



Norwegian University of  
Science and Technology

# A Novel Profit Scoring Method for Classifying Credit Card Applications

**Helge Skarestad**

Master of Science in Physics and Mathematics

Submission date: June 2017

Supervisor: Jo Eidsvik, IMF

Norwegian University of Science and Technology  
Department of Mathematical Sciences



## **Preface**

This master's thesis was written as a part of the study programme Physics and Mathematics at the Department of Mathematical Sciences at the Norwegian University of Science and Technology during the spring of 2017. Upon completion, the thesis leads to a master's degree in Industrial Mathematics. The project has been carried out in cooperation with a Norwegian credit card company and the thesis has therefore been written in a way which assumes that the reader has some background in credit card analyses.

I would like to thank my supervisor Prof. Jo Eidsvik for the helpful guidance and discussions we have had throughout the project. I would also like to thank my external advisors for sparking the ideas behind the thesis and providing me with relevant data sets.

Trondheim, June 2017

Helge Skarestad



## Abstract

This thesis presents a new scoring method for credit card applications. The method balances the risk and the expected profits an applicant represents to a credit card company. In addition, the EMP cut-off selection procedure introduced by Verbraken et al. [European Journal of Operational Research, 238(2), 2014] is adapted to credit card scoring methods. In order to test the methods, they are applied to a set of credit card accounts containing observations of the individual net profit margins. A defaulter is defined as an account which, after a year of credit card usage, represents a negative cumulative profit margin. Models for gains, losses and the probability that an account becomes a defaulter are built. The new scoring method is compared to two conventional scoring methods; the probability of becoming a defaulter and the expected profit conditional on the account becoming a defaulter or not. The cut-off selection technique is compared to the conventional methods of maximising accuracy, using efficiency curves and educated guesses of cut-off values.

The results show that the new credit scoring method can outperform conventional methods in terms of profitability when reliable models for gains and probabilities of defaulting can be built. The new cut-off selection procedure can also outperform conventional methods in terms of profitability as it utilises the distribution of losses and average gain.

## Samandrag

Denne avhandlinga presenterer ein ny metode for å score kredittkortsøknadar. Metoden balanserer risikoen og den venta profitten kvar søker representerer. I tillegg vert EMP-metoden for å velje avskjering originalt introdusert av Verbraken et al. [European Journal of Operational Research, 238(2), 2014] vidareutvikla og tilpassa kredittkortsøknadar. Dei nye metodane vert testa på eit sett med kredittkortkontoar med tilhøyrande individuelle observasjonar av månadleg profittmargin. Eit misleghald av kredittkortlån er definert ved at den kumulative profittmarginen for 12 månadar er negativ. Modellar for vinst, tap og sannsynet for at ein kredittkortsøkar mislegheld lånet vert tilpassa. Den nye score-funksjonen vert samanlikna med to konvensjonelle metodar; sannsynet for at låna vert misleghaldt og den venta profitten føresett at kontoen vert misleghaldt eller ikkje. Den nye teknikken for å velje avskjering vert samanlikna med konvensjonelle metodar som å maksimere nøyaktigheit, avlesing av effektkurver og kvalifiserte gjettingar av cut-off verdier.

Resultata syner at den nye scoringmetoden kan konkurrere ut dei etablerte metodane når pålitelege modellar for vinst og sannsyn for misleghald kan byggjast. Den nye metoden for å velje avskjering kan konkurrere ut konvensjonelle metodar ved å nytte observerte fordelingar av tap og vinst.

# Contents

Preface . . . . .	i
Abstract . . . . .	iii
Samandrag . . . . .	iv
<b>1 Introduction</b>	<b>1</b>
1.1 Background . . . . .	1
1.1.1 Notation . . . . .	3
1.1.2 Literature survey . . . . .	4
1.2 Scope . . . . .	8
1.2.1 Problem formulation . . . . .	9
1.2.2 Approach . . . . .	9
<b>2 Background and descriptive data analysis</b>	<b>10</b>
2.1 Approach . . . . .	10
2.2 Variables . . . . .	11
2.2.1 Grouping . . . . .	12
2.2.2 Missing values and anomalies . . . . .	13
2.2.3 Response variable . . . . .	13
2.3 Exploratory data analysis . . . . .	15
2.4 Problems in credit scoring . . . . .	19

**3 Methods and models 21**

3.1 Scoring functions . . . . . 21

    3.1.1 Probability of default . . . . . 22

    3.1.2 Conditional expected profit . . . . . 22

    3.1.3 Curves in risk-reward plane . . . . . 24

3.2 Setting a cut-off value . . . . . 28

    3.2.1 Educated guesses of cut-off . . . . . 29

    3.2.2 Accuracy . . . . . 29

    3.2.3 Efficiency curves . . . . . 30

    3.2.4 Expected Maximal Profit (EMP) . . . . . 31

**4 Results 39**

4.1 Results from case study . . . . . 39

    4.1.1 Probability of not defaulting . . . . . 40

    4.1.2 Gain given case 1 . . . . . 41

    4.1.3 Loss given case 0 . . . . . 43

    4.1.4 Applying scoring functions and cut-off . . . . . 44

4.2 Sensitivity analysis of scoring functions . . . . . 48

4.3 Comparing risk-reward scores and probability scores . . . . . 51

4.4 Sensitivity analysis of cut-off selections . . . . . 53

**5 Summary 56**

5.1 Summary and conclusions . . . . . 56

5.2 Discussion . . . . . 57

5.3 Recommendations for Further Work . . . . . 59

**Bibliography 60**



# List of Figures

1.1	Example densities and PDF of bad and good accounts . . . . .	4
2.1	Flow chart of approach in profit scoring analyses . . . . .	11
2.2	Frequency of losses among defaulters in data set A . . . . .	15
2.3	Frequency of gains among non-defaulters in data set A . . . . .	15
2.4	Frequency of losses among defaulters in data set B . . . . .	16
2.5	Frequency of gains among non-defaulters in data set B . . . . .	16
3.1	Difference between probability cut-off and risk-reward cut-off . . . . .	25
3.2	Examples of risk-reward cut-off . . . . .	27
3.3	Examples of efficiency curves . . . . .	31
4.1	Model assessment - estimated probabilities in data set A . . . . .	41
4.2	Model assessment - estimated probabilities in data set B . . . . .	41
4.3	Model assessment - estimated gains in data set A . . . . .	43
4.4	Model assessment - estimated gains in data set B . . . . .	43
4.5	Model assessment - estimated losses in data set A . . . . .	44
4.6	Model assessment - estimated losses in data set B . . . . .	44
4.7	Efficiency curves of scoring functions applied to data set A . . . . .	45
4.8	Efficiency curves of scoring functions applied to data set B . . . . .	46

4.9	Efficiency curves and cut-off values from probability scores in data set B . . .	47
4.10	Sensitivity analysis of scoring functions - varying probability estimates . . . .	49
4.11	Sensitivity analysis of scoring functions - varying gain estimates . . . . .	50
4.12	Scenarios where risk-reward scoring beat probability scoring . . . . .	52
4.13	Sensitivity analysis of cut-off methods - varying probability scores . . . . .	54
4.14	Sensitivity analysis of cut-off methods - varying profitability scores . . . . .	55

## List of Tables

2.1	Types of variables in data sets . . . . .	12
2.2	Number of accepted and defaulted accounts in data sets . . . . .	15
2.3	Distributions of profits in some customer segments . . . . .	18
4.1	Profit scoring of training subset of data set B . . . . .	46
4.2	Profit scoring of testing subset of data set B . . . . .	48





# Chapter 1

## Introduction

Credit cards give consumers the possibility to obtain short term loans. The loans are financed by financial institutions against an interest rate. A standard procedure is to collect interest on the loan if the loan is not paid back within a month after the lending date. Loan-providing institutions use statistical models to determine who are given credit and how much money they can loan. It is always desirable to avoid that people are trapped in debt they can't manage. Therefore, the statistical models are continuously refined and improved upon. Even though a model works well today, external factors might influence the way people behave. Some segments of a population may, in time, experience changes in their ability to pay off debt. Such changes need to be accounted for in order for credit card companies to prevail (Thomas, 2000).

### 1.1 Background

When someone wants a credit card, they apply for one at a credit card company. The company decides whether to accept the application and also how large credit is given. The decision tools that companies use are scoring methods. Each applicant is given a score,

typically a number on a fixed scale on the real line. Traditionally, this score has been based on an estimate of the probability that the applicant fails to pay back the loan. Applications with a score lower than a certain threshold are denied credit, while applications with a high score are accepted and credit is granted. Throughout this thesis, such thresholds are referred to as *cut-off values*.

All methods in which credit card applicants are given some kind of score will be referred to as *credit scoring methods*. Some credit scoring methods focus on optimising profit related measures. These methods will be referred to as *profit scoring methods*. Each credit scoring method and cut-off are applied to create *portfolios*. A portfolio is a collection of credit card accounts that a company holds. In the context of credit card loans, a *defaulter* is someone who fails to pay back their loan. Whenever a credit card user becomes a defaulter, the financial institution that provided the loan will have extra expenses with debt collection, extra bank fees, etc.. *Transactors* are people who pay their credit card bills before interest rates add up. *Revolvers* are people who fail to pay the entire debt before extra interest rates add up to the debt. Credit card users that contribute to the profits of credit card companies are all revolvers, but not all revolvers contribute positively to the company's profits. Some revolvers build a debt which is too large to handle and thus become defaulters.

Why is it desirable to develop profit driven scoring methods? As pointed out by [Finlay \(2010\)](#), maximising likelihood is not as important as maximising profit. This reflects the conventional credit scoring method in which a statistical model such as logistic regression is applied to estimate the probability that a credit card applicant will become a defaulter. In a business context, the performance of a portfolio is not measured in how well the logistic model fits the observed credit card accounts, but rather the profit from the portfolio. For instance, classifying two accounts with the same probability of default does not necessarily say anything about the profitability of them. One account might be inactive or a transactor. As the company has expenses with each customer (such as issuing the card, customer service, insurance etc...), idle customers and transactors don't contribute to the profit of the

company. The second account might belong to a revolver. Neither of the accounts default their loans, but the second account contributes greatly to the profit of the company.

A streamlining of a credit card portfolio manifests itself in two different ways; one scenario is that the number of defaulting accounts in the portfolio is lowered and the total profit is increased. For instance by denying credit to all revolvers who will default in the future. Another scenario is that inactive or non-profitable accounts are denied credit such that the average profit per account in the portfolio is increased. In this case, it is desirable to reject credit card applications from both transactors and revolvers that end up in default.

### 1.1.1 Notation

There are some concepts and notations that need explanation before discussing credit scoring analyses. Throughout the thesis, an account which is defaulted will be denoted as a case 0 or belonging to class 0. The opposite case is case 1 belonging to class 1. Class 0 and 1 will also be called bad and good accounts, respectively. A scoring function  $S(\mathbf{x})$  is a transformation of some set of variables  $\mathbf{x}$  into the real line. When each credit card applicant in a data set has a score, the applications can be sorted in terms of this score. The cumulative distribution function  $F_0(t)$  gives the portion of case 0 accounts with scores lower than  $t$ . The sociated density function is  $f_0(t)$ .  $F_1(t)$  and  $f_1(t)$  have the same interpretation, but for case 1 accounts. Figure 1.1 shows an example of how  $F_{1/0}(t)$  and  $f_{1/0}(t)$  might appear. A typical trait is that the scoring function gives lower scores to case 0 accounts than case 1 accounts. For the cut-off  $T$  in Figure 1.1, the majority of bad accounts belong to the area where the scores are less than  $T$ . There is also some overlap between  $f_0$  and  $f_1$  in Figure 1.1, meaning that some good accounts would also be rejected by setting a cut-off  $T$ . In the setting of binary classification,  $AUC$  is the area below the Receiver Operating Characteristic (ROC) curve. The ROC curve is the plot of  $F_0(s)$  on the y-axis vs.  $F_1(s)$  on the x-axis. A perfect credit scoring model has  $AUC$  equal to 1 which can only happen when there is no overlap between  $f_0(t)$  and  $f_1(t)$ .

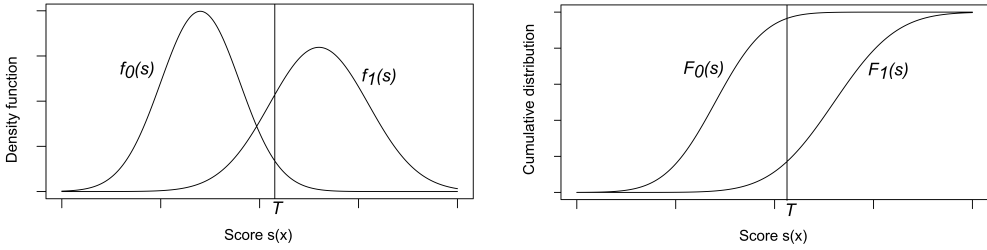


Figure 1.1: Example densities  $f_0(s)/f_1(s)$  and probability distribution functions  $F_0(s)/F_1(s)$  of accounts belonging to class 0/1. Accounts with scores less than  $T$  belong to the area to the left of the vertical line.

$\pi_0$  ( $\pi_1$ ) is the probability that an account ends in default (not-default). In some settings, these probabilities are treated as a priori constants equal to the ratio of bad (good) accounts in a portfolio. In other methods, these probabilities are estimates based on information on each account. In such methods, the probabilities are denoted with a hat and a superscript,  $\hat{\pi}_0^i$  ( $\hat{\pi}_1^i$ ) indicating that they are estimates belonging to application number  $i$ . The two cases 0 and 1 are the only possible outcomes and mutually exclusive. Therefore, the probabilities  $\pi_0$  and  $\pi_1$  always sum to one.

### 1.1.2 Literature survey

As credit scoring methods are used by most loan-providing companies, there is a wide range of literature dealing with scoring functions of varying forms. Among the recent contributions in credit scoring, some articles and books stand out as especially informative and relevant. Some of these texts are large benchmark studies, while others introduce new and significant findings in terms of credit scoring methods and ways of assessing such methods.

One of the earliest argumentations of why models should include measures of profitability is given by [Boyes et al. \(1989\)](#). It is recognised that a profitable portfolio is more important to credit card companies than maximising some likelihood. A profit scoring model



is trained in which probabilities of default are estimated. Expected profits for a set of credit card applications are estimated using the probability estimates and educated guesses regarding profits. The guesses are based on the average profits per non-defaulter and the average losses per defaulter. However, the lack of precise observations of profits for each account belonging to the applicants prevents a suitable assessment of the model.

In the description part of the book by [Thomas et al. \(2002\)](#), the authors describe the book as the first to detail the mathematical models for credit scoring. The text gives an overview of computer intensive methods such as neural networks as well as conventional scoring methods such as logistic regression and classification trees. One chapter is devoted to profit scoring methods. Here, the gain by accepting a good payer and the loss by accepting a bad payer are treated as constants in a similar manner as was done by [Boyes et al. \(1989\)](#). In addition to discussing profit score, a handful of profit-based assessment tools for credit scoring methods are shown. In stead of relying on fixed estimates of losses and gains when investigating profitability of a certain scoring function, it is possible to assess the credit scoring method by using the ratio of losses to gains. Credit scoring methods can thus be compared without making assumptions on losses and gains other than that they are constant in a portfolio.

[Baesens et al. \(2003\)](#) provides a comprehensive benchmark study of 17 credit scoring methods on eight credit card data sets from banks of different nationalities. Each method seeks to estimate the probability that an account will default. The results are reported in three ways. First, all applications with probability higher than 0.5 are accepted. Second, as all applications are assigned to distinct groups due to the variables being categorical, the groups where good/bad ratio is higher than 5/1 are accepted. Third, the AUC of each scoring model is given as an indicator of how well the scoring functions separate non-defaulters from defaulters. The most prominent results in terms of these three assessment methods are that logistic regression, linear programming, support vector machines and neural networks give the best scores. However, for many of the data sets, the differences in performance for the methods are not large. No consideration is given to how one should choose

a cut-off value, it is simply shown that some credit scoring methods give better results than others if the correct cut-off is chosen. Also, there are no profit-driven assessment measure used in the article. The sole interest lies in finding models with optimal explanatory powers regarding the classification case 0 and case 1 accounts.

In two articles [Hand \(2005\)](#) and [Hand \(2009\)](#), it is argued why AUC is an unsuitable measure for comparing scoring methods. Hand rigorously shows that when AUC is used as an indicator of the optimal selection strategy in a credit card portfolio, the method in which AUC is calculated will vary from classifier to classifier. This implies that the conclusions made by [Baesens et al. \(2003\)](#) using AUC to compare different scoring methods are not necessarily trustworthy. This new insight creates a void in the available assessment methods of credit scoring methods. However, Hand provides an alternative to AUC which is coherent for different credit scoring methods. This alternative measure, named *H-measure* hinges on a fixed distribution of average losses and gains in a portfolio. With these new insights into credit scoring, it is established that the average gains and losses follow some kind of distribution. The H-measure is refined in a later article by [Hand and Anagnostopoulos \(2014\)](#).

[Crook et al. \(2007\)](#) give an overview of some current research topics in credit scoring. When discussing profit scoring, *efficiency curves* and *strategy curves* are shown to be relevant assessment tools. Another topic discussed by [Crook et al. \(2007\)](#) is the Basel II accord ([Basel, 2005](#)) which set new standards for risk management in financial institutions. One implication of the Basel II accord is that financial institutions should keep easily interpretable statistical models for estimating risk.

[Finlay \(2008\)](#) provides a study of profit scoring methods. Three different scoring methods are compared. First, ordinary estimates of probability of default are created. Second, in addition to cases 0 and 1 (default and not-default), a third state of indeterminate accounts where profit is approximately zero is added. Third, account worth is modelled as a continuous linear combination of several variables from each account's payment history. The main result is that using such a continuous measure of worth in a profit scoring method can

outperform the conventional credit scoring method in terms of profitability even though the percentage of correctly classified accounts is reduced.

A similar conclusion is drawn by [Finlay \(2010\)](#). The expected profit conditional on eventual status (default/not default) is modelled together with probabilities of default. Again, it is shown that scoring methods built using regression on account profit can outperform conventional credit scoring methods in terms of profit measures. No consideration is given to how the optimal cut-off should be chosen.

[Stewart \(2011\)](#) explains why credit card profit models have been difficult to build and implement. First, banks usually have no clear definition of profit at an account level. Second, profit is often highly correlated with risk. Third, profit distributions are difficult to model e.g. due to heavy tails in the distributions of losses and gains. Fourth, due to changes in overall economy, credit card users change habits from one year to the next depending on outer factors. In addition to discussing the current state of profit scoring models, [Stewart \(2011\)](#) proposes a method to model the revenue of credit cards. This revenue measure is then used to build profit scoring models. By building separate profit models for different bins of risk groups, the profit scoring method may outperform the conventional credit scoring methods in terms of profitability

A new type of assessment method for credit scoring models in bank lending is introduced by [Verbraken et al. \(2013, 2014\)](#). The Expected Maximum Profit (EMP) performance measure is a coherent alternative to AUC. It is shown that EMP is the same as AUC under some conditions, but also how to avoid these conditions. EMP indicates the maximal profit obtainable for a scoring function, when applied to a portfolio of bank loans. In addition to this assessment method, a cut-off selection strategy is introduced which accounts for variations in the expected loss among defaulters. Models for estimating probabilities of default are built using logistic regression and artificial neural networks. Cut-off values are decided using EMP as a cut-off selection technique. The introduction of EMP constitutes a shift in the way of thinking about profit scoring. The distributions of profitability can not only be used when building scoring functions, but also in the cut-off decision.

Twelve years after the previous large benchmark study was conducted, [Lessmann et al. \(2015\)](#) gives an update to [Baesens et al. \(2003\)](#). A large number of newly developed credit scoring methods are added to the benchmark study from 2003. It is noted that most recent studies use a small number of data sets, and that AUC still is an industry standard in measuring the performance of scoring methods. 41 different classification methods are applied to eight different data sets and the performance of each classification function is measured using six different assessment methods. In addition to the benchmark study, the authors address the profitability of portfolios stemming from a selection of methods. It is assumed that the losses and gains among bad and good accounts are constants. One conclusion which is drawn is that although logistic regression is the industry standard, there exist methods that can outperform logistic regression in terms of maximising the predictive powers of classification methods. Examples of such methods are artificial neural networks and regression trees.

## 1.2 Scope

Most studies of credit card scoring models are focused on estimating probabilities of defaulting. Although there are some examples of profit scoring methods, there does not seem to be any methods that directly balance risk and reward. Historically, there has been a lack of precise observations of the profitability of each credit card in a portfolio. This has been a problem in building profit scoring methods, as it is difficult to assess each model. Also, there is a need for cut-off selection methods which optimise profit criteria among credit card applications. The scope of this thesis will be to investigate profit scoring methods. A new method is introduced which seeks to balance the risk and reward by accepting a credit card application. In addition, the cut-off selection procedure of [Verbraken et al. \(2014\)](#) will be redeveloped for use in credit card scoring. Other credit scoring methods and cut-off selection techniques will also be included as means of comparison. The thesis includes a case study of two sets of Norwegian credit card accounts which contain observations of the

profitability of each credit card account.

### 1.2.1 Problem formulation

The purpose of this thesis is to investigate methods of scoring credit card accounts that optimise portfolio profits. The main contributions can be stated in two points:

- (i) *Introduce a new profit scoring method for credit card applications*
- (ii) *Adapt the cut-off selection technique of [Verbraken et al. \(2014\)](#) for use in credit card scoring*

It is understood that the performance of methods and techniques described in this thesis will be conditional on the available data sets. Different types of methods might perform better on different types of data sets. This is not a study which can find the overall best scoring methods, but rather a limited comparison of some methods in some limited scenarios.

### 1.2.2 Approach

The thesis is organised as follows. Chapter 2 is a descriptive data analysis of a set of credit card accounts. Chapter 3 gives an overview of three credit scoring methods of which one is previously undescribed. In addition, five cut-off selection techniques are described of which two are refined versions of the EMP cut-off selection introduced by [Verbraken et al. \(2014\)](#). Chapter 4 provides descriptions of the results of applying the credit scoring methods and cut-off selections to credit card applications. Chapter 5 contains summary and conclusions with a discussion of the results and recommendations for future work.

## Chapter 2

# Background and descriptive data analysis

### 2.1 Approach

Figure 2.1 shows a flow chart which explains the general procedure behind credit scoring analyses. A large data set consisting of credit card applications is cleaned and prepared. Among the applications in the original data set, only a portion were approved by the credit card company and accounts created. The accounts were active for more than a year and their net profit margin were calculated after each month. A training subset of these applications and observations of profit margins are used to build credit scoring models as well as setting suitable cut-off values. The models are applied to the unseen testing set to assess the performance of each method.

The basis of any credit scoring analyses are observations of credit card applicants and the eventual status of these. Scoring functions,  $S(\mathbf{x})$  are maps  $S : \mathbf{x} \mapsto \mathbb{R}$  where  $\mathbf{x}$  contains some relevant observations of economic, demographic or social conditions of a credit card

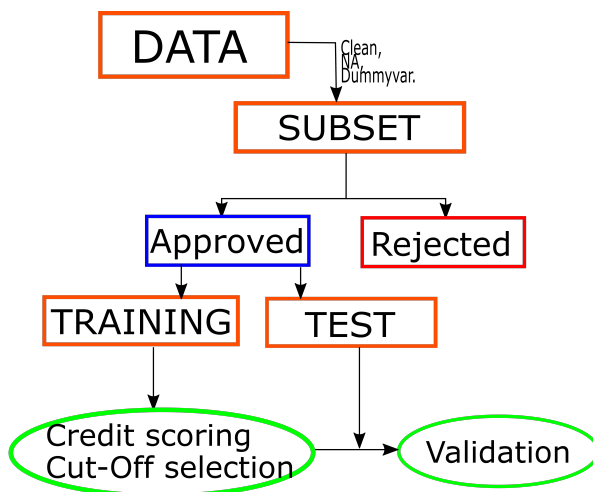


Figure 2.1: Flow chart showing the approach in profit scoring analyses.

applicant. The case study included in this thesis is based on information of accounts belonging to a Norwegian credit card company. For sensitivity reasons, the company is left nameless. The data comprise 93 830 credit card applications that were filed in the period January 2014 through January 2016. It is reasonable to separate the data into two subsets due to differences in the composition of variables. The two different data sets will be denoted set A and B.

## 2.2 Variables

Due to sensitivity reasons, a complete list of the available variables can't be given. However, Table 2.1 gives an overview of the types of variables in each data set. Technical variables show technical features of each application, such as e.g. application date, how the application was filed and how much credit the applicant wanted. Personal variables reflect personal traits of the applicant, such as education level, at what year he/she graduated or whether the applicant owns or rents his/her house/apartment. Demographic variables

give general information such as where the applicant lives or the age of the applicant. Economical variables are meant to give indications to what type of economical situation the applicant is in and consists of e.g. debt, fortune or the number of credit cards the applicant holds from before.

Among the variables listed in Table 2.1, there are both grouped and continuous variables.

### 2.2.1 Grouping

There are two motivations for grouping variables. For continuous variables, there might be non-linear traits which is difficult to capture without allowing for groups. All the continuous variables are grouped to identify potential non-linearity.

The other situation in which an additional grouping of variables should be done is when the existing type of grouping is too coarse. If one group of a variable contains less than 1% of the total observations, then any estimation based on this group would be prone to over-fitting. Therefore, the observations in the specific group are moved to a similar group (if possible) or simply added to the largest group. As a consequence, variables where more than 99% of observations belong to one group are removed. The grouping should be done in a way which maintains a high number of observation in each group and sensible compositions of accounts in each group. Chi-square tests for independence of groups was applied to ensure that the different groups can give reliable indications of the probabilities of default and distribution of profits. County (fylke) is an example of a variable which should be re-grouped. The number of applicants in some counties is too low to give reliable representations. For instance, there were only two accounts from Svalbard. It

Table 2.1: The composition and number of variables in each data set.

Data set	Technical	Personal	Demographic	Economical	Total
A:	11	6	4	26	47
B:	9	7	3	29	48



is preferable not to merge two distant counties. The behaviour of people in two different parts of the country might be influenced by geography, demographics, cultural differences, large cities, etc.. in a way which make them incomparable. The merging of counties should also be balanced with the observed past behaviour of the inhabitants in each county. If two neighbouring counties show totally different credit histories, then they should not be merged.

### **2.2.2 Missing values and anomalies**

Missing values stem from applicants not knowing what to write or forgetting to fill in information. It is thinkable that missing values can carry some information about the applicant. For instance, it might mean something that someone is reluctant in giving away debt information. Without proper statistical backup, such speculations are baseless. Therefore, the missing values in each variable are grouped together.

Some applications contain anomalies. For instance, one application reported a mortgage of 32 billion NOK. This is obviously an error, but in some cases, it can be hard to tell the errors from the real observations. The number of accounts is too high to traverse and manually assess whether the reported debts or incomes make sense. To avoid some of the errors, all stated income and debt variables are truncated at a reasonable level. Some people have reported their total debt to be 1234567 and some people have filled in their phone number instead of their age. In many cases, such anomalies are hard to identify as there are no logic behind the errors people do.

### **2.2.3 Response variable**

A very important feature of the data sets is the cumulative profit margin of each account. These observations are what enables building and testing profit scoring methods. After an application has been approved and an account created, the monthly net margin of each account is calculated. The monthly net margin is the balance between money paid by the

customer to the credit card company and the money he/she owed the company after each month,

$$\text{Monthly Margin} = \text{Money Paid} - \text{Money Owed.}$$

Money Owed consists of the amount the customer has loaned in addition to several variable costs such as interest rates, bank charge and potential debt collection fees. Such costs occur when a customer needs to pay interest rates or transfer money. Therefore, all accounts with margin equal to zero belong to inactive users. Accounts with margin greater than zero belong to customers who have paid more interest rates than the variable costs they have incurred. It is believed that the cumulative sum of the first twelve monthly margins after the account was created is a reliable indicator of the individual customer's behaviour. One year is the time it takes for most bad payers to reveal themselves. Also, one year is sufficiently close in time to the application date. The personal and economic changes the customer might experience in the first twelve months after the application was accepted is limited. Throughout the thesis, the twelve-month cumulative profit margin will simply be referred to as the *profit* or the *margin*.

A common definition of a defaulter (Lucas, 2001; Finlay, 2008) is an account that has fallen behind on three or more consecutive payments. This is not necessarily an accurate definition as there might be accounts that belong to revolvers who contribute immensely to the profit of the company for nine months and then fails to pay a smaller amount in the three last months of the first year. Such an account will come across as a defaulter, but the cumulative profit margin is positive. Since the goal of profit scoring methods is to obtain a portfolio of high profit accounts, it is still desirable to include such an account in the credit card portfolio. In this thesis, a different definition will be adopted. A defaulter will be defined as an account where, after the first twelve months, the margin is negative.

Table 2.2: The number of applications, number of accepted applications, number of defaulted accounts and the percentage of defaulted accounts among the accepted applications in data sets A and B.

Data set	N applic.	N accepted	N defaults	Defaults among accepted
A:	43751	38991	1459	3.74 %
B:	50079	29098	1673	5.75 %

## 2.3 Exploratory data analysis

Table 2.2 shows the total number of applications, how many of these were accepted by the credit card company and the percentage of accepted accounts that were defaulted. An observation worth noticing from Table 2.2 is that data set A contains almost ten thousand more accounts than data set B, but a lower number of defaulters.

Figures 2.2 and 2.3 show histograms of the margin among case 0 and case 1 accounts in data set A, respectively. The largest part of losses among defaulters in data set A are

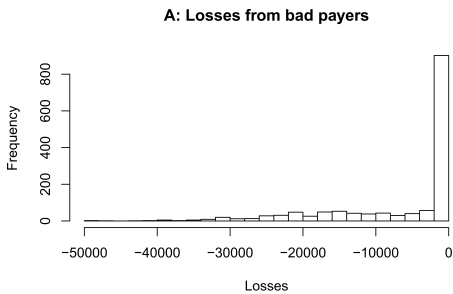


Figure 2.2: Frequency of losses among non-defaulters in data set A (truncated at -50000).

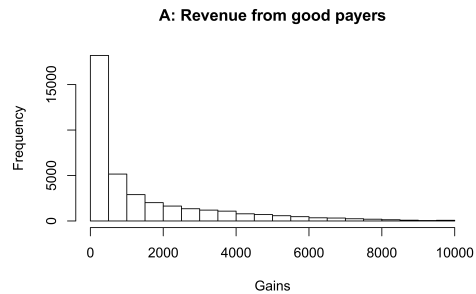


Figure 2.3: Frequency of profits among non-defaulters in data set A (truncated at 10000).

only slightly negative. These accounts typically belong to transactors who pay less interest rates than the variable costs that have occurred. Another large part of the defaulters are the ones who represent losses in the range (-25000,-1000). Such accounts seem to occur at somewhat similar rates. A final observation worth mentioning is the heavy tail. Some

accounts have defaulted on their entire credit amount in addition to interest rates and debt collection fees. Compared to the number of slightly negative losses, the number of giant losses is low. However, due to the magnitude of such large losses, accepting only one of these bad accounts is as damaging to the company profit as accepting 100 accounts with slightly negative losses.

A large part of the profitable accounts represent zero margin or slightly positive margins ( $\leq 250$ ). These accounts belong to people who are mainly transactors, but fall behind on some payments. A difference between distribution of losses and profits is that the ratio of accounts representing intermediate and large profits seem to decay rather smoothly. The decay ends up in a heavy tail. The heavy tails in distributions of losses and gains represent motivations that credit scoring methods should account for the differences in profitability. Accounts representing profits larger than 10000 should be of a higher worth to the company than the accounts representing slightly positive profits.

Figures 2.4 and 2.5 show the same types of histograms as Figures 2.2 and 2.3, but for data set B. In general, the distributions of losses and gains seem to be similar in data sets

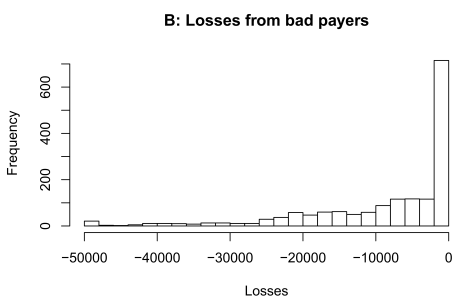


Figure 2.4: Frequency of losses among non-defaulters in data set B (truncated at -50000).

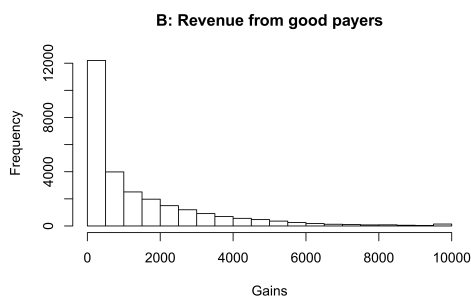


Figure 2.5: Frequency of profits among non-defaulters in data set B (truncated at 10000).

A and B. However, there are some minor differences. For instance, the ratio of accounts representing small losses compared to intermediate losses is lower in data set B.

If the trends shown in Figures 2.2 - 2.5 are to be modelled, then the variables from each application should carry some information about differences in profits and loss. Table 2.3 shows some traits among a selection of segments in data sets A and B. In each segment, the number of accounts and percentage of defaulted accounts is reported along with mean, median, standard deviation and skewness of the profit observations. There are many types of observations which can be gathered from Table 2.3, but in the following, only a few will be mentioned to highlight the presence of profit scoring opportunities and some differences between data set A and B.

Among the three different work situation segments listed in Table 2.3, accounts belonging to employed people stand out as very profitable compared to retirees and students. The percentage of defaulted accounts is lower among employed customers and both mean and median profits are more than twice as high as for retirees and students in both datasets. In data set A, the employed customers also represent a lower skewness indicating that the losses in the segment generally are smaller than those for both students and retirees. Employed customers tend to pay their debts, however not always at the due date.

Home owners are more profitable customers than renters. The percentage of home owners who end up as defaulters is smaller than that of renters and the mean profit among home owners is higher in both data sets. A large part of the losses among renters stem from some extreme observations of defaulted accounts. This is also reflected in the standard deviation, which is higher among renters in both data sets.

There are also some variations across the geographical segments. Oslo shows a clear tendency to be a less profitable part of the country than Finmark and Rogaland, especially in data set B.

The different age groups are clearly showing the non-linear trait which is desirable to capture. Accounts belonging to 45-54-year-old customers are more profitable and have a lower percentage of defaults than accounts belonging to 24-34-year-old or 65-89-year-old customers.

Some general remarks on the distributions of profits can also be made. An important

Table 2.3: Total number of accounts, percentage of defaulted accounts, mean, median, st. dev. and skewness of the profitability in some segments of the two data sets.

		Data set A					Data set B				
		N(%d)	Mean	Median	St.Dev	Skewness	N(%d)	Mean	Median	St.Dev	Skewness
All accounts		38997(3.7)	1104.5	489.5	2984.8	-4.3	29098(5.7)	816.2	563.9	3949.5	-6.2
Work	Employee	22559(3.5)	1286.4	656.8	3064.2	-3.5	21108(5.8)	888.1	709.2	4395.6	-5.8
	Retiree	3940(4.3)	572.2	93.4	1936.9	-4.3	3013(5.2)	433.2	112.3	2107.4	-5.4
	Student	2468(4.2)	636.0	391.8	1822.3	-6.1	1304(6.3)	444.6	326.1	1200.0	-5.4
Housing	Home owner	4229(3.2)	967.5	377.6	2214.9	-4.3	12779(4.0)	1010.0	426.7	3199.2	-6.0
	Renter	2604(4.9)	798.0	575.5	3476.5	-5.6	9638(8.3)	618.3	898.9	5109.1	-5.5
Fylke	Finnmark	1262(3.5)	1393.2	667.7	2635.6	-2.5	894(4.1)	1237.9	834.1	3231.8	-4.1
	Rogaland	6058(3.6)	1080.1	501.4	2784.5	-4.3	4380(5.7)	946.7	646.2	3659.9	-5.1
	Oslo	2890(4.0)	964.4	502.5	3071.8	-5.2	2675(8.7)	262.0	640.7	6007.0	-5.0
Age	25-34	10658(3.9)	958.8	615.5	2712.0	-4.9	7553(7.1)	651.4	848.8	5160.3	-5.5
	45-54	6307(3.1)	1591.6	689.3	3873.6	-3.6	3963(4.1)	1146.5	615.9	4066.1	-6.1
	65-89	4309(4.2)	561.9	79.2	1669.0	-3.8	2528(4.5)	399.8	77.3	1289.4	-0.5

observation is that all skewness is negative. This is also apparent from Figures 2.2-2.5. There is a heavier tail among extreme losses than extreme gains. The differences in percentages of defaulted accounts is much higher in data set B. This means that there might be better potential of creating a credit scoring function relying on probability estimates in data set B. Also, in both data sets, segments with high percentage of defaulters also has low mean profit. As explained by Stewart (2011), large correlation between risk and profit is a reason why profit scoring methods have been difficult to build.

## 2.4 Problems in credit scoring

In all credit scoring analyses, there are some problems that occur when the accept/reject decision is based on a statistical method. Hand (2005) lists three problems. Firstly, the true behaviour of credit card customers is not discovered until the future. Consequently, all implemented scoring methods are outdated as soon as they are implemented. For instance, changes in the overall economy can have grave impact on a credit card portfolio. Secondly, the final profitability of an account is not discovered before the end of the loan period. In practice, this means that accounts that seem profitable after 12 months, may end up in default after 24 months. Thirdly, having identified that an account has high risk of defaulting, it is common to take actions in to reduce the financial risk by e.g. giving a low credit limit. Such actions will not be modelled in this thesis.

The biggest issue with all credit scoring methods is the selection bias which occurs when working with a set of real observations of credit card usage. To observe the profitability of an account or to know whether a credit card user defaults, it is necessary to first sign him/her as a customer. The source of the problem is shown in Table 2.2. Among the total numbers of applications, only 89% and 58% were accepted in data sets A and B, respectively. This previous screening of accounts means that there might be bias among the accepted applications. There is no information available about the would-have-been performance of previously rejected applications. Thus, any new models based on the pre-

viously accepted accounts might have a different performance if implemented as the main selection criterion of a credit card company.

Attempts have been made in the past to correct for sample bias. *Reject inference* is a methodology which has been used in various forms in a range of texts (Boyes et al., 1989; Hand and Henley, 1993; Thomas, 2000; Crook and Banasik, 2004). One solution is to assume that all previously rejected applications would have been defaulters. A more refined method is to use the previously accepted applications to extrapolate the probability of default to previously rejected applications. All the proposed solutions to correct the selection bias have one thing in common. They all rely on some type of assumption and there is no clear way of testing the validity of these assumptions without knowing the would-have-been performance of the accounts. In profit scoring, the situation is even more complex as it is no longer a matter of simply estimating a binary outcome. Insight into the would-have-been margin of the rejected applications is more complex than binary accept/reject analysis. Financial institutions are aware of this problem and the straight-forward solution to it; namely to occasionally accept applications with scores lower than the cut-off. This was done with the accounts in data sets A and B. However, it is not known how many low-scoring applications were accepted.

Having established that there presumably is a bias in data sets A and B and that such bias can't be corrected in a reliable way, the rest of this thesis is put in a similar setting as in Verbraken et al. (2014). The possible selection bias is disregarded and the focus is to add a profit scoring method on top of the credit scoring method which is already implemented in the credit card company. Each of the models proposed in this thesis would come in addition to the current procedure. Thus, the performance of each method should be compared to the base case in which all the previously accepted applications are kept in the portfolio.



# Chapter 3

## Methods and models

The process of constructing a framework for accepting and rejecting credit card applications consists of two steps. First, scoring functions need to be created based on the variables gathered in all applications. Second, cut-off values need to be found such that the spectra of scores stemming from the different scoring functions can be subjected to a binary decision rule. This chapter is dedicated to showing examples how these steps can be performed.

### 3.1 Scoring functions

The three types of scoring functions which will be investigated in this thesis all share some traits. First, a high score should indicate that the application represents more worth than a low score. Second, the Basel II accord sets standards which implies that the scoring rules should be interpretable. The methods described in this thesis should give any credit card analyst the necessary insight to explain to any rejected credit card applicant why he/she was denied a credit card.

### 3.1.1 Probability of default

As the conventional method of credit scoring relies on estimates of the probability of default, it is natural to include such estimates as a means of comparison with other methods. The probability that an application  $i$  will end up as a case 1 twelve months after it is accepted is  $\pi_1^i$ . The estimated probability is denoted  $\hat{\pi}_1^i$  and the score given to the application is

$$S_{\text{Prob}}(\mathbf{x}) = \hat{\pi}_1^i. \quad (3.1)$$

There are several methods available for estimating such probabilities and a comparison between these is beyond the scope of this thesis. However, some remarks about possible choices of methods can be made. Support vector machines and neural networks have been used by e.g. [Verbraken et al. \(2014\)](#) and [Maldonado et al. \(2015\)](#). Although the methods may provide good results, they are not easily interpretable. One of the reflections made by [Lessmann et al. \(2015\)](#) was that random forests should be included in benchmark studies of new credit scoring methods. Random forests produce regression trees that are easier to interpret, but they might be complex as there is a large number of variables gathered from each application. Although the methods mentioned above have proven to be more accurate than logistic regression in various benchmark studies, logistic regression is both easily interpreted and there exist straight forward methods of model selection. This is why logistic regression is applied in this thesis. A thorough introduction to logistic regression is given by [Christensen \(1997\)](#).

### 3.1.2 Conditional expected profit

Having observations of the profitability of each account in the data sets motivates a profit scoring method which builds on the estimates of probability of default. The goal is to extend the existing scoring model via the conditional expectation (3.2), similar to what was

done by [Finlay \(2010\)](#). For account  $i$ , the expected profit conditional on eventual status is

$$E[P^i] = E[P^i|0] \cdot \pi_0^i + E[P^i|1] \cdot (1 - \pi_0^i) \quad (3.2)$$

where  $P^i$  is the profit of account  $i$ . A profit scoring function can thus be built by estimating the three different quantities in the conditional expectation (3.2). From equation (3.2), let  $L^i = -E[P^i|0]$  and  $G^i = E[P^i|1]$ .  $L^i$  denotes the loss which is expected from account  $i$  if the account defaults.  $G^i$  is the expected profit gained from account  $i$  if the account does not default. Models for estimating losses  $\hat{L}^i$  and gains  $\hat{G}^i$  can be built from the population of case 0 and case 1 in each data set. Each application is then given a score according to the scoring function

$$S_{\text{Prof}}(\mathbf{x}) = \hat{G}^i(\mathbf{x}) \cdot \hat{\pi}_1^i(\mathbf{x}) - \hat{L}^i(\mathbf{x}) \cdot (1 - \hat{\pi}_1^i(\mathbf{x})). \quad (3.3)$$

Although this method is intuitive, the quality of the scoring function hinges on the quality of the models behind the estimated parameters.

Figures 2.2-2.5 show the distributions of losses and gains in the populations of case 0 and 1 accounts in both data sets. Starting with the gains, the distributions seem to decay at an exponential rate. Such observations motivate the use of Poisson regression and treating the observed gains as count-data. To account for dispersion in the observed gains, negative binomial models will also be estimated. In such models, the dispersion parameter is treated as a gamma distributed random variable. Negative binomial regression will give the same fitted values as Poisson regression, but each fitted coefficient will have higher standard errors owing to the additional variance introduced in the dispersion parameter. This has implications for the interpretations of each variable in the models. Some might appear to be insignificant even though the ordinary Poisson regression shows that they are. To account for the differences in credit limit among different accounts, the credit limit can be included as an offset in the negative binomial regression. Yet another observation which can be made from Figures 2.3 and 2.5 is that the number of accounts where the gain is zero

seems to be artificially high. Therefore, also zero inflated models should be considered. Insurance claims follow similar distributions as the observed gains and thorough descriptions of suitable regression models can therefore be found in insurance literature such as [Ohlsson and Johansson \(2010\)](#) or [Bølviken \(2014\)](#).

The observed distribution of losses, is less smooth. In both data sets, many losses are close to zero and there are heavy tails. The largest losses are mainly observed among accounts with high credit limit. Therefore, the estimates should be corrected for credit limit. One way to do this is to include granted credit as an offset, but the loss-pattern shown in [Figures 2.2 and 2.4](#) are still present. Either a very small part, or a large part of credit limit is lost. This motivates a binomial model in which the response variable is loss divided by credit limit rounded to 0 or 1.

As the distribution of losses might prove difficult to model, including them in the scoring method could reduce the predictive powers of the scoring function. It is thinkable that in situations where the data on defaulters are scarce, some extreme observations may lead to overfitted models for losses  $\hat{L}^i$  and thus give a bias in estimated profits. In such cases, a profit scoring method which builds on estimated probabilities and gains alone may be more suitable.

### 3.1.3 Curves in risk-reward plane

A fundamental concept in finance is that any stock portfolio can be tailor made to balance the risk of losing money with the potential gain if everything goes well. ([van der Wijst, 2013](#), pp. 60–61). Although a similar idea has been present in the credit scoring literature since before [Boyes et al. \(1989\)](#), there does not seem to be any profit scoring method which builds on such a balance directly. [Finlay \(2010\)](#) describes a spectrum of information a credit card company holds on its accounts. In one end of the spectrum, only the status (default/not-default) of an account is known, but in the other end, everything is known about the borrowers and the lenders can use this to optimise profits. Consider a scenario

where there is much knowledge about good accounts, but scarce data on bad accounts. The gains from good accounts can be modelled in a reliable way, but the probabilities of default and losses from bad accounts are difficult to model. In such a scenario, both probability scores (3.1) and profitability scores 3.3 will give unreliable results. In the following, a new scoring function is designed to balance estimated probabilities and expected gains in a way which makes it less sensitive to poor probability estimates than both (3.1) and (3.3).

The scatter plots in Figure 3.1 show the general idea behind the scoring function. Each

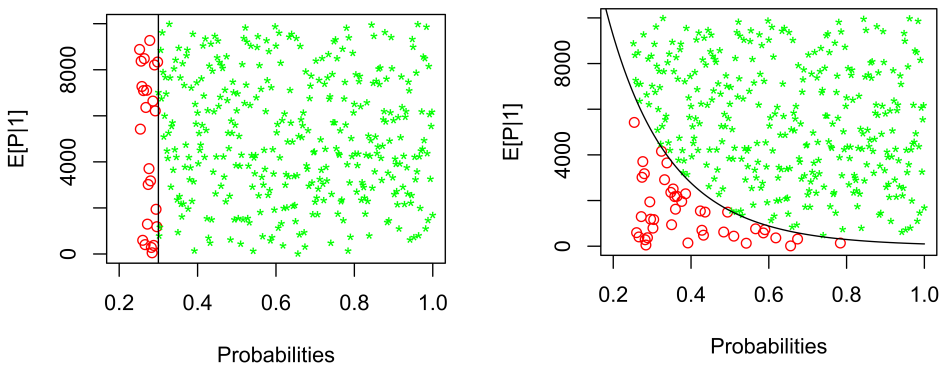


Figure 3.1: Two scatter plots showing a set of accounts in a simulated scenario. The x-axis shows probabilities that the account will be a non-defaulter and the y-axis shows the expected profits given that the account becomes a non-defaulter. The left plot has a one-dimensional cut-off while the plot to the right has a cut-off which balances profitability and probability. In both types of cut-off lines, red circles and green stars represent rejected and accepted applications, respectively.

point in the scatter plots shows the position of a credit card account in the risk-reward plane. The x-axis shows the risk associated to each account, namely the probability that the account becomes a case 1 account. The y-axis shows the potential reward corresponding to each account, namely the expected profit conditional on the account being good. The lines indicate cut-off values. Red circles belong to rejected applications while the green stars indicate that the application was accepted. The scatter plot to the left in Figure 3.1 shows

the conventional credit scoring scenario, where some probability of the application being good has been estimated. The cut-off is selected without consideration to the expected gains. The scatter plot to the right shows how a cut-off line which balances risk and reward can be constructed. Accounts with low probability of becoming good are accepted if they represent a high expected profit if they become good accounts.

Two dimensional lines such as the one shown in the right plot of Figure 3.1 need to be subjected to restrictions and follow certain parametrisations. The lines should be decreasing as the probabilities increase. If the estimated probabilities accurately manage to predict that an account will default, then the lines should grow large as the probabilities go to zero. There are many types of such parametrisations. For instance,  $G \sim S^{-\pi_1}$  or  $G \sim \frac{S}{\pi_1 - \beta}$ , where  $S$  and  $\beta$  are parameters that control the shape of the lines. Another example is shown in Figure 3.2 where the curves follow the parametrisation

$$G = G_0 \exp(S \cdot \pi_1^\nu). \quad (3.4)$$

Here,  $G_0$  is a scaling parameter while  $\nu$  and  $S$  are parameters that control the curvature and position of the line in the plane. In Figure 3.2,  $G_0 = 1000$  and  $\nu = 0.35$  for both curves. In the lower curve (A)  $S = -0.5$  while in the upper curve (B)  $S = -0.4$ . The points 1 and 2 both lie on curve A thus they are assumed to be equally valuable accounts. Points 3 and 4 lie on curve B and are also equally valuable, but more valuable to the company than 1 and 2. The cut-off lines should be decreasing functions of the probability. As account 5 has equal estimated gain as 3, but higher probability of becoming good, then it should be included in the portfolio whenever 3 is included. Conversely, accounts 4 and 6 have the same probability of becoming good, but account 6 has higher estimated profit. Therefore, 6 should always be included whenever 4 is included.

A scoring function which balances risk and reward should be a mapping from the risk-reward plane into the real line such that accounts that are more worth to the company get a higher score. With such a mapping, the scores are univariate and the techniques for cut-off

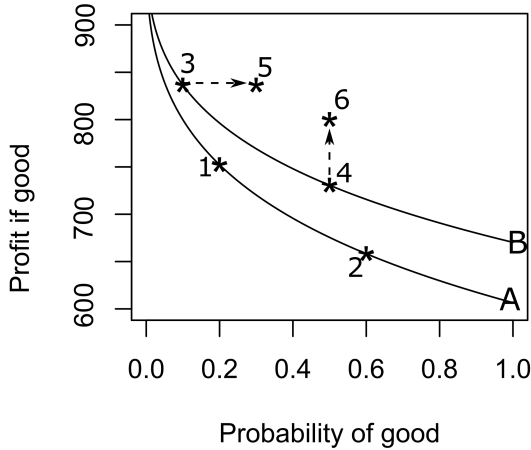


Figure 3.2: Two cut-off lines in the risk-reward plane. The lines follow the parametrisation shown in equation (3.4). For both lines,  $G_0 = 1000$ ,  $\nu = 0.35$  and for line A,  $S = -0.5$  while for line B,  $S = -0.4$ . Points 1-6 show the position of six accounts in the risk-reward plane. Accounts that lie on the same line are considered equally valuable to a credit card company. The cut-off lines should be decreasing functions of the probability. Account 5 should be accepted whenever 3 is accepted and account 6 should be accepted whenever 4 is accepted.

selection simplifies. For fixed values of  $G_0$ ,  $\nu$  and  $S$ , the curves (3.4) are "iso-worth" lines. In other words, each of the accounts with coordinates in the plane that hit this line are assumed to be of the same worth to the credit card company. By fixing the values of  $G_0$  and  $\nu$ , the position of the curves can be controlled by  $S$ . For account  $i$ , let  $S$  be the risk-reward score

$$S_{RR}(\mathbf{x}) = \frac{\log(\hat{G}^i) - \log(G_0)}{(\hat{\pi}_1^i)^\nu}, \quad (3.5)$$

where  $\hat{G}^i$  and  $\hat{\pi}_1^i$  are estimated using  $\mathbf{x}$ . The parametrisation (3.4) allows for many types of cut-off lines and different sets of parameters  $G_0$  and  $\nu$  provide different spectra of scores. The optimal values of  $G_0$ ,  $\nu$  and cut-off  $S$  can be found e.g. via a grid-search in which the best cut-off value in terms of portfolio profit for combinations of  $G_0$  and  $\nu$  is calculated.

There are certain limitations to which values the parameters can take. As the cut-off line  $G(\pi; G_0, S, \nu)$  should be a decreasing function of the probabilities, the derivative of  $G$  with respect to  $\pi$ ,

$$\frac{\partial G}{\partial \pi} = G_0 \exp(S\pi^\nu) \cdot S\nu\pi^{\nu-1},$$

should be negative. For positive  $\nu$  and  $\pi$ , this happens when  $S < 0$ . Considering the scoring function (3.5), negative scores imply that  $G_0$  must be greater than all estimated  $\hat{G}^i$ . A final constraint is that the scores should increase as the probabilities increase. In other words, the derivative of the scoring function with respect to probabilities  $\pi$ ,

$$\frac{\partial S}{\partial \pi} = \frac{\log(G) - \log(G_0)}{\pi^{\nu+1}} \cdot (-\nu),$$

should be positive and letting  $\nu$  be positive is a safe choice to make.

With parameters fulfilling these constraints, accounts where  $S(\mathbf{x}) \ll 0$  either have low expected gain, low probability of becoming good or both. Increase in score can happen if probability comes closer to 1 or if the expected gain increases. For fixed values of  $G_0$  and  $\nu$ , the score of each application can be calculated and a cut-off  $\tilde{S}$  can be chosen. All accounts with risk-reward coordinates above and to the right of the line parametrised by  $G_0$ ,  $\nu$  and  $\tilde{S}$  are accepted.

## 3.2 Setting a cut-off value

Most previous work on credit scoring deals with creating credit scoring methods and the performance of each method is given in how well  $F_0(t)$  and  $F_1(t)$  are separated. For a credit card company, it is also necessary to set an appropriate cut-off. In the following, it is assumed that the cut-off selection is a strict accept/reject decision. Even though it is common that borderline cases are evaluated by credit card underwriters, all decisions regarding credit granting in this thesis are made by binary decision rules.



### 3.2.1 Educated guesses of cut-off

If the score is based on probability of not defaulting,  $S_{\text{Prob}}$ , and the mean losses  $L$  and mean gains  $G$  are known, the expected profit of account  $i$  conditional on eventual status is  $E[P^i] = L(1 - \hat{\pi}_1^i) + G\hat{\pi}_1^i$ . As it is desirable to accept accounts with expected profits greater than 0, the credit card company should only accept applications with estimated probabilities

$$\hat{\pi}_1^i \geq \frac{L}{L+G}. \quad (3.6)$$

This cut-off represents a simplified scenario, in which the distributions of losses and gains are disregarded.

In a similar manner, if the score is the expected profit conditional on eventual status,  $S_{\text{Prof}}$ , and  $\hat{L}^i$ ,  $\hat{G}^i$  and  $\hat{\pi}_1^i$  are estimated, then accepting all applications with scores greater than zero would be a logical choice.

### 3.2.2 Accuracy

In a setting where the score is based on the probability of default without knowledge about distributions of profit, then *accuracy* is an indicator of the performance of the binary classification rule. Accuracy is defined as

$$\text{Accuracy}(t) = \pi_0 F_0(t) + \pi_1 (1 - F_1(t)),$$

which is the sum of the ratio of correctly rejected case 0 accounts and correctly accepted case 1 accounts. An idea is to set the cut-off by maximising the accuracy of the binary accept/reject model. Accepting all applications leads to an accuracy of  $\pi_1$ . Rejecting all applications leads to an accuracy of  $\pi_0$ . The accuracy of a certain cut-off does not account for the fact that there are differences in the cost of accepting case 0 accounts and rejecting case 1 accounts. Although accuracy has previously been used as an assessment method in profit scoring methods (Bravo et al., 2013; Verbraken et al., 2014), the availability of obser-

vations of profits enables more sophisticated cut-off selections.

### 3.2.3 Efficiency curves

Efficiency curves are analogous to accuracy, but in terms of profits and losses. With direct observations of profits and losses, it is possible to identify the cut-off which maximises the profit among accepted accounts. [Oliver and Wells \(2001\)](#) show how to create and interpret efficiency curves. An efficiency curve is a line parametrised by different cut-off values. For each cut-off, it is possible to calculate the total losses and total profits among accounts with scores higher than the cut-off. Losses are the amount of money lost to bad payers and the total profit is the sum of gains from good payers and losses from bad. The efficiency curve is the plot of  $(Profits(t), Losses(t))$  for each cut-off  $t$ . For readability, the axes are scaled by the total profits and losses, respectively. An example of two efficiency curves are shown in [Figure 3.3](#). The point A at (1,1) shows the scenario if all applications are accepted. The solid line stems from a perfect scoring function. Point C corresponds to a cut-off which excludes all bad accounts and the total profits are thus the sum of gains from all good accounts. The accuracy of the cut-off at C is 1. The dotted line comes from a more realistic scoring function and point B marks the maximum profit which is possible to obtain with such scores. Moving from point A to point B would imply that the total losses in the portfolio is reduced by more than 60% and the total profits increased by over 20%.

A good trait about efficiency curves is that they give an overview of the trends in the data sets. The sensitivity of choice of cut-off can thus be investigated. As the cut-off values chosen from efficiency curves balance the observed profits and losses, they are more informative than calculating accuracy. It is possible to accept a reduction in accuracy if the expected profits are higher.

A more recent method of using the distribution of losses and gains to set a cut-off value is introduced by [Verbraken et al. \(2013, 2014\)](#).

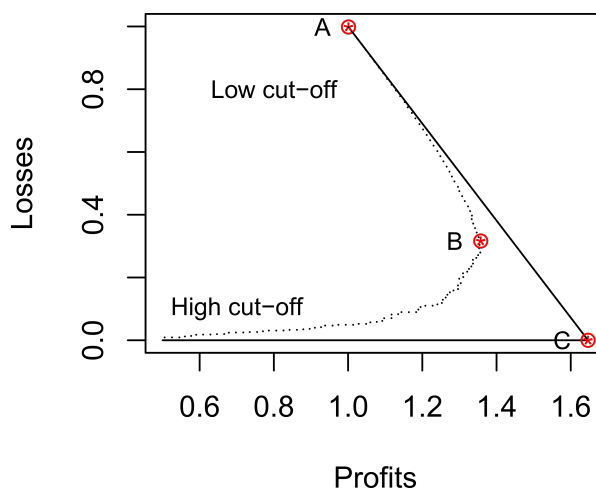


Figure 3.3: An example of a perfect efficiency curve (solid line) and a weaker efficiency curve (dotted line). The perfect curve stems from a scoring function which manages to perfectly separate good and bad accounts, while the weaker curve comes from a non-perfect scoring function. Point A indicates the situation where all applications are accepted. Points B and C indicate the efficient frontier of the two curves. These points correspond to the cut-off values which maximise the possible portfolio profit.

### 3.2.4 Expected Maximal Profit (EMP)

Verbraken et al. (2013, 2014) introduce EMP as an assessment tool of scoring functions and as a method of setting a cut-off value. In the EMP setting, there is knowledge about distribution of losses and gains in the portfolio. The goal is to obtain a reliable assessment criterion for credit scoring methods and a more generalised way of choosing an optimal cut-off. Although the EMP procedure was originally applied to credit scoring in ordinary bank lending and in customer churn analysis, it is possible to redevelop the theory to credit card scoring methods. The following discussion will first give an introduce EMP in a credit card scoring setting and show some additional traits that were omitted from the original articles.

Originally, Verbraken et al. (2014) modelled the money saved by rejecting some accounts. The following deduction of the EMP measure leans on the more intuitive starting

point to measure profits of accepted accounts. The idea is to model the costs and benefits of predicted non-defaulters and choose the cut-off which maximises the profit among accepted accounts. For a fixed cut-off value  $t$ , the expected profit of account  $i$  is

$$E_{\text{scores}}[P^i] = E[P^i | s^i < t] \cdot \text{Prob}(s^i < t) + E[P^i | s^i \geq t] \cdot \text{Prob}(s^i \geq t)$$

where  $s^i$  is the score of account  $i$ . As the profits from a rejected account is 0, the expected profit may be simplified by removing the corresponding term. Among the accepted accounts, some accounts might be bad and some might be good, depending on the quality of the scoring function. A double expectation over distributions of scores and the states 0/1 leads to

$$E_{\text{state}}[E_{\text{scores}}[P^i]] = E[P^i | s^i \geq t, 0] \cdot \text{Prob}(s^i \geq t | 0)\pi_0 + E[P^i | s^i \geq t, 1] \cdot \text{Prob}(s^i \geq t | 1)\pi_1$$

The probabilities of a score being higher or lower than a cut-off  $t$  conditional on eventual status are already known as  $F_0(t)$  and  $F_1(t)$ . Let the expected gain of an account be  $c$  and expected loss be  $b$ , then the double expectation can be written as

$$\mathbf{P}(t; b, c) = -b\pi_0 + c\pi_1 + bF_0(t)\pi_0 - cF_1(t)\pi_1 \quad (3.7)$$

where  $\mathbf{P}$  in bold face denotes the expected profit of the account conditional on its score being above or below  $t$  and conditional on the eventual status of the account. The optimal cut-off value will be

$$T = \underset{t}{\text{argmax}} \mathbf{P}(t; b, c).$$

Differentiating  $\mathbf{P}(t; b, c)$  with respect to  $t$  and equating to zero gives a first order condition which needs to be satisfied at cut-off  $T$ ;

$$\frac{\partial \mathbf{P}(t; b, c)}{\partial t} \Big|_{t=T} = 0 \iff \frac{f_0(T)}{f_1(T)} = \frac{\pi_1}{\pi_0} \cdot \frac{c}{b}. \quad (3.8)$$

The first order condition thus gives an implicit connection between values of gains and losses and a point on the ROC-curve. For different ratios of gains to losses, the optimal cut-off will vary, but it will always correspond to the slope somewhere on the ROC-curve.

Knowing the optimal cut-off value also gives the optimal fraction of accounts that should be rejected. If probabilities are interpreted as ratios, then for a fixed combination of  $b$  and  $c$ ,

$$\eta = \pi_0 F_0(T) + \pi_1 F_1(T). \quad (3.9)$$

$\eta$  can be rewritten as  $\eta = \text{Prob}(s < T \cap 0) + \text{Prob}(s < T \cap 1)$  and can thus be interpreted as  $\eta = \frac{1}{N}(\text{Number of bad with scores } s < T + \text{Number of good with scores } s < T)$

The framework opens for treating loss and gain as random variables. The intuition behind  $b$  and  $c$  is the unit loss and unit gain that can be expected from any account. In a credit card setting, this can be translated to the ratios losses divided by credit limit and gains divided by credit limit. In the general setting, let the joint distribution of  $b$  and  $c$  be  $h(b, c)$  such that the classification profit averaged over all possible  $b$  and  $c$  is

$$\text{EMP} = -E[b]\pi_0 + E[c]\pi_1 + \int_b \int_c [bF_0(T(b, c))\pi_0 - cF_1(T(b, c))\pi_1] h(b, c) dc db.$$

The first two elements represent the scenario where all applications are accepted. The expected profit is the sum of expected loss times the ratio of bad accounts and expected gains times the ratio of good accounts. The double integral models the improvements that can be made by setting optimal cut-off values. The end product of [Verbraken et al. \(2014\)](#) is the double integral without the first two constants. Thus, it is apparent that maximising the expected profits among accepted accounts is the same as minimising the expected profits among rejected accounts. However, the starting point used in this thesis might be a more intuitive one, at least in a business context.

The expected ratio of accounts that should be rejected follows from equation (3.9),

$$\eta_{\text{EMP}} = \int_b \int_c [\pi_0 F_0(T(b, c)) + \pi_1 F_1(T(b, c))] \cdot h(b, c) dc db. \quad (3.10)$$

This expression is the same as given by Verbraken et al. (2014) and implies that the optimal portfolio of accepted accounts consists of the sorted scores  $s_{(l)}, s_{(l+1)}, \dots, s_{(N)}$ , where  $l$  is the smallest integer such that  $l \geq N\eta_{\text{EMP}}$  and  $N$  is the total number of applications.

As the data sets used in this thesis contain direct observations of profit, there will not be need for EMP as an assessment method. The further relevance of the discussion above is the calculation of optimal reject ratios according to equation (3.10).

### Keeping $c$ constant

In regular bank loans,  $c$  is interpreted as the return on investment (ROI) which a bank receives when giving out loans. The ROI does not vary a lot due to rules regarding loan maturity and fixed instalments. Assuming  $c$  is constant leads to a simplification of the double integral (3.10) into a simple integral over possible values of  $b$ .  $b$  can be interpreted as the fraction of total loan amount which is lost after a default. In other words, it is the money a defaulter owes the bank divided by the original loan size. Based on a set of real loan data, Verbraken et al. estimate the distribution of  $b$  to be bimodal with peaks at 0 and 1, and uniform in between;

$$h(b) = \begin{cases} p_0, & \text{for } b = 0 \\ 1 - p_0 - p_1, & \text{for } b \in (0, 1) \\ p_1, & \text{for } b = 1 \end{cases}$$

$p_0$  is the probability that a defaulter pays back the entire loan.  $p_1$  is the probability that a bad account defaults on the entire credit limit or more.  $p_1$  is easily estimated as the fraction of defaulted accounts that have losses higher than or equal to their credit limit.

Having established this simplified scenario, Verbraken et al. moves on to suggest a nu-

merical method to calculate the fraction of denied applications (3.10). This method relies on the convex hull of a Receiver Operating Characteristic (ROC) curve, denoted ROCCH. Fawcett (2006) argues that the optimal operational scenarios are found on the ROCCH as these are scenarios with minimal operational costs. Therefore, the search for optimal cut-off values are restricted to the ROCCH. Assume that the convex hull consists of  $m$  straight lines and  $(r_{1i}, r_{0i})$  is the end point of segment  $i$  ( $i = 1, \dots, m$ ). As the points on the convex hull are optimal, equation (3.8) gives the link between  $b$ ,  $c$  and the optimal cut-off score  $s \in [r_{1i}, r_{1i+1}]$ .

$$b_{i+1} = \frac{\pi_1(r_{1(i+1)} - r_{1i})}{\pi_0(r_{0(i+1)} - r_{0i})} \cdot c$$

$b_0$  is defined to be 0. The maximum value that  $b$  can take is 1. Therefore, the integration over  $b$  only covers the optimal cut-off values corresponding to  $\{b_i; i = 0, \dots, k+1\}$  where  $k = \max(i; b_i < 1)$ .  $b_{k+1}$  is defined to be 1. The integral (3.10) can then be estimated;

$$\eta_{\text{EMP}_b} = [\pi_0 r_{00} + \pi_1 r_{10}] p_0 + [\pi_0 r_{0(k+1)} + \pi_1 r_{1(k+1)}] p_1 + \sum_{i=0}^k \int_{b_i}^{b_{i+1}} [\pi_0 r_{0i} + \pi_1 r_{1i}] h(b) db$$

This method of calculating cut-off values will be called the  $\text{EMP}_b$  method, as a distribution of  $b$  is employed. The first two elements on the right side of the numerical scheme come from the probability point masses at  $b = 0$  and  $b = 1$ . The sum of integrals corresponds to summing the contribution at each  $k+1$  line segments. Observe also that the ROCCH starts in the origin, therefore  $r_{00}$  and  $r_{10}$  are both 0. The use of  $r_{0(k+1)}$  and  $r_{1(k+1)}$  as coordinates of the end-point of the line segment corresponding to  $b_{k+1} = 1$  is misleading. There might be large differences between the end point of the line segment corresponding to  $b_k$  and the next end point. In the implementation,  $r_{0(k+1)}$  and  $r_{1(k+1)}$  are simply taken to be equal to the previous end point  $r_{0k}$  and  $r_{1k}$ . This points to a weakness in the model which occurs whenever the ROC curve is far from convex and  $k$  is a low number.

As a test of the numerical scheme proposed by Verbraken et al., it is interesting to investigate what happens if the scoring function provides perfect separation between defaulters

and non-defaulters. When the underlying scoring function is close to perfect, the ROC-curve consists of an almost vertical line from the origin to (0,1) and a horizontal line from (0,1) to (1,1). The coordinates of the ROCCH have the following form;

$$r_{1\cdot} = (0, \delta, 1) \quad \text{and} \quad r_{0\cdot} = (0, 1, 1)$$

where  $\delta \ll 1$ . All defaulters have scores lower than the lowest-scoring non-defaulter. With this ROCCH, the values of  $b$  that correspond to optimal cut-off values are calculated from the first order condition;  $b_i = (0, \frac{c\pi_1\delta}{\pi_0}, 1)$ , where  $b_2$  is set to be 1 as  $k = 1$ . Then,

$$\begin{aligned} \eta_{\text{EMP}b} &= [\delta\pi_1 + \pi_0]p_1 + \sum_{i=0}^1 \int_{b_i}^{b_{i+1}} [\pi_0 r_{0i} + \pi_1 r_{1i}](1 - p_0 - p_1) db \\ &= [\delta\pi_1 + \pi_0]p_1 + (1 - p_0 - p_1)(\pi_0 + \delta\pi_1)(1 - \frac{c\pi_1\delta}{\pi_0}) \end{aligned}$$

In the situation where  $\delta$  is sufficiently close to 0 such that the elements involving  $\delta$  can be neglected,  $\eta_{\text{EMP}} = (1 - p_0)\pi_0$ . This can be interpreted as the probability that a defaulter fails to pay back the entire loan times the probability that the account is a defaulter.

The second extreme event is that the scoring function completely fails to separate the defaulters from non-defaulters. In which case, the ROC-curve is a straight line from the origin to (1,1) and  $\delta = 1$ . In this case, possible optimal values of  $b$  are  $b_i = (0, 1)$  when  $\frac{c\pi_1}{\pi_0} \geq 1$ . Then, all the elements of  $\eta_{\text{EMP}}$  collapse to 0 as only known end points are in the origin. Although it is intuitively correct that no accounts should be rejected in such scenarios, this also points to a weakness in the method. If  $\frac{c\pi_1}{\pi_0} < 1$ , then the discrete values  $b$  will take are  $b_i = (0, \frac{c\pi_1}{\pi_0}, 1)$  and the ratio of rejected accounts will be greater than 0. In such cases, the cut-off will lead to sub-optimal cut-off values.

The article by [Verbraken et al. \(2014\)](#) was written to suit bank loans and the assumption of a constant  $c$  is not necessarily precise among credit card accounts. Considering the upper tail of Figures 2.3 and 2.5, there are large variations in profit among good accounts. This motivates an alteration to the EMP in which  $c$  is given a distribution.



### Varying both $b$ and $c$

Having direct observations of profitability opens for utilising distributions of both  $b$  and  $c$ . Verbraken et al. (2014) include a sensitivity study in which it is concluded that perturbing  $c$  in the region  $\pm 50\%$  does not have a large impact on the calculated cut-off. As the observed gains divided by granted credit in the data sets vary more than this, it is suitable to consider more extreme perturbations as well. Assume  $b$  and  $c$  are independent such that  $h(b, c) = h_1(b)h_2(c)$ . This implies that the amount lost by a bad account is independent from the gain represented by the account if it was good. In situations where losses and gains are correlated, this assumption might fail. However, assuming  $b$  and  $c$  are independent leads to

$$\begin{aligned}\eta_{\text{EMP}bc} &= \int_c \left[ \int_b [\pi_0 F_0(T(b, c)) + \pi_1 F_1(T(b, c))] h_1(b) db \right] h_2(c) dc \\ &= E_c[\eta_{\text{EMP}b}],\end{aligned}$$

the expected value of  $\eta_{\text{EMP}b}(c)$  as a function of  $c$ . In general,  $E_c[\eta_{\text{EMP}b}(c)]$  is unequal to  $\eta_{\text{EMP}b}(E[c])$ .  $\eta_{\text{EMP}bc}$  therefore needs to be estimated via numerical integration or Monte Carlo methods. This method of calculating a cut-off will be denoted the  $\text{EMP}_{bc}$  to indicate that both  $b$  and  $c$  follow some distribution.

The exponential decay of gains divided by credit limit has a long tail and a Weibull distribution with shape parameter smaller than 1 might thus be suitable. However, as this Weibull density grows to infinity when  $c$  goes to 0, the weights given to low values of  $c$  are too high. Since a large part of observed gains are only slightly positive and as the numerical scheme fails to provide optimal cut-off values when ROC is close to linear and  $c$  is small, it is dangerous to give too much weight to such values of  $c$ . An alternative is to assume that  $c$  is exponentially distributed. However, an exponential density function has too much mass in the lower region and thus the tails are underestimated. A mixture distribution in which the distribution function of  $c$  is a weighted sum of a Weibull density with shape parame-

ter greater than 1 and an exponential density can balance long tails and reasonably high probability mass for low  $c$ .

# Chapter 4

## Results

The risk-reward scoring function and the EMP cut-off selection method are compared to the conventional methods in two ways. First all methods are applied in a case study of Norwegian credit card accounts. Thereafter, a series of sensitivity analyses are described to highlight some traits of the new methods.

### 4.1 Results from case study

Models for losses, gains and probabilities are fitted on a training set which consists of a randomly chosen subset of each data set. The number of observations in the training sets correspond to 75% of all applications in data sets A and B. The fitted values of losses, gains and probabilities are used as input in the scoring functions to calculate scores for the applications in the training set. Cut-off values are calculated from the scores in the training set. Afterwards, the scoring functions and cut-off values are applied to the unseen testing set consisting of the remaining 25% of applications in both data sets.

Forward Stepwise Regression with Backwards Elimination(FSRBE) was applied as the Generalised Linear Model (GLM) selection procedure. Testing hierarchical models within

the FSRBE was done using chi square tests for the difference in deviance between small and large models. In each iteration, the variable which gave the lowest p-value in the chi-square test was added to the model, provided that this p-value was less than 0.1. Backwards elimination was also performed with p-value thresholds of 0.1. A thorough description of the FSRBE is given by [Baesens \(2014\)](#). Pearson residuals are used to assess each model. Pearson residuals are the differences between fitted and observed values, scaled by the standard deviation of the estimates. The standard deviations are calculated according to the assumed distribution of the observed values.

#### 4.1.1 Probability of not defaulting

Logistic models for estimating each account's probability of becoming a non-defaulter were built using logit-links.

In data set A, 12 variables are included in the model. Four of the chosen variables are continuous. Among the eight grouped variables, there are a total of 35 levels in addition to the base levels. 21 levels have significantly non-zero fitted coefficients. The percentwise reduction in deviance from the null model to the fitted model is 2.6% at the cost of 39 degrees of freedom. [Figure 4.1](#) shows Pearson residuals of the fitted values plotted against the fitted values.

The residuals of case 0 accounts are negative while the residuals of case 1 accounts are positive. In [Figure 4.1](#), all fitted probabilities are greater than 0.7 and probabilities fitted to case 0 accounts are greater than 0.8. The variance of the residuals increase as the fitted probabilities increase. This indicates that there is a class imbalance between the cases. Limiting a class imbalance problem was attempted by under- and over-sampling case 1 and case 0 accounts respectively. However, the resulting models failed to improve the fits when applied to the testing set. This was mainly due to loss of relevant observations when under-sampling and overfitting case 0 accounts when over-sampling. The small reduction in deviance and the non-constant variance in residuals indicate that the estimated proba-

bilities will give scoring functions with poor predictive powers.

The model fitted to data set B contains an intercept and 18 variables of which four are continuous. Among the grouped variables there are 60 levels of which 25 levels have significantly non-zero fitted coefficients. The percentwise reduction in deviance from the null model to the fitted model is 6.5% at the cost of 64 degrees of freedom. Figure 4.2 shows Pearson residuals plotted against the fitted probabilities.

Comparing the residuals in Figure 4.2 to Figure 4.1, there are some indications that the fitted probabilities in data set B are more accurate. The fitted probabilities span a wider range in data set B, indicating that the class imbalance problem is not as grave as in set A. Also, the ratio of case 0 to case 1 accounts among the lower region of estimated probabilities is higher. These observations and the larger reduction in deviance, indicate that the scoring functions based on probability estimates will be more reliable in data set B than in A.

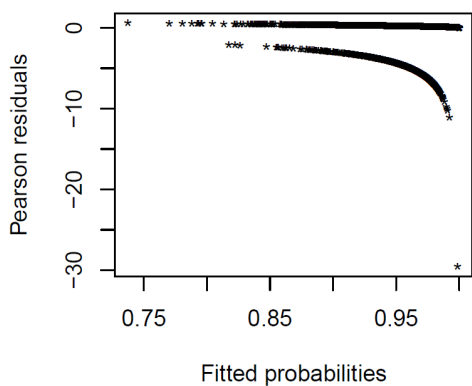


Figure 4.1: Pearson residuals of fitted probabilities in data set A.

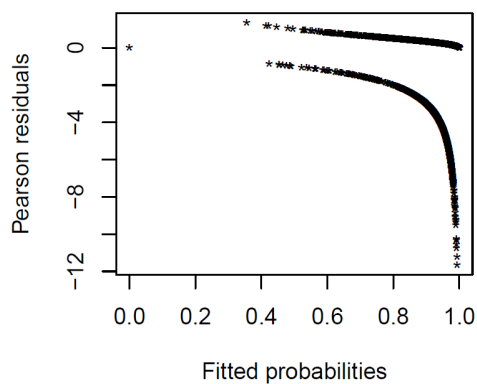


Figure 4.2: Pearson residuals of fitted probabilities in data set B.

#### 4.1.2 Gain given case 1

Negative binomial models with log link and credit limit as offset were fitted to the observed gains in both data sets. In both models for data set A and B, the addition of a dispersion

parameter significantly reduces the deviance of the models.

The model fitted to data set A contains nine variables in addition to the intercept. Two of the included variables are continuous. Among the grouped variables, there are 48 different levels of which 39 levels have significantly non-zero coefficients. The percentwise reduction in deviance from the null-model to the fitted model is 11% using 50 degrees of freedom. Figure 4.3 shows the Pearson residuals plotted against estimated gains for the resulting model.

The variance in the Pearson residuals increases as the estimates decrease. This means that the model fits too high gains to accounts where the observed gain is small. Attempts were made to reduce this problem by accounting for zero inflation. Inactive credit card users seem to appear randomly and thus the models for zero-inflation could not improve the negative binomial model.

In data set B, 10 variables of which one is continuous are included in the model. Among the grouped variables, there are 43 levels of which 29 levels have significantly non-zero fitted coefficients. The percentwise reduction in deviance from the null model to the fitted model was 13% using 44 degrees of freedom. Figure 4.4 shows the Pearson residuals plotted against estimated gains. The Pearson residuals follow a similar pattern as in data set A, but the variance in residuals is slightly more stable in set B. As for set A, attempts to model zero-inflation did not improve the models.

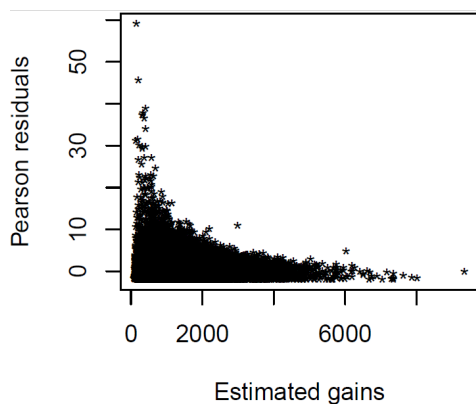


Figure 4.3: Pearson residuals plotted against the fitted gains in data set A.

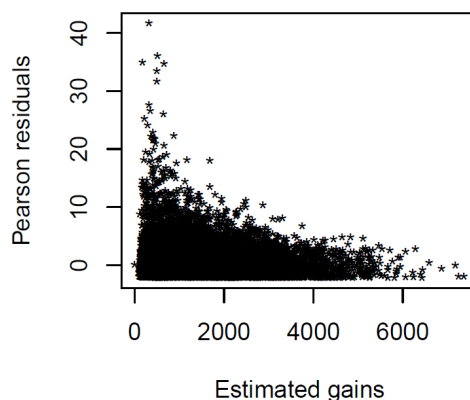


Figure 4.4: Pearson residuals plotted against the fitted gains in data set B.

### 4.1.3 Loss given case 0

Binomial models with log link were fitted to the ratio losses per credit limit rounded to 0 or 1. This is a similar interpretation as the unit loss  $b$  in the EMP-setting. To obtain estimates of losses, the fitted ratios were multiplied with credit limit.

The model fitted to data set A contains intercept and 9 variables, of which three are continuous. Among the grouped variables, there are 27 different levels of which 13 levels have significantly non-zero coefficients. The percentwise reduction in deviance of the fitted model from the null model is 27% at the cost of 30 degrees of freedom. Figure 4.5 shows Pearson residuals of the loss estimates plotted against the fitted losses. Large residuals only appear among low estimated losses, where the estimated loss divided by granted credit is too low. The residuals for large values of estimated losses are small. This means that the procedure manages to encapsulate the heavy tails among observed losses. The variability in Pearson residuals is more stable than for the estimated gains.

The model for losses in data set B is built in the same way as in data set A. There are nine variables in the model, of which two are continuous. Among the grouped variables, there are 23 levels of which 12 levels have significantly non-zero coefficients. The percent-

wise reduction in deviance of the fitted model from the null model is 23% at the cost of 25 degrees of freedom. Figure 4.6 shows Pearson residuals plotted against estimated losses in both data sets. The residuals seem to be more stable in the fitted values belonging to set B than A.

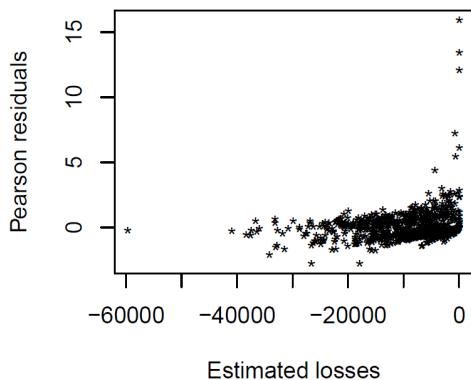


Figure 4.5: Pearson residuals plotted against the fitted losses in data set A.

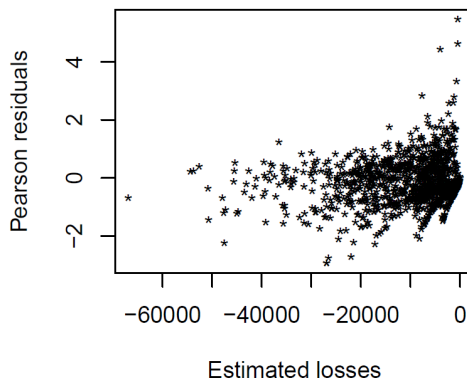


Figure 4.6: Pearson residuals plotted against the fitted losses in data set B.

#### 4.1.4 Applying scoring functions and cut-off

The fitted models for probabilities, losses and gains are combined to create scores for the applications in both training and testing sets.

Figure 4.7 shows the efficiency curves stemming from the scoring functions applied to the training and testing subsets of data set A. The shapes of the efficiency curves reflect the lack of fits in the estimated gains and probabilities of data set A. Letting the cut-off correspond to the most profitable point on the efficiency curve in the training set gives a portfolio with profits slightly greater than accepting all applications. However, since the shapes of the efficiency curves are dissimilar in the training and testing set, the optimal cut-off fails to produce a profitable portfolio in the testing set. The fitted models fail to predict the probabilities, losses and gains in the testing set. Consequently, there are no



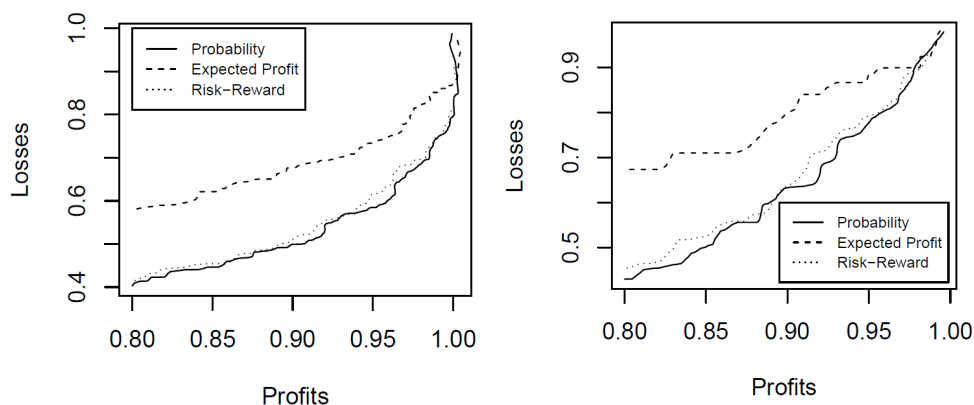


Figure 4.7: Efficiency curves corresponding to scoring functions  $S_{\text{Prob}}$  for probability estimates,  $S_{\text{Prof}}$  for estimates of expected profit and  $S_{\text{RR}}$  for risk-reward score in training set (left plot) and testing set(right plot) of set A.

cut-off values that produce portfolios with greater profits than accepting all applications. Accepting all applicants also leads to maximum accuracy.

In both plots of Figure 4.7, the curves corresponding to risk-reward scores and probability scores follow each other closely. The added consideration to fitted gains in the risk-reward scores does not improve the scoring function. In a situation where models of gain are bad fits and the estimates are misleading, the new risk-reward scoring method can't outperform the probability-based scores.

Although there are no profitable cut-off values, the scoring functions still manage to sort the accounts in terms of worth. The probability scores and risk-reward scores give a streamlining of the portfolio which makes it possible to reduce the ratio of losses to total profits. For instance, one can choose a probability cut-off which reduces the total amount of losses by almost 60% while the total amount of profits is reduced by only 20%.

Figure 4.8 shows the efficiency curves stemming from the spectra of scores in the training and testing subsets of data set B. All the credit scoring methods can give portfolios with higher profits than accepting all applications. The profitability score,  $S_{\text{Prof}}$ , gives the scores with highest potential profits in the training set.

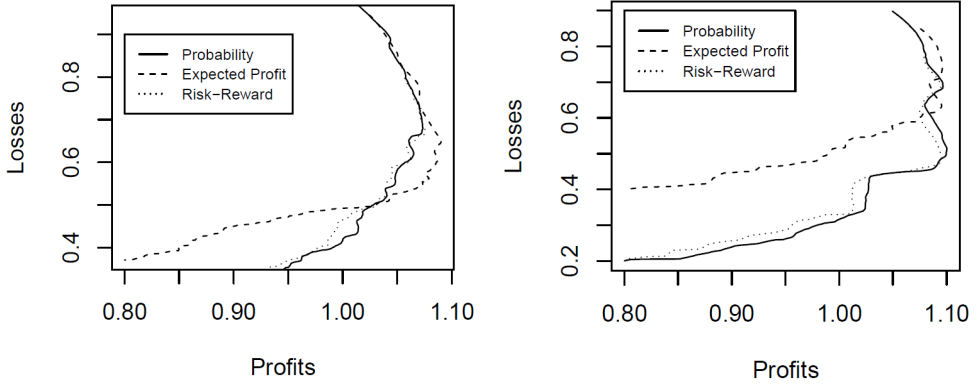


Figure 4.8: Efficiency curves corresponding to scoring functions  $S_{\text{Prob}}$  for probability estimates,  $S_{\text{Prof}}$  for estimates of expected profit and  $S_{\text{RR}}$  for risk-reward score in training set (left plot) and testing set (right plot) of set B.

Figure 4.9 shows a section of the efficiency curves stemming from the profitability score as well as calculated cut-off values. All cut-off values and performance of the resulting portfolios in training and testing sets are shown in Tables 4.1 and 4.2.

Table 4.1: Cut-off values and resulting portfolios when letting the scores be estimated expected profit conditional on eventual status, training subset of data set B. Each cut-off value corresponds to a point in the left plot of Figure 4.9

Cut-off selection	Cut-off	Accuracy	Rejected		Accepted	
			N	Avg. profit	N	Total profit
Accuracy	-9370.1	94.3	22	-385115	21802	18329141
Efficiency Curve	-300.1	91.6	976	-1656328	20848	19600354
$\text{EMP}_b$	-1004.0	93.0	536	-1321606	21288	19265632
$\text{EMP}_{bc}$	-43.0	90.4	1312	-1541952	20510	19485978
Educated guess	0.0	90.2	1381	-1534923	20443	19478949

It is apparent from Figure 4.9 and Tables 4.1 and 4.2 that the cut-off which maximises accuracy is not necessarily the most profitable one. All the other cut-off methods give lower accuracy, but higher profits. By definition, the efficiency curve cut-off gives the most profitable portfolio in the training set, but this cut-off is not the best in the testing set. The

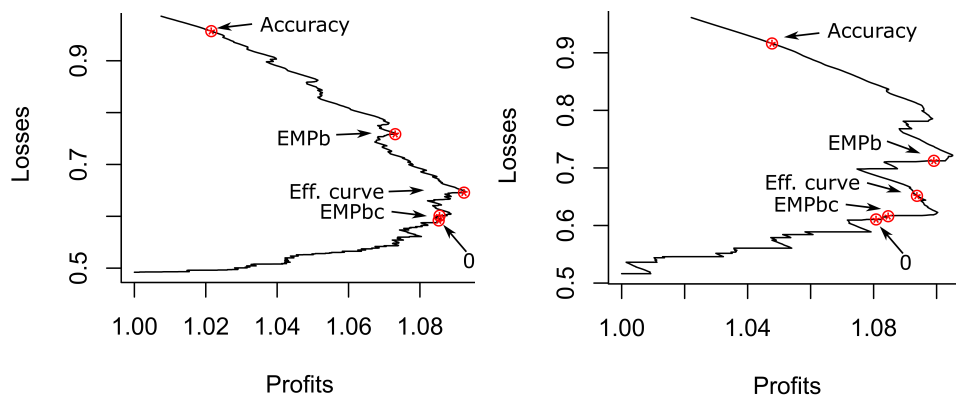


Figure 4.9: Efficiency curves corresponding to scoring functions  $S_{Prof}$  in the training set (left plot) and testing set(right plot) of set B. Cut-off values are indicated as points on the curves.  $EMP_b$  denotes the EMP cut-off calculated using constant  $c$ , while  $EMP_{bc}$  shows the cut-off calculated using an exponentially distributed  $c$ . Tables 4.1 and 4.2 show the corresponding cut-off values and resulting portfolios.

$EMP_b$  cut-off (with constant  $c$ ) gives the most profitable portfolio in the testing set. Efficiency curves are jagged and the efficiency curve cut-off identifies a peak on the profit axis. The EMP cut-off rely on the distributions of losses and gains and can therefore sometimes be more reliable than the efficiency curve cut-off. Another observation is that the educated guess of cut-off does not match the spectra of scores. There is some bias in the scores such that the most profitable cut-off values are negative.

This example of a scoring function and cut-off values shows that using the distributions of losses and profits in the selection of cut-off can give improvements to the educated guesses and maximum accuracy cut-off values.

Table 4.2: Cut-off values and resulting portfolios when letting the scores be estimated expected profit conditional on eventual status, testing subset of data set B. Each cut-off value corresponds to a point in the right plot of Figure 4.9

Cut-off selection	Cut-off	Accuracy	Rejected		Accepted	
			N	Total profit	N	Total prof.
Accuracy	-9370.1	94.2	9	-204040	7265	6008649
Efficiency Curve	-300.2	91.4	314	-542290	6960	6346899
EMP <sub>b</sub>	-1004.8	92.7	190	-584032	7084	6388642
EMP <sub>bc</sub>	-39.4	90.3	407	-498502	6867	6303112
Educated guess	0.0	90.0	437	-473362	6387	6277971

## 4.2 Sensitivity analysis of scoring functions

In a scenario such as for data set A, where one or more of the models for probabilities, gains and losses are bad fits, the resulting scores will fail to produce profitable cut-off values. In order to assess how the performance of the scoring functions reduce when the estimated probabilities deteriorate, a sensitivity analysis was performed. The procedure is summarised in the steps described below;

- Simulate 1000 accounts with a randomly drawn profit margin.
- Assign "estimates" of losses and gains to each account. The "estimates" are drawn from a normal distribution such that "estimated" losses were closer to zero among good accounts than among bad and "estimated" gains were closer to zero among bad accounts than among good.
- In a series of iterations:
  - Assign "estimates" of probabilities to each account such that the accuracy of the probabilities reduce in each iteration.
  - Calculate the Mean Squared Difference(MSD) between "estimates" of probabilities and the classes due to profit margins
  - Calculate the scores of each account according to  $S_{\text{Prob}}$ ,  $S_{\text{Prof}}$  and  $S_{\text{RR}}$

- Choose the most profitable cut-off from the resulting efficiency curves

Figure 4.10 shows the results of the sensitivity analysis. The x-axis shows the MSD calcu-

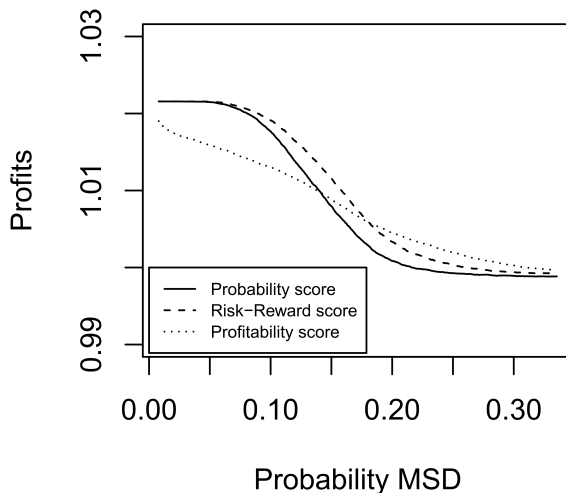


Figure 4.10: Sensitivity analysis of scoring functions  $S_{Prof}$ ,  $S_{Prob}$  and  $S_{RR}$  for probability estimates of varying quality. The estimates of losses and gains were fixed. The x-axis shows the mean squared difference between true classes and probability estimates. The y-axis shows the ratio of profit obtained by using the efficiency curve cut-off and the profit if no model was applied.

lated in each iteration and the y-axis shows profit from the optimal cut-off values scaled by the profit if all accounts were accepted. Rejecting only the defaulters would lead to a 2.16% increase in profits.

The dotted line in Figure 4.10 shows the performance of the profitability-based score. The estimated profit given eventual status hinges on probability estimates and is sensitive to errors in the estimates. As the estimates of gains and losses are good, the drop in performance is slow and steady. The solid line shows the performance of the probability score. When the mean squared difference reaches 0.25, the optimal choice of cut-off is to accept all accounts. This value of mean squared difference is special, as it occurs when all probability estimates are 0.5. The dashed line in Figure 4.10 comes from the risk-reward score.

Risk-reward scores are similar to the probability scores, but gives slightly more profitable portfolios.

Figure 4.11 shows a similar sensitivity analysis as displayed in Figure 4.10 in which the "estimates" of losses and probabilities are fixed while the "estimates" of gains vary. The

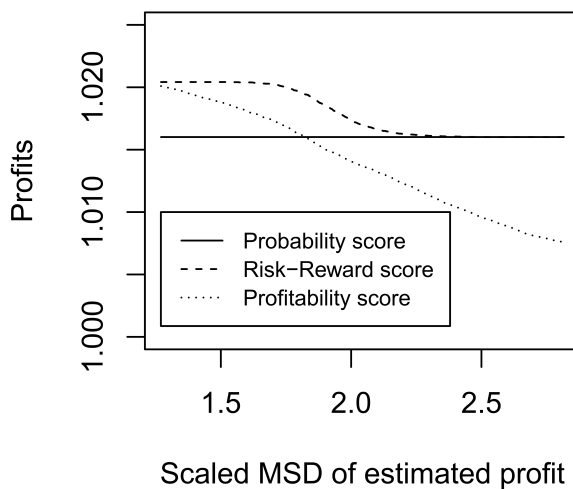


Figure 4.11: Sensitivity analysis of scoring functions  $S_{Prof}$ ,  $S_{Prob}$  and  $S_{RR}$  in which the quality of estimated gains varies. The expected profits,  $S_{Prof}$ , are used as a proxy of the quality of the gain estimates. The x-axis shows the mean squared difference between expected profits and observed profit margin scaled by the empirical variance of profit margins. The y-axis shows the ratio of profit among accepted due to best possible cut-off and the profit if all applicants were accepted.

x-axis in Figure 4.11 shows the MSD between the profitability score and the true profits scaled by the empirical variance of the profits. The profitability scoring function is thus used as an indicator of the quality of the gain estimates. This is why the performance of the profitability score, shown in dotted lines, seems to decrease linearly. It can also be seen how the risk-reward scoring function gives equally or more profitable portfolios than the probability score. As the estimates of gains deteriorate, the risk-reward scoring function reduces to the probability scoring function and they give equal cut-off values.

### 4.3 Comparing risk-reward scores and probability scores

The plots in Figure 4.12 show four different sets of simulated accounts plotted in the risk-reward plane. Case 0 accounts are indicated with red dots while case 1 accounts are indicated with blue circles. The predictive powers of the estimated gains and probabilities vary between the plots and thus the optimal cut-off lines also change. Solid lines indicate the most profitable two-dimensional cut-off in each scenario while the dashed lines indicate the optimal cut-off of probability score.

In the top left plot of Figure 4.12, estimated gains and estimated probabilities perfectly separate defaulters from non-defaulters. Here,  $G_0$  is chosen such that the cut-off line crosses the y-axis directly above the highest estimated gain among the case 1 accounts.  $\nu$  is close to 0 such that the line decreases slowly. As the estimated probabilities perfectly manage to separate case 0 from case 1 accounts, both cut-off methods produce portfolios consisting of all the non-defaulters.

In the top right plot, the estimated gains perfectly separate case 0 from case 1 accounts while the estimated probabilities fail to separate them. A cut-off based on estimates of probabilities will accept some defaulters, while the risk-reward cut-off only accepts non-defaulters.  $G_0$  and  $\nu$  are similar to those used in the top left plot.

In the lower left plot, the estimated probabilities perfectly separate the classes, but the estimated gains fail to do this. Both cut-off lines give perfect portfolios. Here,  $G_0$  is a very high number to account for a large uncertainty in the estimates of gain and  $\nu$  is close to 0.5 such that the cut-off line decreases steadily.

In the bottom right plot, both models of probability and profits fail to separate defaulters from non-defaulters. In such situations, the risk-reward scoring function might outperform the probability score. As can be seen from the plot, the risk-reward cut-off rejects a cluster of bad accounts that are accepted by the probability cut-off.

As there is more freedom in choosing a two-dimensional cut-off, the risk-reward scoring method will in general give equal or better performance than the probability cut-off.

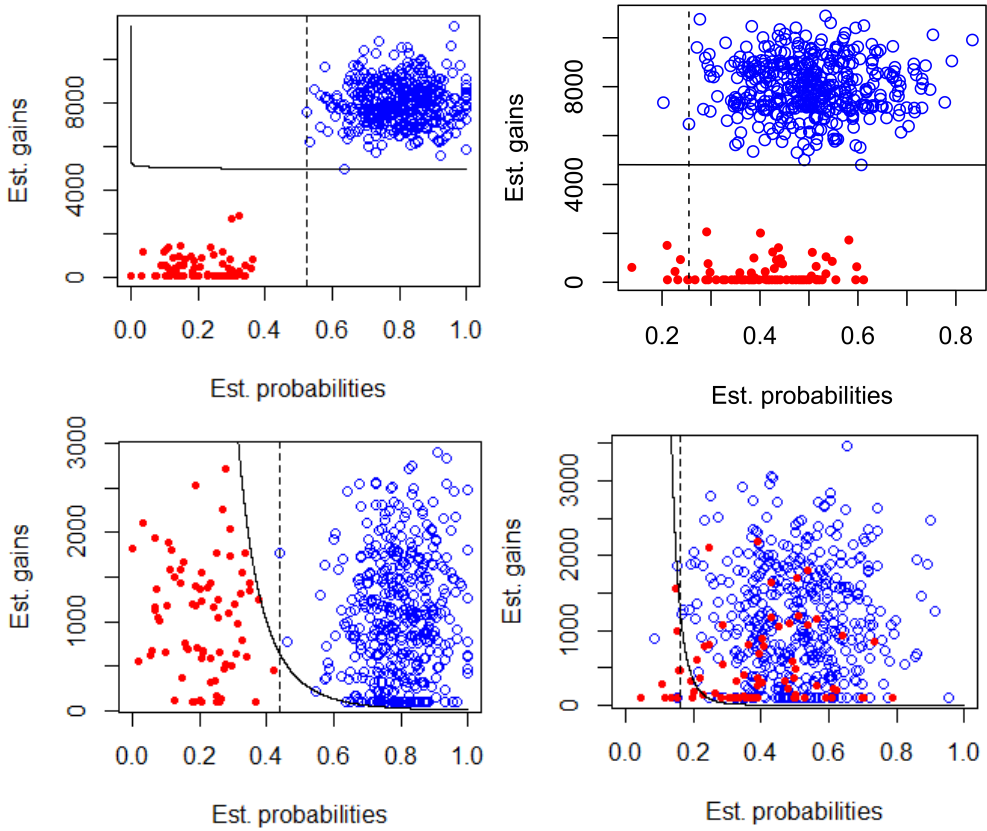


Figure 4.12: Four simulated scenarios in which estimates of expected gains and probabilities of not defaulting are varying in quality. Blue circles represent case 1 accounts while red dots represent case 0 accounts. In each scenario, the dashed line indicates the probability cut-off while the solid line indicates the two-dimensional cut-off. Each cut-off has been chosen such that they maximise profit in the resulting portfolio. The top left plot shows the scenario where both estimates of probability and gain give perfect separation of case 0 and 1 accounts. The top right plot shows probability estimates of low quality, but the estimated gains perfectly separate the classes. Lower left plot shows good probability estimates while the estimates of gains fail to separate the classes. The lower right plot shows the situation where both estimates of probabilities and gains fail to separate the classes.

The greatest advantage of using the risk-reward cut-off instead of the probability cut-off comes in situations where the estimates of probability are sub-optimal and there exist good



estimates of gains. A weakness in the method occurs when gains and losses are positively correlated such that accounts which seem to give high gains end up as large losses instead.

## 4.4 Sensitivity analysis of cut-off selections

In order to assess the performance of the cut-off selection methods, an additional set of sensitivity analyses were performed. In a series of iterations, each account in the training subset of data set B is given a score. The scores are drawn such that they are decreasing in quality for each iteration. For each spectra of scores, cut-off values are trained and the profits obtained by applying the cut-off values to the underlying set of accounts are calculated.

Figure 4.13 shows the results when the scores are probability estimates. Each line corresponds to the profits obtained by applying the cut-off methods;

- (A) The efficient frontier cut-off
- (B)  $EMP_b$  with constant  $c$
- (C)  $EMP_{bc}$  assuming exponentially distributed  $c$
- (D)  $EMP_{bc}$  assuming mixed distribution for  $c$
- (E) The educated guess of cut-off

The mixture distribution function which is adapted for  $c$  in (D) is a linear combination of exponential and Weibull distribution functions. The resulting distribution has both heavy tails and a reasonable density for low values of  $c$ . The x-axis in Figure 4.13 shows the mean squared difference between the scores and the true classes. The y-axis shows the ratio between profits obtained by using the different cut-off values and profits obtained if all accounts are accepted.

By definition, the efficient frontier cut-off gives the most profitable portfolios. The  $EMP_b$  cut-off gives profits that are inseparable from the efficient frontier cut-off.  $EMP_{bc}$

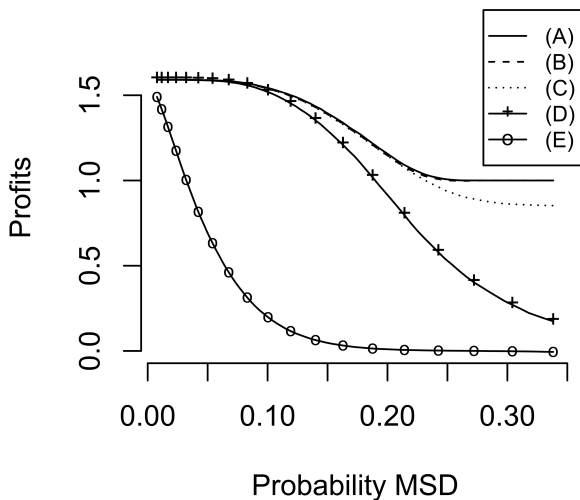


Figure 4.13: Sensitivity analysis of cut-off methods applied to probability scores from training set B. The cut-off methods used are (A) the efficient frontier cut-off, (B)  $EMP_b$  with constant  $c$ , (C)  $EMP_{bc}$  assuming exponentially distributed  $c$ , (D)  $EMP_{bc}$  assuming mixed distribution for  $c$  (E) educated guess. The y-axis shows the ratio of profits obtained when using each cut-off method and the profit when all accounts are accepted. The x-axis shows the mean squared difference between the scores and actual classes.

cut-off assuming exponentially distributed  $c$  is also close to optimal for probability scores with MSD less than 0.20. Both the  $EMP_{bc}$  cut-off assuming a mixed distribution for  $c$  and the educated guess of cut-off value fail to recognise the most profitable cut-off values. When the MSD reaches 0.25 and the most profitable cut-off is to accept all accounts, only  $EMP_b$  does this together with the efficiency curve cut-off.

Figure 4.14 shows the results of a similar sensitivity analysis where the scores are estimates of expected profit conditional on class. The x-axis shows the mean squared difference between the scores and the true margins scaled by the variance of the true margins. The y-axis and the cut-off methods used are the same as in Figure 4.13.

The results are similar to those in Figure 4.13. Cut-off selection methods (B) and (C) give portfolios that are close to optimal. The quality of educated guesses in cut-off (E) quickly

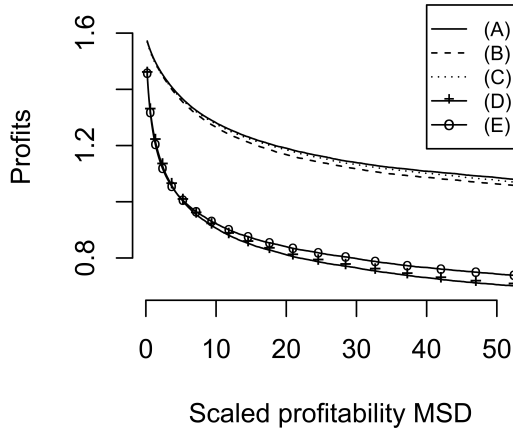


Figure 4.14: Sensitivity analysis of cut-off methods applied to varying profitability scores in the training subset of B. The cut-off methods used are (A) the efficient frontier cut-off, (B)  $EMP_b$  with constant  $c$ , (C)  $EMP_{bc}$  assuming exponentially distributed  $c$ , (D)  $EMP_{bc}$  assuming mixed distribution for  $c$  and (E) letting the cut-off be 0. The y-axis shows the ratio of profits obtained when using each cut-off method and the profit when all accounts are accepted. The x-axis shows the mean squared error of the respective scores to actual margin.

deteriorates. This is due to bias in the scoring functions. In both Figures 4.13 and 4.14, the  $EMP_{bc}$  cut-off assuming mixed distribution for  $c$  give too high cut-off values.

## Chapter 5

# Summary and recommendations for further work

The purpose of this thesis has been to investigate methods of scoring credit card accounts with a profitability objective. In particular, the risk-reward credit scoring method was introduced and the cut-off selection technique of [Verbraken et al. \(2014\)](#) was adapted for use in credit card scoring.

### 5.1 Summary and conclusions

Sensitivity analyses of scoring functions showed that the risk-reward scoring method can outperform conventional methods. The best results came from situations where there existed reliable models for estimating the expected gain from good accounts. Sensitivity analyses of cut-off methods showed that the EMP cut-off selection procedure gives more profitable portfolios than cut-off values based on maximising accuracy or educated guesses.

In addition to the results regarding the new scoring method and cut-off selection techniques, the project also contained a case study in which the scoring methods and cut-off

selection techniques were tested on a two sets of Norwegian credit card accounts. An important part of the data were monthly observations of the net profit margin each account had represented to the company. A defaulter was defined as someone who represented a negative 12-month cumulative profit margin. Logistic models were built to estimate probabilities that the accounts would end in default. As the ratio of accounts that defaulted were low, the corresponding models were prone to selection bias. Negative binomial models were built to estimate the gain among good accounts. The negative binomial regressions did not give optimal models. This was due to a high number of accounts with slightly positive profit margin. To estimate the losses observed from defaulters, a logistic model was fitted to the rounded ratio of observed losses and respective credit limit.

The scoring functions provided profitable cut-off values only in one of the testing sets. There, the profitability scoring function  $S_{\text{Prof}}$  gave the most profitable cut-off values. Among the cut-off values applied to this set of scores, the efficiency curve cut-off and EMP cut-off assuming constant gains gave the most profitable portfolios. These are both utilising the distribution of losses among defaulters. The fact that the EMP cut-off was slightly more profitable in the testing set, suggests that the EMP cut-off was slightly less prone to over fitting than the efficiency curve cut-off.

## 5.2 Discussion

Having observations of cumulative monthly profit margin is an important step towards more profitable scoring methods. In the conventional credit scoring scenario, when scores are built on estimates of probabilities, the focus is to obtain as high accuracy as possible. Having observations of the profitability of each account, it is possible to accept a reduction in accuracy if the resulting portfolios are more profitable. The refined definition of a defaulter might be more profitable as it is thinkable that some accounts represent a gain in the long run even though they fall behind on some monthly payments.

Based on the results of the case study and the sensitivity analyses, it is clear that the

usefulness of the scoring functions and cut-off selection methods vary depending on the setting. If there are reliable models for probabilities and gains, then the risk-reward scoring method can give more profitable portfolios. It has been shown via numerical examples that the risk-reward scoring function can be interpreted as a generalisation of the probability scores. If a reliable model for losses can be built, then the profitability scoring function can outperform both probability and risk-reward scoring. This is what happened in the case study. If there is a high correlation between estimated losses and gains, such that the accounts with large losses are similar to the accounts with large gains, then the risk-reward scores reduce to probability scores.

Both EMP and efficient frontier cut-off selection methods are independent from the nature of the scoring function. This is important when applying the risk-reward scoring function, as there is no apparent and intuitive cut-off present. It has been shown via sensitivity analyses that the EMP cut-off gives similar portfolios as the efficiency curve cut-off. However, the case study indicated that the efficiency curves tend to be jagged and the corresponding cut-off selection is thus prone to be over fitted. Therefore, EMP is a good choice provided that the underlying parameters in the EMP method can be estimated in a reliable way.

A cut-off equal to zero for the scoring  $S_{\text{Prof}}$  would be the best choice if all estimates were perfect. Similarly, letting the cut-off be the educated guess for the scoring  $S_{\text{Prob}}$  would be an optimal choice if the estimated probabilities were close to perfect and the variations in losses and gains were small. As this is not the case, and rarely is, it is better to use the trends in losses and gains to train a cut-off.  $\text{EMP}_{bc}$  assuming mixed distribution for  $c$  greatly exaggerate the cut-off values. This is due to the high probability mass of slightly positive  $c$  that produce instabilities in the numerical scheme for calculating the cut-off. As indicated in section 3.2.4, when integrating over a range of different values of  $c$ , problems in the discretisation of optimal values of  $b$  can occur. These instabilities produce too large cut-off values and thus non-optimal portfolios.  $\text{EMP}_b$  and  $\text{EMP}_{bc}$  with exponential distribution for  $c$  gave close to optimal cut-off values in the sensitivity analyses.

### **5.3 Recommendations for Further Work**

Possible extensions to the work done in this thesis is to examine alternatives to the choices that have been made during the project.

A comparison between the different definitions of a defaulter could shed more light on the value of having observations of monthly net profit margins.

Concerning the scoring functions, there exist many different methods of estimating probabilities, gains and losses. Other models might be better fits to the data and thus give scoring functions of higher quality. In addition, there are several possible choices of parametrisations of the risk-reward scoring function. Other versions than the one chosen in this thesis might provide more profitable cut-off values.

There were many choices made regarding the cut-off selection methods as well. Tuning the cut-off selection methods can lead to more profitable portfolios. For instance, the efficiency curves could be smoothed and interpolated in order to avoid letting the cut-off be the peak of a jagged line. In addition, it is thinkable that experimenting with different choices of distributions to include in the EMP cut-off selection procedure also can improve the performance.

# Bibliography

- Baesens, B. (2014). *Analytics in a Big Data World*. John Wiley & Sons, Incorporated. Available from: ProQuest Ebook Central. [8 February 2017].
- Baesens, B., Van Gestel, T., Viaene, S., Stepanova, M., Suykens, J., and Vanthienen, J. (2003). Benchmarking state-of-the-art classification algorithms for credit scoring. *Journal of the Operational Research Society*, 54(6):627–635.
- Basel (2005). Basel committee on banking supervision: International convergence of capital measurement and capital standards: A revised framework. Bank for International Settlements.
- Bølviken, E. (2014). *Computation and Modelling in Insurance and Finance*. Cambridge University Press.
- Boyes, W., Hoffman, D., and Stuart, A. (1989). An econometric analysis of the bank credit scoring problem. *Journal of Econometrics*, 40(1):3–14.
- Bravo, C., Maldonado, S., and Weber, R. (2013). Granting and managing loans for micro-entrepreneurs: New developments and practical experiences. *European Journal of Operational Research*, 227(2):358–366.
- Christensen, R. (1997). *Log-Linear Models and Logistic Regression*. Springer.



- Crook, J. and Banasik, J. (2004). Does reject inference really improve the performance of application scoring models? *Journal of Banking & Finance*, 28(4):857–874.
- Crook, J., Edelman, D., and Thomas, L. (2007). Recent developments in consumer credit risk assessment. *European Journal of Operational Research*, 183(3):1447–1465.
- Fawcett, T. (2006). An introduction to roc analysis. *Pattern Recognition Letters*, 27(8):861–874.
- Finlay, S. (2008). Towards profitability: a utility approach to the credit scoring problem. *Journal of the Operational Research Society*, 59(7):921–931.
- Finlay, S. (2010). Credit scoring for profitability objectives. *European Journal of Operational Research*, 202(2):528–237.
- Hand, D. (2005). Good practice in retail credit scorecard assessment. *The Journal of the Operational Research Society*, 59(9):1109–1117.
- Hand, D. (2009). Measuring classifier performance: a coherent alternative to the area under the roc curve. *Machine Learning*, 77(1):103–124.
- Hand, D. and Anagnostopoulos, C. (2014). A better beta for the h measure of classification performance. *Pattern Recognition Letters*, 40(1):41–46.
- Hand, D. and Henley, W. (1993). Can reject inference ever work? *IMA Journal of Mathematics Applied in Business & Industry*, 5(1):43–55.
- Lessmann, S., Baesens, B., Seow, H., and Thomas, L. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *European Journal of Operational Research*, 247(1):124–136.
- Lucas, A. (2001). Statistical challenges in credit card issuing. *Applied Stochastic Models in Business and Industry*, 17(1):83–92.

- Maldonado, S., Flores, A., Verbraken, T., Baesens, B., and Weber, R. (2015). Profit-based feature selection using support vector machines – general framework and an application for customer retention. *Applied Soft Computing*, 35(1):740–748.
- Ohlsson, E. and Johansson, B. (2010). *Non-Life Insurance Pricing with Generalized Linear Models*. Springer Berlin Heidelberg.
- Oliver, R. and Wells, E. (2001). Efficient frontier cutoff policies in credit portfolios. *The Journal of the Operational Research Society*, 52(9):1025–1033.
- Stewart, R. (2011). A profit-based scoring system in consumer credit: making acquisition decisions for credit cards. *The Journal of the Operational Research Society*, 62(9):1719–1725.
- Thomas, L. (2000). A survey of credit and behavioural scoring: forecasting financial risk of lending to consumers. *International Journal of Forecasting*, 16(2):149–172.
- Thomas, L., Edelman, D., and Crook, J. (2002). *Credit Scoring and Its Applications*. SIAM.
- van der Wijst, N. (2013). *Finance, a quantitative introduction*. Cambridge University Press.
- Verbraken, T., Bravo, C., Weber, R., and Baesens, B. (2014). Development and application of consumer credit scoring models using profit-based classification measures. *European Journal of Operational Research*, 238(2):505–513.
- Verbraken, T., Verbeke, W., and Baesens, B. (2013). A novel profit maximizing metric for measuring classification performance of customer churn prediction models. *IEEE Transactions on Knowledge and Data Engineering*, 25(5):961–973.