial and demographic variables

	Change from	To	Standard deviation	Number of std.	dF/dX*std in % pts.	Change in default ratio
DEFICITCAPITAL2	303,75	23819	4,359	1	0,26%	20,68%
TAXED_PROPERTY2	0,14	3,53	1,379	1	-0,12%	-9,83%
INHABITANTS2	55606	198494	1,284	1	-0,02%	-1,98%
INCOME2	70068	2064423	3,383	1	-0,35%	-28,48%
AGE	34,3	46,3	11,982	1	-0,34%	-27,52%

Comments: AGE is calculated from AGE and AGE2

Appendix A2-2a: Probit regression: Behavioral variables

		Number of obs =	169 453
		Wald chi2(53) =	2 078,78
		Prob > chi2 =	0,0000
Log pseudolikelihood =	-9 524,11	Pseudo R2 =	0,2135

(Std.Err. adjusted for 126 518 clusters in ID)

Variables	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
AGE	-0,0513631	0,0057106	-8,99	0,0000	-0,0626	-0,0402
AGE2	0,0004726	0,0000719	6,58	0,0000	0,0003	0,0006
INCOME2	-0,0462135	0,0022294	-20,73	0,0000	-0,0506	-0,0418
GENDER*	0,0123066	0,0287325	0,43	0,6680	-0,0440	0,0686
FAILEDBUYS	0,0394959	0,0050671	7,79	0,0000	0,0296	0,0494
T BABY*	-0,0462520	0,0520848	-0,89	0,3750	-0,1483	0,0558
T CARS*	-0,3606470	0,1375739	-2,62	0,0090	-0,6303	-0,0910
T CHILDREN*	0,0831357	0,0719773	1,16	0,2480	-0,0579	0,2242
T COMPUTERS*	-0,1105478	0,0594387	-1,86	0,0630	-0,2270	0,0059
T COSMETICS*	0,0503777	0,0470194	1,07	0,2840	-0,0418	0,1425
T ENTERTAINMENT*	0,0341291	0,0703387	0,49	0,6280	-0,1037	0,1720
T EROTIC*	0,0178628	0,0550048	0,32	0,7450	-0,0899	0,1257
T FASHION*	-0,0703644	0,0420762	-1,67	0,0940	-0,1528	0,0121
T FITNESS*	-0,2185962	0,0599855	-3,64	0,0000	-0,3362	-0,1010
T GADGETS*	0,1581342	0,0621960	2,54	0,0110	0,0362	0,2800
T GIFTS*	0,0934830	0,1104385	0,85	0,3970	-0,1230	0,3099
T HEALTH*	-0,0489645	0,0510115	-0,96	0,3370	-0,1489	0,0510
T HOME*	-0,1922514	0,0519990	-3,70	0,0000	-0,2942	-0,0903
T LEISURE*	-0,3263539	0,0928766	-3,51	0,0000	-0,5084	-0,1443
T PETS*	0,0437530	0,0800037	0,55	0,5840	-0,1131	0,2006
AVERAGESALES2	-0,0882474	0,0136983	-6,44	0,0000	-0,1151	-0,0614
SUM2	0,2326930	0,0137660	16,90	0,0000	0,2057	0,2597
TIMELASTCREDITCHECK	0,0020272	0,0002311	8,77	0,0000	0,0016	0,0025
PREVIOUSUNPAID2	0,8778525	0,0455992	19,25	0,0000	0,7885	0,9672
PREVIOUSPAID2	-0,7764174	0,0688955	-11,27	0,0000	-0,9115	-0,6414
PREVIOUSUNPAIDR2	0,2109733	0,1442328	1,46	0,1440	-0,0717	0,4937
PREVIOUSPAIDR2	0,1688685	0,1657654	1,02	0,3080	-0,1560	0,4938
PREVIOUSUNPAIDD2	1,8748700	0,3322728	5,64	0,0000	1,2236	2,5261
PREVIOUSPAIDD2	1,0090440	0,1845094	5,47	0,0000	0,6474	1,3707
00 - 01*	0,2631337	0,0623066	4,22	0,0000	0,1410	0,3853
01 - 02*	0,2971992	0,0761071	3,91	0,0000	0,1480	0,4464
02 - 03*	0,5525452	0,0850308	6,50	0,0000	0,3859	0,7192
03 - 04*	0,4855490	0,1166035	4,16	0,0000	0,2570	0,7141
04 - 05*	0,4035998	0,1486839	2,71	0,0070	0,1122	0,6950
05 - 06*	0,4506229	0,1466879	3,07	0,0020	0,1631	0,7381
06 - 07*	0,5136729	0,1072454	4,79	0,0000	0,3035	0,7239
07 - 08*	0,1987248	0,0960208	2,07	0,0380	0,0105	0,3869
08 - 09*	-0,0136962	0,0833299	-0,16	0,8690	-0,1770	0,1496
09 - 10*	-0,0031066	0,0675061	-0,05	0,9630	-0,1354	0,1292
10 - 11*	0,0532065	0,0608673	0,87	0,3820	-0,0661	0,1725
11 - 12*	0,1150486	0,0576182	2,00	0,0460	0,0021	0,2280
12 - 13*	0,1241956	0,0557435	2,23	0,0260	0,0149	0,2335
13 - 14*	0,1163351	0,0558039	2,08	0,0370	0,0070	0,2257
14 - 15*	0,0680069	0,0575206	1,18	0,2370	-0,0447	0,1807
15 - 16*	0,0895505	0,0569580	1,57	0,1160	-0,0221	0,2012
16 - 17*	0,0332644	0,0585951	0,57	0,5700	-0,0816	0,1481

17 - 18*	-0,0050432	0,0579625	-0,09	0,9310	-0,1186	0,1086
18 - 19*	0,1011165	0,0548507	1,84	0,0650	-0,0064	0,2086
19 - 20*	0,0014837	0,0561765	0,03	0,9790	-0,1086	0,1116
20 - 21*	0,0391977	0,0542709	0,72	0,4700	-0,0672	0,1456
22 - 23*	0,0814740	0,0523647	1,56	0,1200	-0,0212	0,1841
23 - 00*	0,1327250	0,0572012	2,32	0,0200	0,0206	0,2448
_CONS	-2,8978710	0,1017939	-28,47	0,0000	-3,0974	-2,6984

Note: 1 failure and 0 successes completely determined .

Appendix A2-2b: Probit regression: Marginal effects on behavioral and demographic variables

Log pseudolikelihood =	-9 524,11	Number of obs	169 453
Mean default	1,3485%	Wald chi2(53)	2 078,78
		Prob > chi2	0,0000
		Pseudo R2	0,2135

(standard errors adjusted for clustering on ID)

Variable	dF/dx	Robust Std. Err.	z	P>|z|	x-bar	dF/dX*std in % pts.	Change in default ratio
AGE	-0,000809	0,000089	-8,99	0,0000	33,738	-0,95%	-70,37%
AGE2	0,000007	0,000001	6,58	0,0000	1275,8	0,68%	50,65%
INCOME2	-0,000742	0,000044	-20,73	0,0000	11,0915	-0,26%	-19,20%
GENDER*	0,000198	0,000464	0,43	0,6680	0,4262	0,02%	1,47%
FAILEDBUYS	0,000634	0,000084	7,79	0,0000	0,2907	0,15%	11,33%
T BABY*	-0,000708	0,000758	-0,89	0,3750	0,0852	-0,07%	-5,25%
T CARS*	-0,003767	0,000862	-2,62	0,0090	0,0096	-0,38%	-27,94%
T CHILDREN*	0,001479	0,001412	1,16	0,2480	0,0182	0,15%	10,97%
T COMPUTERS*	-0,001599	0,000772	-1,86	0,0630	0,1208	-0,16%	-11,86%
T COSMETICS*	0,000855	0,000842	1,07	0,2840	0,0701	0,09%	6,34%
T ENTERTAINMENT*	0,000571	0,001226	0,49	0,6280	0,0229	0,06%	4,24%
T EROTIC*	0,000293	0,000920	0,32	0,7450	0,0492	0,03%	2,17%
T FASHION*	-0,001061	0,000594	-1,67	0,0940	0,1422	-0,11%	-7,87%
T FITNESS*	-0,002764	0,000589	-3,64	0,0000	0,0564	-0,28%	-20,50%
T GADGETS*	0,003065	0,001435	2,54	0,0110	0,0368	0,31%	22,73%
T GIFTS*	0,001688	0,002232	0,85	0,3970	0,0073	0,17%	12,52%
T HEALTH*	-0,000745	0,000736	-0,96	0,3370	0,0617	-0,07%	-5,52%
T HOME*	-0,002536	0,000563	-3,70	0,0000	0,0846	-0,25%	-18,80%
T LEISURE*	-0,003560	0,000655	-3,51	0,0000	0,0146	-0,36%	-26,40%
T PETS*	0,000742	0,001429	0,55	0,5840	0,0151	0,07%	5,50%
AVERAGESALES2	-0,001417	0,000223	-6,44	0,0000	1,9534	-0,16%	-11,56%
SUM2	0,003736	0,000225	16,90	0,0000	5,8194	0,34%	25,56%
TIMELASTCREDITCHECK	0,000033	0,000004	8,77	0,0000	8,1153	0,12%	8,74%
PREVIOUSUNPAID2	0,014095	0,000900	19,25	0,0000	0,0673	0,34%	25,10%
PREVIOUSPAID2	-0,012467	0,000974	-11,27	0,0000	0,2019	-0,55%	-41,01%
PREVIOUSUNPAIDR2	0,003388	0,002311	1,46	0,1440	0,0027	0,02%	1,14%
PREVIOUSPAIDR2	0,002711	0,002645	1,02	0,3080	0,0375	0,05%	3,69%
PREVIOUSUNPAIDD2	0,030104	0,005497	5,64	0,0000	0,0002	0,03%	2,55%
PREVIOUSPAIDD2	0,016202	0,003000	5,47	0,0000	0,0143	0,18%	13,20%
00 - 01*	0,005833	0,001837	4,22	0,0000	0,0254	0,58%	43,26%
01 - 02*	0,006936	0,002446	3,91	0,0000	0,0125	0,69%	51,44%
02 - 03*	0,017965	0,004694	6,50	0,0000	0,0061	1,80%	133,23%
03 - 04*	0,014531	0,005680	4,16	0,0000	0,0038	1,45%	107,76%
04 - 05*	0,010886	0,006100	2,71	0,0070	0,0024	1,09%	80,73%
05 - 06*	0,012915	0,006649	3,07	0,0020	0,0023	1,29%	95,78%
06 - 07*	0,015923	0,005495	4,79	0,0000	0,0046	1,59%	118,08%
07 - 08*	0,004096	0,002477	2,07	0,0380	0,0112	0,41%	30,37%
08 - 09*	-0,000216	0,001295	-0,16	0,8690	0,0239	-0,02%	-1,60%
09 - 10*	-0,000050	0,001076	-0,05	0,9630	0,0391	0,00%	-0,37%
10 - 11*	0,000908	0,001101	0,87	0,3820	0,0514	0,09%	6,73%
11 - 12*	0,002105	0,001192	2,00	0,0460	0,0572	0,21%	15,61%
12 - 13*	0,002298	0,001177	2,23	0,0260	0,0555	0,23%	17,04%
13 - 14*	0,002129	0,001155	2,08	0,0370	0,0615	0,21%	15,79%
14 - 15*	0,001178	0,001070	1,18	0,2370	0,0638	0,12%	8,74%
15 - 16*	0,001589	0,001110	1,57	0,1160	0,0645	0,16%	11,78%
16 - 17*	0,000554	0,001013	0,57	0,5700	0,0612	0,06%	4,11%

17 - 18*	-0,000081	0,000920	-0,09	0,9310	0,0585	-0,01%	-0,60%
18 - 19*	0,001820	0,001098	1,84	0,0650	0,0602	0,18%	13,49%
19 - 20*	0,000024	0,000905	0,03	0,9790	0,0674	0,00%	0,18%
20 - 21*	0,000657	0,000948	0,72	0,4700	0,0733	0,07%	4,87%
22 - 23*	0,001432	0,001005	1,56	0,1200	0,0670	0,14%	10,62%
23 - 00*	0,002482	0,001234	2,32	0,0200	0,0526	0,25%	18,41%
obs. P |	0,013485						
pred. P |	0,005625	(at x-bar)					

(*) dF/dx is for discrete change of dummy variable from 0 to 1
z and P>|z| correspond to the test of the underlying coefficient being 0

Appendix A2-2c: Adjusted and non-adjusted marginal effects on significant non-dummy behavioral and demographic variables

Variable	Change from	To	Standard deviation	Number of std.	dF/dX*std in % pts.	Change in default ratio
AGE	33,7	45,5	2,408	1	-0,27%	-19,72%
INCOME2	65 609	2 150 843	1,100	1	-0,26%	-19,20%
FAILEDBUYS	0,03	1,98	0,922	0,81	0,12%	9,17%
AVERAGESALES2	6,05	20,18	36,269	1	-0,16%	-11,56%
SUM2	335,76	846,09	0,240	1	0,34%	25,56%
TIMELASTCREDITCHECK	8,1	44,4	0,444	1	0,12%	8,74%
PREVIOUSUNPAID2	0,07	1,10	0,045	2,8	0,95%	70,29%
PREVIOUSPAID2	0,02	1,08	0,183	1,6	-0,88%	-65,62%
PREVIOUSUNPAIDR2	0,00	0,05	0,011	1	0,02%	1,14%
PREVIOUSPAIDR2	0,04	0,25	0,110	1	0,05%	3,69%
PREVIOUSUNPAIDD2	0,00	0,99	4,359	60	2,06%	153,08%
PREVIOUSPAIDD2	0,01	1,00	1,379	6,2	1,10%	81,82%

Comments: AGE is calculated from AGE and AGE2
PREVIOUSPAID2; the from has been decreased by 0,4 std while the to has been increased by 1,2 std
FAILEDBUYS; the from has been decrease by 0,11 std while the to has been increased by 0,7 std

Appendix A2-3a: Probit regression: All variables

		Number of obs =	169453
		Wald chi2(133) =	2421,1
		Prob > chi2 =	0
Log pseudolikelihood =	-9262,7	Pseudo R2 =	0,2351

(Std. Err. adjusted for 126 518 clusters in ID)

Variable	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
FAILEDBUYS	0,038002	0,004941	7,69	0,0000	0,02832	0,04769
T BABY*	-0,024587	0,052231	-0,47	0,6380	-0,12696	0,07778
T CARS*	-0,379638	0,139193	-2,73	0,0060	-0,65245	-0,10683
T CHILDREN*	0,139432	0,073531	1,90	0,0580	-0,00469	0,28355
T COMPUTERS*	-0,015777	0,062770	-0,25	0,8020	-0,13880	0,10725
T COSMETICS*	0,019710	0,047345	0,42	0,6770	-0,07309	0,11250
T ENTERTAINMENT*	0,076143	0,071917	1,06	0,2900	-0,06481	0,21710
T EROTIC*	0,034958	0,055822	0,63	0,5310	-0,07445	0,14437
T FASHION*	-0,088190	0,042330	-2,08	0,0370	-0,17115	-0,00523
T FITNESS*	-0,220184	0,061212	-3,60	0,0000	-0,34016	-0,10021
T GADGETS*	0,115771	0,061325	1,89	0,0590	-0,00442	0,23597
T GIFTS*	0,076410	0,107728	0,71	0,4780	-0,13473	0,28755
T HEALTH*	-0,041630	0,051399	-0,81	0,4180	-0,14237	0,05911
T HOME*	-0,163266	0,053491	-3,05	0,0020	-0,26811	-0,05843
T LEISURE*	-0,296351	0,095539	-3,10	0,0020	-0,48360	-0,10910
T PETS*	0,056850	0,081353	0,70	0,4850	-0,10260	0,21630
AVERAGESALES2	-0,095050	0,013656	-6,96	0,0000	-0,12182	-0,06828
SUM2	0,250083	0,013753	18,18	0,0000	0,22313	0,27704
TIMELASTCREDITCHECK	0,002079	0,000231	9,01	0,0000	0,00163	0,00253
PREVIOUSUNPAID2	0,884746	0,042839	20,65	0,0000	0,80078	0,96871
PREVIOUSPAID2	-0,774785	0,067443	-11,49	0,0000	-0,90697	-0,64260
PREVIOUSUNPAIDR2	0,179725	0,142584	1,26	0,2070	-0,09973	0,45918
PREVIOUSPAIDR2	0,190220	0,161159	1,18	0,2380	-0,12565	0,50609
PREVIOUSUNPAIDD2	1,841502	0,320391	5,75	0,0000	1,21355	2,46946
PREVIOUSPAIDD2	0,953274	0,181592	5,25	0,0000	0,59736	1,30919
00 - 01*	0,254951	0,063059	4,04	0,0000	0,13136	0,37854
01 - 02*	0,281378	0,076624	3,67	0,0000	0,13120	0,43156
02 - 03*	0,528471	0,087996	6,01	0,0000	0,35600	0,70094
03 - 04*	0,461992	0,117049	3,95	0,0000	0,23258	0,69140
04 - 05*	0,382874	0,150473	2,54	0,0110	0,08795	0,67780
05 - 06*	0,460474	0,147213	3,13	0,0020	0,17194	0,74901
06 - 07*	0,509989	0,107707	4,73	0,0000	0,29889	0,72109
07 - 08*	0,205976	0,096353	2,14	0,0330	0,01713	0,39482
08 - 09*	-0,000205	0,084322	0,00	0,9980	-0,16547	0,16506
09 - 10*	-0,000779	0,068736	-0,01	0,9910	-0,13550	0,13394
10 - 11*	0,058325	0,061558	0,95	0,3430	-0,06233	0,17898
11 - 12*	0,123238	0,057634	2,14	0,0320	0,01028	0,23620
12 - 13*	0,130981	0,056356	2,32	0,0200	0,02053	0,24144
13 - 14*	0,122147	0,056034	2,18	0,0290	0,01232	0,23197
14 - 15*	0,063773	0,058064	1,10	0,2720	-0,05003	0,17758
15 - 16*	0,087004	0,057039	1,53	0,1270	-0,02479	0,19880
16 - 17*	0,023797	0,058447	0,41	0,6840	-0,09076	0,13835
17 - 18*	-0,021086	0,058073	-0,36	0,7170	-0,13491	0,09273
18 - 19*	0,090950	0,055583	1,64	0,1020	-0,01799	0,19989
19 - 20*	0,005190	0,056436	0,09	0,9270	-0,10542	0,11580
20 - 21*	0,031231	0,054732	0,57	0,5680	-0,07604	0,13850

22 - 23*	0,074609	0,053064	1,41	0,1600	-0,02939	0,17861
23 - 00*	0,128733	0,058098	2,22	0,0270	0,01486	0,24260
DEFICITCAPITAL2	0,013472	0,003472	3,88	0,0000	0,00667	0,02028
TAXED_PROPERTY2	-0,051341	0,024740	-2,08	0,0380	-0,09983	-0,00285
COUNTRYMAILL*	-0,117856	0,044402	-2,65	0,0080	-0,20488	-0,03083
BLEKINGE*	0,106029	0,120127	0,88	0,3770	-0,12942	0,34147
DALARNA*	0,093816	0,099218	0,95	0,3440	-0,10065	0,28828
GOTLAND*	0,152776	0,145209	1,05	0,2930	-0,13183	0,43738
GÄVLEBORG*	0,027919	0,098556	0,28	0,7770	-0,16525	0,22108
HALLAND*	0,096499	0,107023	0,90	0,3670	-0,11326	0,30626
JÄMTLAND*	0,199695	0,133073	1,50	0,1330	-0,06112	0,46051
JÖNKÖPING*	0,136226	0,099383	1,37	0,1700	-0,05856	0,33101
KALMAR*	0,163055	0,108242	1,51	0,1320	-0,04910	0,37521
KRONOBERG*	0,020940	0,126842	0,17	0,8690	-0,22767	0,26954
NORRBOTTEN*	-0,055264	0,095746	-0,58	0,5640	-0,24292	0,13239
SKÅNE*	0,187530	0,061808	3,03	0,0020	0,06639	0,30867
SÖDERMANLAND*	0,056908	0,087632	0,65	0,5160	-0,11485	0,22866
UPPSALA*	-0,057603	0,087890	-0,66	0,5120	-0,22986	0,11466
VÄRMLAND*	-0,038848	0,106351	-0,37	0,7150	-0,24729	0,16960
VÄSTERBOTTEN*	-0,306250	0,120682	-2,54	0,0110	-0,54278	-0,06972
VÄSTERNORRLAND*	0,144761	0,093583	1,55	0,1220	-0,03866	0,32818
VÄSTMANLAND*	0,184918	0,083407	2,22	0,0270	0,02144	0,34839
VÄSTRA GÖTALAND*	0,205409	0,056833	3,61	0,0000	0,09402	0,31680
ÖREBRO*	0,078004	0,084331	0,92	0,3550	-0,08728	0,24329
ÖSTERGÖTLAND*	0,068013	0,088615	0,77	0,4430	-0,10567	0,24170
GENDER*	0,024764	0,028682	0,86	0,3880	-0,03145	0,08098
BLEKINGE LÄN*	0,022544	0,122670	0,18	0,8540	-0,21789	0,26297
GOTLAND LÄN*	-0,020606	0,151978	-0,14	0,8920	-0,31848	0,27727
GÄVLEBORG LÄN*	0,062124	0,089251	0,70	0,4860	-0,11281	0,23705
GÖTEBORG AND BOHUS LÄN*	-0,091205	0,071386	-1,28	0,2010	-0,23112	0,04871
HALLAND LÄN*	-0,052786	0,113935	-0,46	0,6430	-0,27609	0,17052
IMMIGRANT/FOREIGNER*	0,186600	0,080634	2,31	0,0210	0,02856	0,34464
IMMIGRANTAFTER1990*	-0,057382	0,132762	-0,43	0,6660	-0,31759	0,20283
IMMIGRANTORADOPTED*	0,288998	0,176198	1,64	0,1010	-0,05634	0,63434
JÄMTLAND LÄN*	-0,035813	0,128226	-0,28	0,7800	-0,28713	0,21551
JÖNKÖPING LÄN*	-0,032231	0,097718	-0,33	0,7420	-0,22376	0,15929
KALMAR LÄN*	-0,005847	0,099071	-0,06	0,9530	-0,20002	0,18833
KOPPARBERG LÄN*	-0,033065	0,096646	-0,34	0,7320	-0,22249	0,15636
KRISTIANSTAD LÄN*	0,010624	0,081732	0,13	0,8970	-0,14957	0,17082
KRONOBERG LÄN*	-0,048019	0,108084	-0,44	0,6570	-0,25986	0,16382
MALMÖHUS LÄN*	-0,066759	0,070226	-0,95	0,3420	-0,20440	0,07088
NORRBOTTEN LÄN*	0,035800	0,087259	0,41	0,6820	-0,13522	0,20682
SKARABORG LÄN*	-0,144087	0,090838	-1,59	0,1130	-0,32213	0,03395
SÖDERMANLAND LÄN*	-0,060714	0,088732	-0,68	0,4940	-0,23462	0,11320
UPPSALA LÄN*	-0,026303	0,100650	-0,26	0,7940	-0,22357	0,17097
VÄRMLAND LÄN*	0,085771	0,105837	0,81	0,4180	-0,12167	0,29321
VÄSTERBOTTEN LÄN*	0,145496	0,103535	1,41	0,1600	-0,05743	0,34842
VÄSTERNORRLAND LÄN*	-0,019911	0,092503	-0,22	0,8300	-0,20121	0,16139
VÄSTMANLAND LÄN*	-0,068533	0,086453	-0,79	0,4280	-0,23798	0,10091
ÄLVSBORG LÄN*	-0,115962	0,076047	-1,52	0,1270	-0,26501	0,03309
ÖREBRO LÄN*	0,124090	0,084778	1,46	0,1430	-0,04207	0,29025
ÖSTERGÖTLAND LÄN*	-0,068991	0,085457	-0,81	0,4190	-0,23648	0,09850
SUBM_PHONE*	-0,108747	0,031547	-3,45	0,0010	-0,17058	-0,04692
MARRIED*	-0,004499	0,036006	-0,12	0,9010	-0,07507	0,06607
BREDBANDSBOLAGET*	-0,187506	0,137836	-1,36	0,1740	-0,45766	0,08265
COMHEM*	-0,073950	0,126189	-0,59	0,5580	-0,32128	0,17338
GLOCALNET*	0,261253	0,184983	1,41	0,1580	-0,10131	0,62381
GMAIL*	-0,191309	0,128912	-1,48	0,1380	-0,44397	0,06135
HOME*	0,088958	0,134680	0,66	0,5090	-0,17501	0,35293
HOTMAIL*	0,198283	0,034653	5,72	0,0000	0,13037	0,26620
MSN*	0,225934	0,113152	2,00	0,0460	0,00416	0,44771
OTHER*	-0,040110	0,049295	-0,81	0,4160	-0,13673	0,05651

SPRAY*	-0,070530	0,085465	-0,83	0,4090	-0,23804	0,09698
STUDENT*	-0,238628	0,195632	-1,22	0,2230	-0,62206	0,14480
SWIPNET*	-0,397435	0,275079	-1,44	0,1490	-0,93658	0,14171
TELE2*	-0,203719	0,211856	-0,96	0,3360	-0,61895	0,21151
TELIA*	-0,050786	0,070849	-0,72	0,4730	-0,18965	0,08808
YAHOO*	0,212674	0,077890	2,73	0,0060	0,06001	0,36534
CO*	-0,262559	0,153362	-1,71	0,0870	-0,56314	0,03803
INHABITANTS2	-0,010710	0,011441	-0,94	0,3490	-0,03313	0,01171
INCOME2	-0,041639	2,979399	-17,80	0,0000	-0,04622	-0,03705
AGE	-0,056749	0,006892	-8,23	0,0000	-0,07026	-0,04324
AGE2	0,000550	0,000082	6,74	0,0000	0,00039	0,00071
JANUARY*	0,098341	0,061844	1,59	0,1120	-0,02287	0,21955
FEBRUARY*	0,175395	0,061539	2,85	0,0040	0,05478	0,29601
APRIL*	0,061164	0,060871	1,00	0,3150	-0,05814	0,18047
MAY*	0,094897	0,066149	1,43	0,1510	-0,03475	0,22455
JUNE*	0,179394	0,060521	2,96	0,0030	0,06078	0,29801
JULY*	0,113422	0,060724	1,87	0,0620	-0,00559	0,23244
AUGUST*	0,119656	0,060255	1,99	0,0470	0,00156	0,23775
SEPTEMBER*	0,124316	0,063090	1,97	0,0490	0,00066	0,24797
OCTOBER*	0,095199	0,063056	1,51	0,1310	-0,02839	0,21879
NOVEMBER*	0,056159	0,067088	0,84	0,4030	-0,07533	0,18765
DECEMBER*	0,132126	0,062824	2,10	0,0350	0,00899	0,25526
_CONS	-2,242464	0,210448	-10,66	0,0000	-2,65493	-1,82999

Note: 1 failure and 0 successes completely determined .

Appendix A2-3b: Probit regression: Marginal effects on all variables

		Number of obs	169 453
		Waldchi2(133)	2 421,10
Log pseudolikelihood =	-9262,7	Prob> chi2	0,0000
Mean default =	1,35%	Pseudo R2	0,2351

(standard errors adjusted for clustering on ID)

Variable	dF/dx	Robust Std. Err.	z	P>|z|	x-bar	dF/dX*std in % pts.	Change in default ratio
FAILEDBUYS	0,00053	0,00007	7,69	0,0000	0,29	0,13%	9,52%
T BABY*	-0,00034	0,00070	-0,47	0,6380	0,09	-0,03%	-2,49%
T CARS*	-0,00336	0,00071	-2,73	0,0060	0,01	-0,34%	-24,94%
T CHILDREN*	0,00233	0,00145	1,90	0,0580	0,02	0,23%	17,29%
T COMPUTERS*	-0,00022	0,00085	-0,25	0,8020	0,12	-0,02%	-1,62%
T COSMETICS*	0,00028	0,00069	0,42	0,6770	0,07	0,03%	2,09%
T ENTERTAINMENT*	0,00117	0,00121	1,06	0,2900	0,02	0,12%	8,71%
T EROTIC*	0,00051	0,00085	0,63	0,5310	0,05	0,05%	3,79%
T FASHION*	-0,00114	0,00051	-2,08	0,0370	0,14	-0,11%	-8,46%
T FITNESS*	-0,00242	0,00052	-3,60	0,0000	0,06	-0,24%	-17,92%
T GADGETS*	0,00187	0,00113	1,89	0,0590	0,04	0,19%	13,85%
T GIFTS*	0,00118	0,00183	0,71	0,4780	0,01	0,12%	8,77%
T HEALTH*	-0,00056	0,00066	-0,81	0,4180	0,06	-0,06%	-4,13%
T HOME*	-0,00193	0,00053	-3,05	0,0020	0,08	-0,19%	-14,31%
T LEISURE*	-0,00290	0,00063	-3,10	0,0020	0,01	-0,29%	-21,52%
T PETS*	0,00086	0,00131	0,70	0,4850	0,02	0,09%	6,35%
AVERAGESALES2	-0,00133	0,00020	-6,96	0,0000	1,95	-0,15%	-10,87%
SUM2	0,00351	0,00020	18,18	0,0000	5,82	0,32%	23,99%
TIMELASTCREDITCHECK	0,00003	0,00000	9,01	0,0000	8,12	0,11%	7,85%
PREVIOUSUNPAID2	0,01241	0,00078	20,65	0,0000	0,07	0,30%	22,10%
PREVIOUSPAID2	-0,01087	0,00086	-11,49	0,0000	0,20	-0,48%	-35,75%
PREVIOUSUNPAIDR2	0,00252	0,00199	1,26	0,2070	0,00	0,01%	0,85%
PREVIOUSPAIDR2	0,00267	0,00225	1,18	0,2380	0,04	0,05%	3,63%
PREVIOUSUNPAIDD2	0,02583	0,00464	5,75	0,0000	0,00	0,03%	2,19%
PREVIOUSPAIDD2	0,01337	0,00258	5,25	0,0000	0,01	0,15%	10,89%
00 - 01*	0,00492	0,00161	4,04	0,0000	0,03	0,49%	36,49%
01 - 02*	0,00567	0,00211	3,67	0,0000	0,01	0,57%	42,02%
02 - 03*	0,01480	0,00416	6,01	0,0000	0,01	1,48%	109,74%
03 - 04*	0,01188	0,00487	3,95	0,0000	0,00	1,19%	88,13%
04 - 05*	0,00888	0,00526	2,54	0,0110	0,00	0,89%	65,88%
05 - 06*	0,01184	0,00609	3,13	0,0020	0,00	1,18%	87,81%
06 - 07*	0,01396	0,00491	4,73	0,0000	0,00	1,40%	103,55%
07 - 08*	0,00376	0,00223	2,14	0,0330	0,01	0,38%	27,91%
08 - 09*	0,00000	0,00118	0,00	0,9980	0,02	0,00%	-0,02%
09 - 10*	-0,00001	0,00096	-0,01	0,9910	0,04	0,00%	-0,08%
10 - 11*	0,00088	0,00099	0,95	0,3430	0,05	0,09%	6,49%
11 - 12*	0,00199	0,00106	2,14	0,0320	0,06	0,20%	14,79%
12 - 13*	0,00214	0,00106	2,32	0,0200	0,06	0,21%	15,87%
13 - 14*	0,00197	0,00103	2,18	0,0290	0,06	0,20%	14,62%
14 - 15*	0,00096	0,00094	1,10	0,2720	0,06	0,10%	7,13%
15 - 16*	0,00135	0,00097	1,53	0,1270	0,06	0,13%	9,99%
16 - 17*	0,00034	0,00086	0,41	0,6840	0,06	0,03%	2,54%
17 - 18*	-0,00029	0,00078	-0,36	0,7170	0,06	-0,03%	-2,14%
18 - 19*	0,00142	0,00095	1,64	0,1020	0,06	0,14%	10,50%
19 - 20*	0,00007	0,00080	0,09	0,9270	0,07	0,01%	0,54%

20 - 21*	0,00045	0,00082	0,57	0,5680	0,07	0,05%	3,36%
22 - 23*	0,00114	0,00088	1,41	0,1600	0,07	0,11%	8,44%
23 - 00*	0,00210	0,00109	2,22	0,0270	0,05	0,21%	15,57%
DEFICITCAPITAL2	0,00019	0,00005	3,88	0,0000	5,72	0,08%	6,11%
TAXED_PROPERTY2	-0,00072	0,00034	-2,08	0,0380	0,13	-0,10%	-7,36%
COUNTRYMAILL*	-0,00147	0,00049	-2,65	0,0080	0,11	-0,15%	-10,90%
BLEKINGE*	0,00170	0,00219	0,88	0,3770	0,02	0,17%	12,60%
DALARNA*	0,00148	0,00174	0,95	0,3440	0,03	0,15%	10,94%
GOTLAND*	0,00261	0,00298	1,05	0,2930	0,01	0,26%	19,36%
GÄVLEBORG*	0,00041	0,00148	0,28	0,7770	0,03	0,04%	3,00%
HALLAND*	0,00152	0,00189	0,90	0,3670	0,03	0,15%	11,30%
JÄMTLAND*	0,00361	0,00303	1,50	0,1330	0,02	0,36%	26,79%
JÖNKÖPING*	0,00226	0,00193	1,37	0,1700	0,04	0,23%	16,72%
KALMAR*	0,00280	0,00224	1,51	0,1320	0,02	0,28%	20,78%
KRONOBERG*	0,00030	0,00187	0,17	0,8690	0,02	0,03%	2,24%
NORRBOTTEN*	-0,00073	0,00118	-0,58	0,5640	0,04	-0,07%	-5,38%
SKÅNE*	0,00318	0,00126	3,03	0,0020	0,12	0,32%	23,55%
SÖDERMANLAND*	0,00086	0,00141	0,65	0,5160	0,03	0,09%	6,34%
UPPSALA*	-0,00075	0,00107	-0,66	0,5120	0,04	-0,08%	-5,59%
VÄRMLAND*	-0,00052	0,00136	-0,37	0,7150	0,03	-0,05%	-3,86%
VÄSTERBOTTEN*	-0,00301	0,00078	-2,54	0,0110	0,03	-0,30%	-22,31%
VÄSTERNORRLAND*	0,00242	0,00185	1,55	0,1220	0,03	0,24%	17,97%
VÄSTMANLAND*	0,00326	0,00181	2,22	0,0270	0,03	0,33%	24,15%
VÄSTRA GÖTALAND*	0,00348	0,00115	3,61	0,0000	0,15	0,35%	25,84%
ÖREBRO*	0,00120	0,00143	0,92	0,3550	0,03	0,12%	8,92%
ÖSTERGÖTLAND*	0,00103	0,00146	0,77	0,4430	0,04	0,10%	7,67%
GENDER*	0,00035	0,00041	0,86	0,3880	0,43	0,03%	2,59%
BLEKINGE LÄN*	0,00033	0,00182	0,18	0,8540	0,02	0,03%	2,41%
GOTLAND LÄN*	-0,00028	0,00202	-0,14	0,8920	0,01	-0,03%	-2,09%
GÄVLEBORG LÄN*	0,00094	0,00145	0,70	0,4860	0,04	0,09%	6,96%
GÖTEBORG AND BOHUS LÄN*	-0,00116	0,00082	-1,28	0,2010	0,07	-0,12%	-8,59%
HALLAND LÄN*	-0,00069	0,00140	-0,46	0,6430	0,02	-0,07%	-5,14%
IMMIGRANT/FOREIGNER*	0,00331	0,00177	2,31	0,0210	0,02	0,33%	24,55%
IMMIGRANTAFTER1990*	-0,00075	0,00161	-0,43	0,6660	0,01	-0,07%	-5,55%
IMMIGRANTORADOPTED*	0,00592	0,00500	1,64	0,1010	0,00	0,59%	43,92%
JÄMTLAND LÄN*	-0,00048	0,00164	-0,28	0,7800	0,02	-0,05%	-3,56%
JÖNKÖPING LÄN*	-0,00043	0,00127	-0,33	0,7420	0,03	-0,04%	-3,22%
KALMAR LÄN*	-0,00008	0,00137	-0,06	0,9530	0,03	-0,01%	-0,60%
KOPPARBERG LÄN*	-0,00045	0,00125	-0,34	0,7320	0,03	-0,04%	-3,30%
KRISTIANSTAD LÄN*	0,00015	0,00118	0,13	0,8970	0,03	0,02%	1,12%
KRONOBERG LÄN*	-0,00063	0,00135	-0,44	0,6570	0,02	-0,06%	-4,71%
MALMÖHUS LÄN*	-0,00087	0,00085	-0,95	0,3420	0,08	-0,09%	-6,46%
NORRBOTTEN LÄN*	0,00052	0,00133	0,41	0,6820	0,05	0,05%	3,88%
SKARABORG LÄN*	-0,00170	0,00089	-1,59	0,1130	0,03	-0,17%	-12,61%
SÖDERMANLAND LÄN*	-0,00079	0,00107	-0,68	0,4940	0,03	-0,08%	-5,87%
UPPSALA LÄN*	-0,00036	0,00132	-0,26	0,7940	0,03	-0,04%	-2,65%
VÄRMLAND LÄN*	0,00134	0,00183	0,81	0,4180	0,03	0,13%	9,90%
VÄSTERBOTTEN LÄN*	0,00244	0,00205	1,41	0,1600	0,03	0,24%	18,09%
VÄSTERNORRLAND LÄN*	-0,00027	0,00124	-0,22	0,8300	0,03	-0,03%	-2,02%
VÄSTMANLAND LÄN*	-0,00089	0,00103	-0,79	0,4280	0,03	-0,09%	-6,56%
ÄLVSBORG LÄN*	-0,00142	0,00081	-1,52	0,1270	0,04	-0,14%	-10,54%
ÖREBRO LÄN*	0,00203	0,00160	1,46	0,1430	0,03	0,20%	15,03%
ÖSTERGÖTLAND LÄN*	-0,00089	0,00102	-0,81	0,4190	0,05	-0,09%	-6,62%
SUBM_PHONE*	-0,00158	0,00048	-3,45	0,0010	0,61	-0,16%	-11,70%
MARRIED*	-0,00006	0,00050	-0,12	0,9010	0,29	-0,01%	-0,47%
BREDBANDSBOLAGET*	-0,00210	0,00120	-1,36	0,1740	0,02	-0,21%	-15,54%
COMHEM*	-0,00095	0,00147	-0,59	0,5580	0,01	-0,09%	-7,01%
GLOCALNET*	0,00515	0,00491	1,41	0,1580	0,01	0,51%	38,16%
GMAIL*	-0,00212	0,00111	-1,48	0,1380	0,02	-0,21%	-15,75%
HOME*	0,00140	0,00236	0,66	0,5090	0,01	0,14%	10,37%
HOTMAIL*	0,00315	0,00063	5,72	0,0000	0,27	0,32%	23,37%
MSN*	0,00424	0,00276	2,00	0,0460	0,01	0,42%	31,48%

OTHER*	-0,00054	0,00064	-0,81	0,4160	0,13	-0,05%	-4,02%
SPRAY*	-0,00091	0,00101	-0,83	0,4090	0,03	-0,09%	-6,74%
STUDENT*	-0,00248	0,00145	-1,22	0,2230	0,00	-0,25%	-18,41%
SWIPNET*	-0,00344	0,00130	-1,44	0,1490	0,01	-0,34%	-25,50%
TELE2*	-0,00221	0,00173	-0,96	0,3360	0,01	-0,22%	-16,42%
TELIA*	-0,00067	0,00089	-0,72	0,4730	0,07	-0,07%	-4,99%
YAHOO*	0,00390	0,00182	2,73	0,0060	0,02	0,39%	28,93%
CO*	-0,00266	0,00107	-1,71	0,0870	0,01	-0,27%	-19,76%
INHABITANTS2	-0,00015	0,00016	-0,94	0,3490	10,91	-0,02%	-1,43%
INCOME2	-0,00058	0,00004	-17,80	0,0000	11,09	-0,20%	-15,11%
AGE	-0,00080	0,00010	-8,23	0,0000	33,74	-0,93%	-69,22%
AGE2	0,00001	0,00000	6,74	0,0000	1275,8	0,71%	52,49%
JANUARY*	0,00154	0,00107	1,59	0,1120	0,08	0,15%	11,39%
FEBRUARY*	0,00299	0,00126	2,85	0,0040	0,08	0,30%	22,16%
APRIL*	0,00092	0,00097	1,00	0,3150	0,09	0,09%	6,79%
MAY*	0,00147	0,00113	1,43	0,1510	0,09	0,15%	10,92%
JUNE*	0,00307	0,00125	2,96	0,0030	0,08	0,31%	22,74%
JULY*	0,00180	0,00109	1,87	0,0620	0,08	0,18%	13,36%
AUGUST*	0,00191	0,00109	1,99	0,0470	0,08	0,19%	14,19%
SEPTEMBER*	0,00200	0,00116	1,97	0,0490	0,08	0,20%	14,82%
OCTOBER*	0,00148	0,00109	1,51	0,1310	0,08	0,15%	10,99%
NOVEMBER*	0,00084	0,00106	0,84	0,4030	0,08	0,08%	6,22%
DECEMBER*	0,00215	0,00118	2,10	0,0350	0,07	0,21%	15,93%

obs. P |	0,0134846	
pred. P |	0,0048312	(at x-bar)

(*) dF/dx is for discrete change of dummy variable from 0 to 1
z and P>|z| correspond to the test of the underlying coefficient being 0

Appendix A2-3c: Adjusted and non-adjusted marginal effects on all significant non-dummy variables

Variable	Change from	To	Standard deviation	Number of std.	dF/dX*std in % pts.	Change in default
FAILEDBUYS	0,03	1,98	2,408	0,81	0,10%	7,71%
AVERAGESALES2	6,05	20,18	1,100	1	-0,15%	-10,87%
SUM2	335,76	846,09	0,922	1	0,32%	23,99%
TIMELASTCREDITCHECK	8	44	36,269	1	0,11%	7,85%
PREVIOUSUNPAID2	0,07	0,95	0,240	2,5	0,74%	55,24%
PREVIOUSPAID2	0,02	1,08	0,444	1,6	-0,77%	-57,20%
PREVIOUSUNPAIDR2	0,00	1,03	0,045	15,5	0,18%	13,15%
PREVIOUSPAIDR2	0,04	1,01	0,183	3,6	0,18%	13,06%
PREVIOUSUNPAIDD2	0,00	0,99	0,011	60	1,77%	131,32%
PREVIOUSPAIDD2	0,01	1,00	0,110	6,2	0,91%	67,51%
DEFICITCAPITAL2	303,75	23 819	4,359	1	0,08%	6,11%
TAXED_PROPERTY2	0,14	3,53	1,379	1	-0,10%	-7,36%
COUNTRYMAILL*	54 990	198 811	1,285	1	-0,02%	-1,43%
INCOME2	65 609	2 150 843	3,490	1	-0,20%	-15,11%
AGE	33,7	45,5	11,728	1	-0,23%	-16,73%

Comments: AGE is calculated from AGE and AGE2
PREVIOUSPAID2; the from has been decreased by 0,4 std while the to has been increased by 1,2 std
FAILEDBUYS; the from has been decrease by 0,11 std while the to has been increased by 0,7 std

A3 Descriptive statistics

Appendix A3-1a Reporting descriptive statistics before calculus

Variable	Min	P - 25	P - 50	P - 75	Max	Mean	Std.
DEFAULT	0	0	0	0	1	0,01348	0,11534
FAILEDBUYS	0	0	0	0	102	0,29069	2,40813
AVERAGESALES	0,0027	1,7762	5,427	15,558	49,239	11,7919	14,7892
SUM	20	189	334	596	35250	523,839	692,676
TIMELASTCHECK	0	0	0	0	601	8,11527	36,2692
PREVIOUSUNPAID	0	0	0	0	19	0,11505	0,50154
PREVIOUSPAID	0	0	0	0	50	0,42003	1,357
PREVIOUSUNPAIDR	0	0	0	0	4	0,00412	0,07079
PREVIOUSPAIDR	0	0	0	0	16	0,06444	0,37881
PREVIOUSUNPAIDD	0	0	0	0	2	0,00025	0,01801
PREVIOUSPAIDD	0	0	0	0	13	0,0232	0,20234
DEFICIT_CAPITAL	0	0	2476	13905	3E+06	9430	17181,4
TAXED_PROPERTY	0	0	0	0	6E+07	28562,9	500922
COUNTRYMAIL	0	0	0	0	1	0,10644	0,3084
GENDER	0	0	0	1	1	0,42625	0,49453
SUBM_PHONE	0	0	1	1	1	0,61156	0,4874
MARRIED	0	0	0	1	1	0,29187	0,45463
CO	0	0	0	0	1	0,00939	0,09647
MAILADDRESS	0	0	0	0	1	0,01229	0,11016
INHABITANTS	2549	21369	53952	111207	786509	133050	211080
INCOME	-3E+06	96116	188620	257622	5E+07	197217	256939
AGE	17	24	32	41	98	33,7382	11,7284

Appendix A3-1b Reporting descriptive statistics after calculus

Variable	Min	P - 25	P - 50	P - 75	Max	Mean	Std.
DEFAULT	0	0	0	0	1	0,01348	0,11534
FAILEDBUYS	0	0	0	0	102	0,29069	2,40813
AVERAGESALES2	0,0027	1,0211	1,8605	2,8069	3,9168	1,95341	1,09981
SUM2	3,0445	5,247	5,8141	6,3919	10,47	5,81938	0,92242
TIMELASTCHECK	0	0	0	0	601	8,11527	36,2692
PREVIOUSUNPAID2	0	0	0	0	2,9957	0,06731	0,24015
PREVIOUSPAID2	0	0	0	0	3,9318	0,20192	0,44364
PREVIOUSUNPAIDR2	0	0	0	0	1,6094	0,00274	0,04539
PREVIOUSPAIDR2	0	0	0	0	2,8332	0,03754	0,18338
PREVIOUSUNPAIDD2	0	0	0	0	1,0986	0,00017	0,01143
PREVIOUSPAIDD2	0	0	0	0	2,6391	0,01431	0,10983
DEFICIT_CAPITAL2	0	0	7,8148	9,5401	14,775	5,71948	4,3588
TAXED_PROPERTY2	0	0	0	0	17,94	0,13131	1,37862
COUNTRYMAIL	0	0	0	0	1	0,10644	0,3084
GENDER	0	0	0	1	1	0,42625	0,49453
SUBM_PHONE	0	0	1	1	1	0,61156	0,4874
MARRIED	0	0	0	1	1	0,29187	0,45463
CO	0	0	0	0	1	0,00939	0,09647
MAILADDRESS	0	0	0	0	1	0,01229	0,11016
INHABITANTS2	7,8438	9,9697	10,896	11,619	13,575	10,9149	1,2852
INCOME2	-9,2103	11,473	12,147	12,459	17,627	11,0915	3,48989
AGE	17	24	32	41	98	33,7382	11,7284
AGE2	289	576	1024	1681	9604	1275,82	918

Appendix A3:2a Reporting proportions on dummy variables

Type	Prop.
OTHER - BC	16,65%
BABY	8,52%
CARS	0,96%
CHILDREN	1,82%
COMPUTERS	12,08%
COSMETICS	7,01%
DATING	3,88%
ENTERTAINMENT	2,29%
EROTIC	4,92%
FASHION	14,22%
FITNESS	5,64%
GADGETS	3,68%
GIFTS	0,73%
HEALTH	6,17%
HOME	8,46%
LEISURE	1,46%
PETS	1,51%

- BC = Base case

Maillan	Prop.
STOCKHOLM - BC	19,56%
BLEKINGE	1,88%
DALARNA	3,20%
GOTLAND	0,63%
GÄVLEBORG	3,38%
HALLAND	2,77%
JÄMTLAND	1,52%
JÖNKÖPING	3,54%
KALMAR	2,46%
KRONOBERG	2,03%
NORRBOTTEN	3,99%
SKÅNE	12,23%
SÖDERMANLAND	3,12%
UPPSALA	3,55%
VÄRMLAND	3,26%
VÄSTERBOTTEN	3,23%
VÄSTERNORR~D	3,35%
VÄSTMANLAND	3,18%
VÄSTRA GÖTALAND	15,49%
ÖREBRO	3,21%
ÖSTERGÖTLAND	4,41%

Domain name	Prop.
NONE - BC	38,48%
BREDBAND	2,19%
COMHEM	1,21%
GLOCALNET	0,64%
GMAIL	1,71%
HOME	0,82%
HOTMAIL	27,18%
MSN	0,80%
OTHER	13,21%
SPRAY	3,21%
STUDENT	0,38%
SWIPNET	0,66%
TELE2	0,59%
TELIA	6,86%
YAHOO	2,06%

Ordertime	Prop.
21 - 22 - BC	7,45%
00 - 01	2,54%
01 - 02	1,25%
02 - 03	0,61%
03 - 04	0,38%
04 - 05	0,24%
05 - 06	0,23%
06 - 07	0,46%
07 - 08	1,12%
08 - 09	2,39%
09 - 10	3,91%
10 - 11	5,14%
11 - 12	5,72%
12 - 13	5,55%
13 - 14	6,15%
14 - 15	6,38%
15 - 16	6,45%
16 - 17	6,12%
17 - 18	5,85%
18 - 19	6,02%
19 - 20	6,74%
20 - 21	7,33%
22 - 23	6,70%
23 - 00	5,26%

Month	Prop.
MARCH - BC	9,36%
JANUARY	8,14%
FEBRUARY	7,97%
APRIL	9,26%
MAY	9,22%
JUNE	8,26%
JULY	8,28%
AUGUST	8,31%
SEPTEMBER	8,18%
OCTOBER	8,06%
NOVEMBER	7,51%
DECEMBER	7,46%

Lancode	Prop.
STOCKHOLMS LÄN - BC	18,28%
BLEKINGE LÄN	1,95%
GOTLANDS LÄN	0,77%
GÄVLEBORGS LÄN	3,57%
GÖTEBORGS OCH BOHUS LÄN	7,29%
HALLANDS LÄN	2,06%
IMIGRANT OR FOREIGN CITIZEN	2,09%
IMMIGRANT AFTER 1990	0,58%
IMMIGRANT OR ADOPTED	0,19%
JÄMTLANDS LÄN	1,71%
JÖNKÖPINGS LÄN	3,44%
KALMAR LÄN	2,92%
KOPPARBERGS LÄN	3,34%
KRISTIANSTADS LÄN	3,26%
KRONOBERGS LÄN	2,13%
MALMÖHUS LÄN	7,67%
NORRBOTTENS LÄN	4,78%
SKARABORGS LÄN	2,94%
SÖDERMANLANDS LÄN	3,09%
UPPSALA LÄN	2,74%
VÄRMLANDS LÄN	3,29%
VÄSTERBOTTENS LÄN	3,24%
VÄSTERNORRLANDS LÄN	3,46%
VÄSTMANLANDS LÄN	3,25%
ÄLVSBORGS LÄN	4,32%
ÖREBRO LÄN	3,10%
ÖSTERGÖTLANDS LÄN	4,55%

A4 Multicollinearity

Appendix A4-1a: Correlation table: Multicollinearity between variables exceeding ABS(0.2)

	FAILEDBUYS	*PREVIOUSUNPAID2*	*PREVIOUSPAID2*	*PREVIOUSUNPAIDR2*	*PREVIOUSPAIDR2*
PREVIOUSUNPAID2	0,2259	1			
PREVIOUSPAID2		0,261	1		
PREVIOUSUNPAIDR2		0,2241		1	
PREVIOUSPAIDR2			0,4947		1
PREVIOUSUNPAIDD2				0,2542	
PREVIOUSPAIDD2			0,3214		0,6863

Appendix A4-1b: Correlation table: Multicollinearity between variables exceeding ABS(0.2)

	*HOTMAIL**	*COUNTRYMAILL**	*T COMPUTERS**
*OTHER**	-0,2384		
INHABITANTS2		-0,2407	
AVERAGESALES2			0,5951
*GENDER**			0,2362

Appendix A4-1c: Correlation table: Multicollinearity between variables exceeding ABS(0.2)

	AGE	*DEFICITCAPITAL2*	*MARRIED**	*INCOME2*
*MARRIED**		0,2685	1	
INCOME2		0,2738		1
AGE	1	0,3322	0,3885	0,3079
AGE2	0,9827	0,2571	0,354	0,2583

Appendix A4-1d: Correlation table Län: Multicollinearity between Län-variables exceeding ABS(0.2)

	*SKÅNE**	*KALMAR**	*GOTLAND**	*BLEKINGE**	*JÖNKÖPING**
*MALMÖHUS LÄN**	0,6191				
*KRISTIANSTAD LÄN**	0,3531				
*KALMAR LÄN**		0,5899			
*GOTLAND LÄN**			0,5726		
*BLEKINGE LÄN**				0,6134	
*JÖNKÖPING LÄN**					0,5718
	*HALLAND**	*VÄSTRA GÖTALAND**	*ÖREBRO**	*ÖSTERGÖTLAND**	*KRONOBERG**
*HALLAND LÄN**	0,5529				
*GÖTEBORG AND BOHUS LÄN**		0,4797			
*SKARABORG LÄN**		0,2838			
*ÄLVSBORG LÄN**		0,364			
*ÖREBRO LÄN**			0,6215		
*ÖSTERGÖTLAND LÄN**				0,6096	
*KRONOBERG LÄN**					0,5327
	*SÖDERMANLAND**	*VÄSTMANLAND**	*UPPSALA**	*DALARNA**	*GÄVLEBORG**
*SÖDERMANLAND LÄN**	0,5575				
*VÄSTMANLAND LÄN**		0,6012			
*UPPSALA LÄN**			0,4944		
*KOPPARBERG LÄN**				0,6463	
*GÄVLEBORG LÄN**					0,6786
	*VÄRMLAND**	*JÄMTLAND**	*VÄSTERBOTTEN**	*VÄSTERNORRLAND**	*NORRBOTTEN**
*VÄRMLAND LÄN**	0,7096				
*JÄMTLAND LÄN**		0,589			
*VÄSTERBOTTEN LÄN**			0,685		
*VÄSTERNORRLAND LÄN**				0,6576	
*NORRBOTTEN LÄN**					0,7052

Appendix B: Tables

Table 8:

List of Swedish and International credit reporting agencies (alphabetical order)

Swedish	International
Business Check	Dun & Bradstreet
Creditsafe	Experian
Dun & Bradstreet (Soliditet)	Equifax
Upplysningscentralen	TransUnion

Source: Upplysningscentralen, Dun & Bradstreet, Creditsafe, Business Check, Experian, Equifax, TransUnion

Table 9:

Type of characteristics

Financial / Demographic	Behavioural
Sex	Number of late payments
Age	Purpose of loan
Occupation	Exceeded credit limit
Annual income	Prior month's purchase record
Running water	Amount of loan

Table 10

Hypotheses: Direct financial reality

#	Hypothesis	Variable(s)
H1	High income is negatively correlated with probability of default	*INCOME2*
H2	A high debt burden is positively correlated with probability of default	*DEFICIT_CAPITAL2*
H3	Personal wealth decreases the probability of default	*TAXED_PROPERTY2*
H4	Marriage is negatively correlated with probability of default	*MARRIED*

Table 11

Hypotheses: Indirect financial ability

#	Hypothesis	Variable(s)
H5	Age is relevant in determining the probability of default	*AGE; AGE2*
H6	Men are more likely to default than women	*GENDER*
H7	People from the countryside are less likely to default	*COUNTRYMAIL*
H8	People’s willingness and/or ability to pay varies between regions	*MAILLAN*
H9	People’s probability of default should differ depending on where they were born	*LANCODE*
H10	City size has an impact on probability of default	*INHABITANTS2*
H11	People living on a care of-address are more likely to default	*CO*
H12	People's probability of default should not depend on in which month they were born	*MONTH*
H13	Payment history is relevant when estimating the probability of default	*PREVIOUSUNPAID2; PREVIOUSPAID2; PREVIOUSUNPAIDR2; PREVIOUSPAIDR2; PREVIOUSUNPAIDD2; PREVIOUSPAIDD2*

Table 12

Hypotheses: Moral hazard

#	Hypothesis	Variable(s)
H14	People that submit voluntary information are less likely to default	*SUBM_PHONE TYPE;*
H15	Probability of default should differ depending on type of store	*AVERAGESALES2*
H16	People that try to maximise their credit have a higher probability of default	*FAILEDBUYS*
H17	Loan size increases probability of default	*SUM2*
H18	People's email-addresses tell us something about the probability of default	*DOMAIN_NAME*
H19	People ordering at awkward times of the day are more likely to default	*ORDERTIME*

Table 13

Regressions

Regression	Type of variables	Comment
1	Demographic	
2	Behavioural and Demographic	Cluster on (*ID*)
3	All variables	Cluster on (*ID*)

Table 14

Hypotheses: Direct financial ability

#	Hypothesis	Variable(s)	Decision
H1	High income is negatively correlated with probability of default	*INCOME2*	Accepted
H2	A high debt burden is positively correlated with probability of default	*DEFICIT_CAPITAL2*	Accepted
H3	Personal wealth decreases the probability of default	*TAXED_PROPERTY2*	Accepted
H4	Marriage is negatively correlated with probability of default	*MARRIED*	Rejected

Table 8

Hypotheses: Direct financial ability

#	Variable(s)	Change in prob. of default	Exp(Mean (μ))	Exp(Mean + 1 std dev. (μ + δ²))
H1	*INCOME2*	*-15.11%*	65,609	2,150,843
H2	*DEFICIT_CAPITAL2*	*6.11%*	304	23,819
H3	*TAXED_PROPERTY2*	*-7.36%*	0.14	3.53

Table 9

Hypotheses: Indirect financial ability

#	Hypothesis	Variable(s)	Decision
H5	Age is relevant in determining the probability of default	*AGE; AGE2*	Accepted
H6	Men are more likely to default than women	*GENDER*	Rejected
H7	People from the countryside are less likely to default	*COUNTRYMAIL*	Accepted
H8	People’s willingness and/or ability to pay varies between regions	*MAILLAN*	Accepted
H9	People’s probability of default should differ depending on where they were born	*LANCODE*	Accepted
H10	City size has an impact on probability of default	*INHABITANTS2*	Rejected
H11	People living on a care of-address are more likely to default	*CO*	Rejected
H12	People's probability of default should not depend on in which month they were born	*MONTH*	Rejected
H13	Payment history is relevant when estimating the probability of default	*PREVIOUSUNPAID2; PREVIOUSPAID2; PREVIOUSUNPAIDR2; PREVIOUSPAIDR2; PREVIOUSUNPAIDD2; PREVIOUSPAIDD2*	Accepted

Table 10

Hypotheses: Moral hazard

#	Hypothesis	Variable(s)	Decision
H14	People that submit voluntary information are less likely to default	*SUBM_PHONE*	Accepted
H15	Probability of default should differ depending on type of store	*TYPE; AVERAGESALES2*	Accepted
H16	People that try to maximise their credit have a higher probability of default	*FAILEDBUYS*	Accepted
H17	Loan size increases probability of default	*SUM2*	Accepted
H18	People's email-addresses tell us something about the probability of default	*DOMAIN_NAME*	Accepted
H19	People ordering at awkward times of the day are more likely to default	*ORDERTIME*	Accepted

Appendix C: Equations

Equation 5

$$Pn = E(yn \mid Xn) = F(\alpha + \beta Xn)$$

Equation 6

$F(\alpha + \beta X_n) = \int_{-\infty}^{\alpha+\beta Xn} f(z)dz$ is the cumulative standard normal distribution function

Equation 7

$$f(z) = [1/(2\pi)]^{1/2} \exp(-z^2/2)$$

Equation 8

DEFAULT = β_0 + β_1(*FAILEDBUYS*) + β_2(*STORE CATEGORY*) + β_3(*AVERAGESALES2*) + β_4(*SUM2*) + β_5(*TIMELASTCREDITCHECK*) + β_6(*PREVIOUSUNPAID2*) + β_7(*PREVIOUSPAID2*) + β_8(*PREVIOUSUNPAIDR2*) + β_9(*PREVIOUSPAIDR2*) + β_{10}(*PREVIOUSUNPAIDD2*) + β_{11}(*PREVIOUSPAIDD2*) + β_{12}(*TIMEOFPURCHASE**)+ β_{13}(*DEFICITCAPITAL2*) + β_{14}(*TAXED_PROPERTY2*) + β_{15}(*COUNTRYMAILL**) + β_{16}(*LANBORN**) + β_{17}(*GENDER**) + β_{18}(*LIVING LAN**) + β_{19}(*SUBM_PHONE**) + β_{20}(*MARRIED**) + β_{21}(*EMAILDOMAIN**) + β_{22}(*CO**) + β_{23}(*INHABITANTS2*) + β_{24}(*INCOME2*) + (β_{25}*β_{26}(*AGE*)) + β_{27}(*BIRTH MONTH**)

Appendix D: Figures

Figure 2*: Plot of probit function* – P_n take values from 0 to 1 on the X-axis and X_n take values from *(-1-α)/β* to *(1-α)/β* on the Y-axis.