Acceptance and Profitability Modelling for Consumer Loans

Pingchuan Ma

Doctor of Philosophy

Business School

University of Edinburgh

2008

(Confidential)

Abstract

This thesis explores and models the relationships between offers of credit products, credit scores, consumers' acceptance decisions and expected profits generated using data that records actual choices made by customers and their monthly account status after being accepted. Based on Keeney and Oliver's theoretical work, this thesis estimates the expected profits for the lender at the time of application, draws the iso-profit curves and iso-preference curves, derives optimal policy decisions subject to various constraints and compares the economic benefits after the segmentation analysis.

This thesis also addresses other research issues that have emerged during the exploration into profitability and acceptance. We use a Bivariate Sample Selection model to test the existence of sample selection bias and found that acceptance inference may not be necessary for our data. We compared the predictive performance of Support Vector Machines (SVMs) vs. Logistic Regression (LR) on default data as well as on acceptance data, without finding that SVMs outperform LR. We applied different Survival Analysis models on two events of interest, default and paying back early. Our results favoured semi-parametric PH-Cox models separately estimated for each hazard.

Acknowledgements

Special thanks to my PhD supervisors Professor Jonathan Crook and Professor Jake Ansell. They guided me all the way through the research and the writing of this thesis. I owe them lots of gratitudes for having me work with them. Their knowledge and enthusiasm for producing high-quality work have set such high standards that I will spend the rest of my life trying to reach.

I am grateful to the generous funding support from The Credit Research Centre, Management School at University of Edinburgh and the Overseas Research Students Awards Scheme. Also I want to thank Dr Tom Archibald, Dr Galina Andreeva and Dr Tony Bellotti for their valuable advices during my PhD study.

I am very grateful for all the supports from my family and friends in the UK and back to China. Especially, I would like to pay my special thanks to my wife Li Ding. Without her, I cannot imagine how I get where I am today.

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

(*Pingchuan Ma*)

12/02/2009

To my parents.

Table of Contents

1 Introduction **1**

1.1 Aim of thesis . . . 1

1.2 Importance of the research . . . 3

1.3 Contributions to knowledge . . . 5

1.4 Thesis structure . . . 8

2 Literature Review **11**

2.1 Introduction . . . 11

2.1.1 Consumer's preference . . . 13

2.1.2 Consequence to the lender . . . 14

2.1.3 Set of win-win offers . . . 19

2.1.4 Lender's preferences . . . 21

2.2 Previous research in acceptance modelling . . . 23

2.3 Previous research in profit scoring . . . 25

2.3.1 Maximize profit using default based score . . . 26

2.3.2 Relationship between Profit, Volume and Loss in a portfolio . 27

2.3.3 Risk based pricing . . . 32

2.3.4 Other profit scoring approaches . . . 35

2.3.5 Profit affected by Basel II 42

2.4 Conclusion 43

3 Default Risk Modelling **44**

3.1 Introduction 44

3.2 Previous research in default risk modelling 44

3.3 Logistic regression 48

3.3.1 Variable selection using step-wise selection 49

3.3.2 Existence of MLE(maximum likelihood estimation) 50

3.4 Support vector machines 51

3.4.1 Formulation 52

3.4.2 Practical concerns 56

3.5 Performance measures 58

3.5.1 Definition of the ROC and area under ROC curve 58

3.5.2 Area under ROC curve compared with Accuracy Ratio as a performance measure 61

3.6 Data preparation 63

3.6.1 Definition of default 63

3.6.2 Bands separation 63

3.6.3 Data transformation 65

3.6.4 Training and holdout sample separation 66

3.7 Results 67

3.7.1 Logistic regression results 67

3.7.2 Unweighted SVM results 70

3.7.3 Weighted SVM results 74

3.7.4 Summary 75

3.8 Simultaneous Equations Model 76
3.8.1 Investigation of the relationships between default, rate and score using Bayesian network structure search 76
3.8.2 Simultaneous equations model 78
3.8.3 Two-stage least squares method 80
3.8.4 Estimates of the simultaneous equations model and prediction results 81
3.9 Conclusion 84

4 Acceptance Modelling **86**
4.1 Introduction 86
4.2 Previous research in acceptance modelling 87
4.3 Model design 90
4.4 Data preparation 92
4.5 Logistic regression results 93
4.5.1 Performance across bands 93
4.5.2 Features selected from stepwise selection 94
4.5.3 ROC curve and MLE results 95
4.6 SVMs results 95
4.6.1 RBF kernel 97
4.6.2 Polynomial kernel 98
4.7 Modelling acceptance elasticity of interest rate 99
4.7.1 Elasticities for bands 100
4.8 Bivariate Probit sample selection model 106
4.8.1 Background and previous research 106
4.8.2 Estimation results 109

4.9 Indifference curves 117
4.9.1 Indifference curves based on Logit model 117
4.9.2 Indifference curves using a different approach 120
4.10 Conclusion 122
5 Survival Analysis **124**
5.1 Introduction 124
5.2 Introduction to survival analysis 124
5.2.1 Nonparametric model 125
5.2.2 Parametric regression survival models 126
5.2.3 Cox proportional hazards model 127
5.3 Data description 130
5.3.1 Description of the data using the Kaplan-Meier model 131
5.3.2 Differences between customers from two different brands . . 133
5.3.3 Difference between hazard functions with different loan terms 136
5.4 Results of default modelling 138
5.4.1 Selection of explanatory variables 139
5.4.2 Parametric regression using the Weibull distribution 139
5.4.3 Parametric regression using the Exponential distribution . . . 142
5.4.4 Cox proportional hazard model 145
5.4.5 Test of proportional hazard assumption 148
5.4.6 Conclusion 152
5.5 Results of paying back early modelling 153
5.5.1 Parametric proportional hazards modelling results 154
5.5.2 PH Cox model results 156
5.5.3 Parametric accelerated failure time models 156

5.5.4 Model comparison 158
5.5.5 Conclusion 162
5.6 Results of competing risks modelling 163
5.7 Conclusion 165

6 Profitability Modelling **167**
6.1 Introduction 167
6.2 Estimating equations 168
6.2.1 Conditional expected profit 169
6.2.2 Acceptance probability 177
6.2.3 Unconditional expected profit 180
6.2.4 Iso-profit curves plotted together with iso-preference curves . 182
6.3 Optimal decision policies 183
6.3.1 Optimal policies if choice is of APR rate s.t. p(a) 186
6.3.2 Optimal policies if choice is of APR given loan amount 189
6.4 Segmentation 192
6.4.1 Sensitivity test of the segmentation based on application channel 192
6.4.2 Economic benefit of the segmentation 202
6.5 Conclusion 207

7 Conclusion **209**
7.1 Summary of findings 209
7.1.1 How can we model the profitability of making a loan, unconditional on the acceptance by the applicants, and how can iso-profit and iso-acceptance contours be empirically estimated and presented? 210

7.1.2 Is acceptance inference needed? 211
7.1.3 How do novel approaches like support vector machines (SVMs) perform (compared to logistic regression) in predicting default and acceptance ? 212
7.1.4 How to model the chance of default and paying back early and how to incorporate them into a profit estimation? 213
7.2 Contributions to knowledge reiterated 214
7.3 Limitations and future research 216

References **218**

A Appendix for Chapter 3 **225**

B Appendix for Chapter 4 **238**

C Appendix for Chapter 5 **245**

D Appendix for Chapter 6 **264**

List of Figures

2.1 Consumers' iso-preference curves. The lines represent contours around a utility hill where the third dimension is utility. Figure based on Keeney and Oliver (2005) 13

2.2 Lender's profits are different along the consumer iso preference curve. Figure based on Keeney and Oliver (2005) 14

2.3 Offers yielding maximum expected profit for the lender. Figure based on Keeney and Oliver (2005) 15

2.4 Maximum expected profit at each probability of take. Figure based on Keeney and Oliver (2005) 16

2.5 Unconditional expected profit. Figure based on Keeney and Oliver (2005) 17

2.6 Iso-profit contours for the lender in the consumer iso-preference map. Figure based on Keeney and Oliver (2005) 19

2.7 Set of win win offers. Figure based on Keeney and Oliver (2005) . . 20

2.8 Lender's iso-preference curves. Figure based on Keeney and Oliver (2005) 22

2.9 Trade off value per acquisition. Figure based on Keeney and Oliver (2005) 22

2.10 Efficient Frontier when the objective is to minimize expected loss subject to given expected profit, Figure based on Oliver and Wells (2001) 29
2.11 Efficient Frontier when the objective is to maximize expected profit subject to given expected volume, Figure based on Oliver and Wells (2001) 30

3.1 SVM on two-class linearly separable data 53
3.2 Sample ROC curve 60
3.3 AuROC of SVM using Linear kernel 71
3.4 SVM using Polynomial kernel. 72
3.5 SVM using RBF kernel. 73
3.6 AuROC of SVM using RBF kernel 73
3.7 SVM using Sigmoid kernel. 74
3.8 Area under the ROC of Weighted SVM using RBF kernel. 75
3.9 The relationships between 3 variables in a graph found after a global search 77
3.10 The relationships between 3 variables in a graph found after a local search 77
3.11 Area under ROC given by CDSIMEQ 84

4.1 Our data samples sets 91
4.2 The ROC curve for the Acceptance model all bands combined 96
4.3 Grid search of the best predictive model parameters for SVM RBF kernel 97
4.4 Grid search of the best predictive model parameters for SVM Polynomial kernel with dimension parameter $d = 3$ 98
4.5 Band 10 has shown the most elasticity at interest rate of 13% 103

4.6 Band 20 has shown the most elasticity at interest rate of 12% 103
4.7 Band 30 has shown the most elasticity at interest rate of 13% 104
4.8 Band 40 has shown the most elasticity at interest rate of 14% 104
4.9 Band 50 has shown the most elasticity at interest rate of 17% 105
4.10 Band 60 has shown the most elasticity at interest rate of 29% 105
4.11 Combining all bands together has seen the most elasticity at interest rate around 14% 106
4.12 Compare ROC curves of Probit and Heckprob models 116
4.13 Indifference curves 118
4.14 Eigenvalues after PCA 121
4.15 Indifference curves using a different approach 122

5.1 Example of hazard functions for Weibull models 127
5.2 Observed loan terms 131
5.3 Kaplan-Meier survival functions, default and paying back early 132
5.4 Kaplan-Meier hazard functions, default and paying back early 133
5.5 Kaplan-Meier hazard function for default 133
5.6 The KM hazard functions of the Brand1 customers for paying back early and default 134
5.7 The KM hazard functions of the Brand1 customers for default 135
5.8 The KM hazard functions of the Brand2 customers for paying back early and default 135
5.9 The KM hazard functions of the Brand2 customers for default 136
5.10 Compare the Hazard functions for loans with different terms. 137
5.11 Compare the default hazard functions for loans with different terms. . 138
5.12 Hazard function, Weibull distribution, default data 140

5.13 Survivor function, Weibull distribution, default data 140
5.14 Area under ROC, Weibull distribution, default data 142
5.15 Survivor function, Exponential distribution, default data 142
5.16 Hazard function, Exponential distribution, default data 143
5.17 Area under ROC, Exponential distribution, default data 143
5.18 Baseline function S(t), Cox model, default data 145
5.19 Area under ROC, Cox PH model, default data 145
5.20 Scaled Schoenfeld Residuals for spsetld, violating PH Assumption, default data 150
5.21 Scaled Schoenfeld Residuals for AGE, non violating PH Assumption, default data 150
5.22 Area under ROC, Cox model with time dependent covariates, default data 152
5.23 Hazard function, Weibull model, paying back early data 155
5.24 Hazard function, Exponential model, paying back early data 155
5.25 Hazard function, PH Cox model, paying back early data 156
5.26 Hazard function, Lognormal model, paying back early data 157
5.27 Hazard function, Loglogistic model, paying back early data 157
5.28 Using Cox-Snell residuals to check the fit of Weibull model 159
5.29 Using Cox-Snell residuals to check the fit of Exponential model . . . 159
5.30 Using Cox-Snell residuals to check the fit of Lognormal model 160
5.31 Using Cox-Snell residuals to check the fit of Loglogistic model 160
5.32 Using Cox-Snell residuals to check the fit of PH Cox model 161

6.6 The iso-profit curves plotted together with the iso-preference curves. 183
6.7 Decision diagram 184

6.8 Contour of unconditional expected profit with peak point marked . . . 185
6.9 The acceptance line $p(a) = 0.6$ on the iso-profit contours. 187
6.10 The loan amount that maximise the expected profit when $p(a) \geq 0.6$. 188
6.11 The optimal rate APR given Loan Amount L if p(a) is not considered. 190
6.12 The optimal rate APR given Loan Amount L if p(a) is the constraint. 191
6.13 Acceptance probabilities for customers applying through Internet . . . 199
6.14 Acceptance probabilities for customers applying through Non-Internet 199
6.15 Unconditional expected profits for customers applying through Internet 200
6.16 Unconditional expected profits for customers applying through Non-Internet 200
6.17 Iso-profit and iso-acceptance curves (marked with p(a) from 0.1 to 0.9)for customers applying through Internet 201
6.18 Iso-profit and iso-acceptance curves (marked with p(a) from 0.1 to 0.9) for customers applying through Non-Internet 201

List of Tables

3.1 Compare the performance of different classifiers using accuracy, enlarged based on Crook et al. (2007) 46
3.2 Confusion Matrix 59
3.3 Example of a Cost Matrix 62
3.4 Average interest rates across bands 64
3.5 The predictive performance measured in AuROC across bands. 68
3.6 Variable name explanation 69

4.1 Comparison of the predictive performance for different risk bands . . 94
4.2 The price elasticities across different bands 101
4.3 The elasticities of other variables on all bands combined 101
4.6 The predictive performance of lean model and a complex model . . . 116

5.1 Size of different loan term groups 136
5.2 Weibull model estimates for default 141
5.3 Parametric regression results using Exponential Distribution on default 144
5.4 PH Cox model estimates on default 146
5.5 Test of proportional hazards assumption 149
5.6 Comparison of the model predictive performance on holdout set . . . 161

5.7 Comparison of the model performance 165

6.1 Matrix of optimal decision policies by lender 186

6.2 Compare the AuROC values on the holdout set on different models . . 195

6.3 Benefits of Segmentation Model 1 203

6.4 Benefits of Segmentation Model 3 204

A.1 Maximum likelihood estimates of parameters by Logistic Regression across all bands. To be continued in table A.2 226

A.2 Following table A.1. Maximum likelihood estimates of parameters by Logistic Regression across all bands. 227

A.3 Maximum likelihood estimates of parameters by Logistic Regression after stepwised selection based on 223 cases in band 10. 228

A.4 Maximum likelihood estimates of parameters by Logistic Regression after stepwised selection based on 389 cases in band 20. 228

A.5 Maximum likelihood estimates of parameters by Logistic Regression after stepwised selection based on 997 cases in band 30. 229

A.6 Maximum likelihood estimates of parameters by Logistic Regression after stepwised selection based on 1327 cases in band 40. 230

A.7 Maximum likelihood estimates of parameters by Logistic Regression after stepwised selection based on 3209 cases in band 50. 231

A.8 Maximum likelihood estimates of parameters by Logistic Regression after stepwised selection based on 15766 cases in band 60. To be continued. 232

A.9 Following table A.8. Maximum likelihood estimates of parameters by Logistic Regression after stepwised selection based on 15766 cases in band 60. 233
A.10 Maximum likelihood estimates of parameters by Logistic Regression after stepwised selection based on data of band 0 and 20 233
A.11 Equation for APR in Simultaneous Equations with apr adjustment variable 234
A.12 Equation for Default in Simultaneous Equations with apr adjustment variable 235
A.13 Equation for APR in Simultaneous Equations without apr adjustment variable 236
A.14 Equation for Default in Simultaneous Equations without apr adjustment variable 237

B.2 Probit Default Model with variables stepwise-selected with significance value of 0.002 242
B.3 Probit Default Model with variables stepwise-selected with significance value of 0.05 242
B.5 Parameters estimated from OLS regression 244

C.1 Use time dependent covariates to test the PH assumption, group 1 . . 246
C.2 Use time dependent covariates to test the PH assumption, group 2 . . 247
C.3 Use time dependent covariates to test the PH assumption, group 3 . . 248
C.4 Cox with time dependent covariates 249
C.5 Estimates from Weibull Model on paying back early 250
C.6 Estimates from Exponential Model on paying back early 251

C.7 Estimates from PH Cox on paying back early 252

D.1 Model 1 acceptance Non-segmentation 264

D.2 Model 1 acceptance on Internet segment 266

D.3 Model 1 acceptance on Non-Internet segment 267

D.4 Model 1 default Non-Segmentation 268

D.5 Model 1 default on Internet segment 269

D.6 Model 1 default on Non-Internet segment 270

D.7 Model 1 paying back early Non-Segmentation 271

D.8 Model 1 paying back early Internet segment 274

D.9 Model 1 paying back early Non-Internet segment 275

D.10 Model 2 acceptance Non-Segmentation 277

D.11 Model 2 acceptance on Internet segment 279

D.12 Model 2 acceptance on Non-Internet segment 280

D.13 Model 2 default Non-Segmentation 282

D.14 Model 2 default on Internet segment 283

D.15 Model 2 default on Non-Internet segment 283

D.16 Model 2 paying back early Non-Segmentation 284

D.17 Model 2 paying back early on Internet segment 286

D.18 Model 2 paying back early on Non-Internet segment 287

D.19 Model 3 Acceptance on Internet segment 288

D.20 Model 3 Acceptance on Non-Internet segment 289

D.21 Model 3 default on Internet segment 291

D.22 Model 3 default on Non-Internet segment 291

D.23 Model 3 paying back early on Internet segment 292

D.24 Model 3 paying back early on Non-Internet segment 293

D.25 Model 4 Acceptance on Internet segment 295
D.26 Model 4 Acceptance on Non-Internet segment 297
D.27 Model 4 default on Internet segment 299
D.28 Model 4 default on Non-Internet segment 299
D.29 Model 4 paying back early on Internet segment 300
D.30 Model 4 paying back early on Non-Internet segment 302

Chapter 1

Introduction

1.1 Aim of thesis

Over the last two decades the growth in consumer debt has been rapid. In the US the total consumer credit outstanding has tripled from 0.8 to 2.5 trillions of dollars from 1990 to 2008[1]. The UK total consumer credit outstanding quadrupled from £52 to £229 billions from 1993 to 2008[2]. One of the forces facilitating such a fast pace of growth is the wide application of Credit Scoring techniques, which automatically assesses the risk and profit involved in lending to an individual applicant and therefore make millions of lending decisions economically possible.

When a customer fills in the application for a credit product (a fixed term loan, for example), the lender will firstly evaluate his/her credit worthiness by assessing the risk of default. If the credit score is higher than the cut-off threshold set by the lender, an offer (involving an interest rate) will be made to the customer. Subject to the attrac-

[1]http://www.federalreserve.gov/releases/g19/Current/

[2]Bank of England statistics, code LPMVZRD

tiveness perceived by the customer, the offer might be accepted or rejected. Having accepted the offer and received the loan amount (in the case of a fixed loan amount), the customer will be obliged to make monthly payments until the end of the term when the balance is cleared. During this payment period, some will choose to close the account by paying back the remaining balance before the end of the term. Some will stop making payments and default. The rest of the customers will keep making payments until the end of the term.

In a competitive retail lending market, given the business objectives of either maximization profit or market share, lenders need to develop models taking into consideration profit and market share. This thesis explores and models the relationships between offers of credit products, credit scores, consumers' acceptance decisions and expected profit generated using data that records the actual choices made by customers and their monthly account status after being accepted. Specifically, this thesis attempts to address the following issues:

1. How can we model the profitability of making a loan, unconditional on the acceptance by the applicants, and how can iso-profit and iso-acceptance contours be empirically estimated and presented?

2. Is acceptance inference needed?

3. How do novel approaches like support vector machines (SVMs) perform (compared to logistic regression) in predicting default and acceptance ?

4. How to model the chance of default and paying back early and how to incorporate them into a profit estimation?

1.2 Importance of the research

Profit Scoring has been a promising research direction in the Credit Scoring literature (see Hopper and Lewis (1992), Oliver (1993), Marshall and Oliver (1995), Hand and Kelly (2001) Li and Hand (2002), Somers and Whittaker (2007), Trench et al. (2003), Andreeva et al. (2007), Keeney and Oliver (2004), Keeney and Oliver (2005) for example). Most of the research in Profit Scoring models the profitability either of each customer or of a whole portfolio (except Keeney and Oliver (2005)). Those analyses are based on data collected from existing customers. Unfortunately, customers who have rejected offers made to them by lenders and therefore who are not existing customers in the previous analysis have been neglected in the profit predictions. The analysis of the profitability without considering the acceptance of offers, we argue, is not complete in a competitive market where no lender can guarantee all of its offers are accepted.

In previous acceptance modelling research Jung et al. (2003), Seow and Thomas (2005) and Thomas et al. (2006) modelled acceptance behaviour using data relating to a hypothetical student bank account where participants (first year students at the University of Southampton) chose offers of different features. The hypothetical nature of the data collected together with the small sample size limits the applicability of their results. A large data set of actual responses of applicants to real offers made to them is exactly what is needed for acceptance modelling.

Another question that remains unanswered in previous acceptance modelling research is the possible need for "acceptance inference". Similar to the scenario of the need for Reject Inference, customers who rejected offers might do so because they have better scores (low default probabilities) which enable them to shop around to find good deals. On the other hand, customers who accept offers may do so just because of low scores (high risk of default) and have limited choices. If such sample selection leads to biased parameter estimates of the probability of default , models built on the applicants who have accepted the offers will be different from models built on all the applicants (which include those who have accepted and rejected the offers put to them).

Apart from acceptance modelling, two other important factors affecting profitability are the likelihood of default and of paying back early. In the literature on default modelling, many studies have proposed and compared different approaches to separating the defaulters from the non-defaulters by assuming that the probability of default is dependent on a set of predictive variables. Recently more and more lenders have come to use Risk Based Pricing instead of charging a flat interest rate for all customers. Risk Based Pricing generally involves charging riskier customers higher interest rates. The probabilities of default perceived by the lenders' credit scoring systems are therefore reflected in the interest rates charged. If future predictions are to be made based on the models built on such data, the existence of a reverse influence of the probability of Default on the Interest Rate cannot simply be ruled out.

As another crucial factor contributing to profitability estimation, the probability of paying back early has not received as much attention as the probability of default in the literature. In fact, the average probability of paying back early is observed to be more than 10 times larger than the average probability of default in our data. This contrast indicates the high level of competition between lenders during the period in which this set of data was collected. Without modelling the probability of paying back early it will not be possible to accurately calculate expected profits .

The scarcity of studies that model the probability of paying back early leads to the lack of investigations estimating the probabilities of default and paying back early under a competing risk framework in the credit scoring literature. These two events of interest can be assumed to be independent and estimated separately. But once this assumption of independence is questioned, it will be interesting to see how the competing risk approach can be applied and whether improvements can be made to our model as a result of using this approach.

1.3 Contributions to knowledge

This thesis makes a number of contributions to the literature. First, it is the first empirical academic study to estimate expected profits *at the time of application*. Previous literature predicts the profits of customers who have already accepted an offer (for example, see Somers and Whittaker (2007), Trench et al. (2003) and Andreeva et al. (2007)). We estimate expected profits by combining the results from acceptance modelling, survival analyses of default and of paying back early.

Second, the research is based on a unique data set reflecting the actual acceptance choices made by customers of a real financial product, and which records their default performance and early repayment behaviour. Previous research used a data set recording undergraduate students' acceptance choices towards offers of a hypothetical bank account (see Jung et al. (2003) and Seow and Thomas (2005)). The findings from our model will be closer to what will be observed in the practical retail lending industry than those in the literature.

Third, this thesis provides iso-preference curves and iso-profit curves as an empirical implementation of Keeney and Oliver's theoretical model. However, our iso-preference curves, which were drawn based on estimates from the data, indicate that the customers prefer lower loan amounts, rather than larger amounts which may be contrary to the assumptions of a preference for higher credit lines in the K-O model.

Fourthly, using iso-preference curves and iso-profit contours, this thesis illustrates how to maximize unconditional profit under different objectives which the lender may choose. Previous literature such as Keeney and Oliver (2005) discusses optimal strategies using assumed numerical cost and profit figures as example cases while this thesis uses results estimated from industry sourced data.

Fifthly, this thesis also provides a segmentation analysis by separately estimating the profits on Internet and Non-Internet groups. The optimal interest rates are then chosen separately for each segment for each given loan amount requested. Our results demonstrate that when offerings in the fixed term loan market are segmented in this way, markedly different policy decisions would be made, compared with those drawn from non-segmented data.

Sixthly, we explore the possible existence of sample selection bias due to estimating a default model using a sample that omits those who rejected a loan offer made to them after application. Previous literature has suggested why a limited improvement can be achieved through reject inference unless very high cut-off values are used (see Crook and Banasik (2004), Banasik and Crook (2005)). The finding in this thesis suggests it is highly unlikely that our default models suffer from sample selection bias when only the customers having accepted the offer have performance data recorded and are included in the default models. Utilising the acceptance data, a bivariate probit sample selection model does not give higher predictive performance compared with a simple probit model based on the default data only for borrowers who accepted a loan offer. We also find a significant correlation between the residuals of the default and acceptance models only when a lean model is used. This suggests that acceptance inference might not be necessary.

A further contribution is a comparison between classification methods: SVM (support vector machines) vs logistic regression, which has been carried out to model default and acceptance probabilities. The SVM, albeit found to record good performance in the literature (Baesens (2003), Baesens et al. (2003)) , does not predict as well as the logistic regression on our Default data in terms of the Area under ROC curves. But SVMs have never been applied in the acceptance modelling literature before and we find that in this context, SVM gives equally good results as the logistic regression model. The varied performance on different data by the SVM can be explained by the class distribution in the data where the Default data is much more unbalanced than the Acceptance data. This makes Default Modelling a more challenging task for the SVM as it is more sensitive to the class distribution.

Although frequently used in statistical medical research for modelling multiple failure events, competing risk survival models have rarely been used in the Credit Scoring literature (Banasik et al. (1999)). This thesis presents a comparative study of the predictive performances of competing risk survival models. We find little improvement in the predictions over previous survival models estimated separately for each hazardous event. This suggests that little benefit can be achieved by using the competing risk survival models on this type of data.

1.4 Thesis structure

The structure of the rest of the dissertation is as follows. Chapter 2 reviews the recent literature. The details of the theoretical model on which much of this project is based, the K-O model, will be introduced and discussed. This chapter continues to review the previous research in the areas of acceptance modelling and Profit Scoring. Both are essential to the implementation of the K-O model.

Chapter 3 presents the modelling of default. Two different approaches to default modelling, logistic regression and support vector machines, are compared. SVM, with its more complex model structure, does not seem to be as competitive as logistic regression in the prediction of default. Relaxing the assumption that there is one way dependence between the probability of default and the interest rate, a simultaneous equations model was used to investigate the mutual influence between the rate and the default. However, the predictive performance of this model was not as good as was achieved by a logistic regression model.

Chapter 4 shows the results of modelling consumer acceptance behaviour. The predictive performances of logistic regression models and SVMs are compared. The results show that SVMs, although giving a better predictive performance than they do in the prediction of default, do not outperform logistic regression when modelling acceptance. Chapter 4 also looks into the need for acceptance inference. An attempt has been made to improve the default estimation by applying bivariate probit with sample selection models, assuming that the residuals of the two equations are normally distributed. The insignificant improvement leads to the conclusion that our default models do not seem to suffer from the sample selection bias [3]. Finally, indifference curves have been drawn in APR vs. Loan Amount space. A difference between the shape of the indifference curves and that assumed in Keeney-Oliver model is observed and commented on.

Chapter 5 is dedicated to applying survival analysis to model two types of hazardous events that affect the profitability of customers who accepted a loan offer, default and paying back early. In this chapter non-parametric Kaplan-Meier estimates are used to compare and illustrate the differences observed between the hazard and survivor functions for the two types of hazardous events on the whole data and on different data segments. Afterwards, different parametric models and semi-parametric PH-Cox models have been fitted and assessed before their predictive performances are compared. Finally, competing risk models have been applied to estimate the probabilities of default and paying back early jointly. Their predictive performances are also compared.

Chapter 6 calculates the expected profits for the lender at the time of application using estimates from the acceptance modelling and survival models of default and paying back early. The equation to calculate the expected conditional profits together with the assumptions made are explained in detail. Plugging into the profit equation the

[3] Although this is conditional on the validity of the normality assumption.

estimates from the acceptance and PH-Cox models, the expected unconditional profits at the time of application (and so before an offer is made) are calculated and plotted in three dimensional space of profit, interest rate and loan amount. Dependent on different constraints on the optimization objectives, optimal decision policies have been discussed. This chapter also analyses the difference between models built on each of the Internet and Non-Internet segments, comparing the economic benefit of this segmentation under two different modelling assumptions.

Chapter 7 will conclude this dissertation by summarising the findings, noting some limitations of the work and then discussing possible extensions of the work in the future.

Chapter 2

Literature Review

2.1 Introduction

Today's consumer credit markets are growing very fast. The total consumer credit outstanding (combining revolving and non-revolving) totalled over 2.5 trillion dollars in the US according to the Federal Reserve Statistical Release at Q4 2007 [1]. Widely applied credit scoring techniques have helped financial institutions to design new products for customers and to accept them at much lower costs than before their use.

Lenders have traditionally focused on modelling the risk of default to make a decision of accepting or rejecting a new applicant. Using data on previous applicants and assuming the relationship between the probability of default and the predictive independent variables remains constant over time, lenders build models to predict the probability of default. Default could be defined as the chance that an applicant misses 3 or more consecutive payments in the next 12 months although other definitions are possible (see Kelly and Hand (1999)).

[1] http://www.federalreserve.gov/releases/G19/Current/

However, lenders have recognized business objectives other than the risk of default. Some research (Hopper and Lewis (1992)) has discussed the idea that profit as a measure of performance is an alternative to the probability of default. They discussed strategies that included consideration of individual account profitability instead of portfolio profitability. Keeney and Oliver (2004) and Keeney and Oliver (2005) pointed out that both profit and market share are fundamental objectives to achieve. They built a theoretical model to identify the set of win-win situations to integrate both the consumers' preferences for price and credit line and the lenders' preferences for profit and market share for a revolving credit product. The implementation of their model depends on the availability of information on

- the consumer's preferences;
- the probabilities of the consumers to accept offers from lenders;
- estimates of the consequence to the lender conditioned on the offer being accepted by the consumer;
- the lender's preference for portfolio performance.

The next four sections in this chapter will discuss in detail Keeney and Oliver's analysis of consumers' preferences, how the consumers' iso-preference can be presented using the probability of offer acceptance, the consequence to the lender under such situations and how the win-win situation set can be identified considering both the lender's and the consumer's preferences and how the lender shall express his/her preference when selecting optimal offers for the consumer. After that, implementation issues regarding how the acceptance probabilities and how the customer preferences can be estimated will be discussed.

2.1.1 Consumer's preference

Keeney and Oliver (2004)& Keeney and Oliver (2005) assumed that in a two dimensional space of Credit Line and APR Rate, a consumer wishes to get as much Credit Line as possible and to be charged as low an APR Rate as possible. Therefore, in CreditLine-APR space, it is reasonable to assume that the combinations of Credit Line and APR in the top left of Fig 2.1 are preferred by a consumer than points in the bottom right. Some points shall share a similar preference to the consumer. Connecting those points which yield the same preference by the consumer we gain iso-preference curves. The consumer who receives any offer on an iso-preference curve shall have the same probability of accepting this offer.

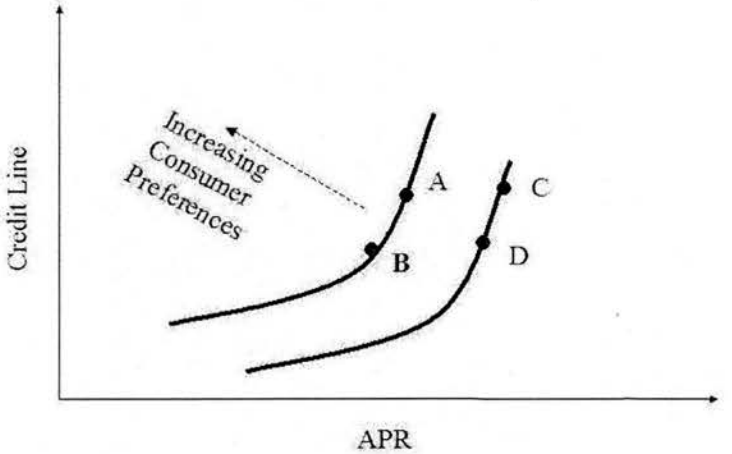

Figure 2.1: Consumers' iso-preference curves. The lines represent contours around a utility hill where the third dimension is utility. Figure based on Keeney and Oliver (2005)

Different consumers will have different iso-preference maps. The shape of the iso-preference curves represents the preferences of an individual consumer for the offer characteristics, and so the trade off between those characteristics. A consumer with extremely high price sensitivity will show nearly vertical iso-preference curves on the

2 dimensional space of Credit Line and APR Rate and a consumer with almost no sensitivity to APR will have almost horizontal iso-preference curves.

The iso-preference curves Keeney and Oliver illustrated reside in a two dimensional feature space. In a real world case, the iso-preference curves could reside in a larger dimensional space. Other features like insurance take up, length of loan, gifts such as free travel money may also be included in a consumer's utility function as empirically shown by Jung et al. (2003).

2.1.2 Consequence to the lender

A consumer is indifferent between all points on the same iso-preference curve. However, for the lender offering them, the story is different when the lender's objectives are towards the profit generated from the consumer accepting the offer.

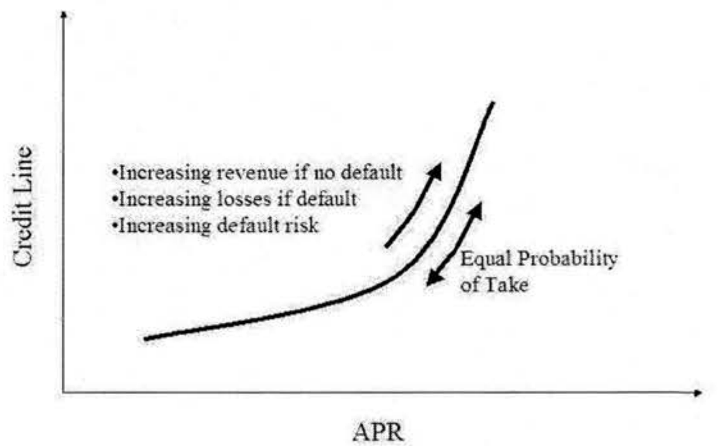

Figure 2.2: Lender's profits are different along the consumer iso preference curve. Figure based on Keeney and Oliver (2005)

In Fig 2.2, although the probability of a consumer accepting the offer is equal along the iso-preference curve, the revenue generated by the customer will increase when moving along the line from bottom left to top right. This is because the lender receives payments from a higher interest rate and it is applied to a larger loan amount. At the same time, the expected losses are also increasing because the probability of default is growing together with the amount of money that would be lost if default occurred. In summary, Keeney and Oliver argued that "a contribution to the expected profit is initially small, increases to a single high point and then decreases monotonically along any individual iso-preference curve".

So along each iso-preference curve of the consumer, there should be a point indicating the maximum expected profit generated by the consumer for the lender if the consumer takes the offer. As shown in the Fig 2.3 below, the points A and B are both the points that yield maximum expected profit while A' and B' are the points giving less profit to the lender than point A and B.

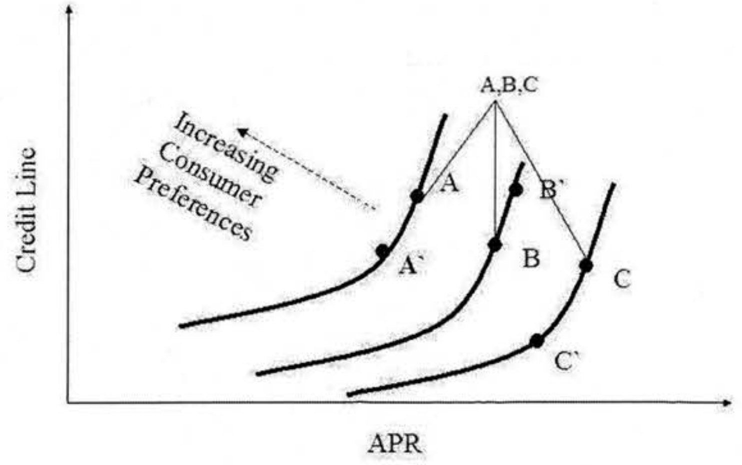

Figure 2.3: Offers yielding maximum expected profit for the lender. Figure based on Keeney and Oliver (2005)

Keeney and Oliver also argue that if we compare the maximum expected profit conditional on acceptance of the loan by the applicants of different APRs, the expected profit is low, rises, and reaches a maximum and then declines. The reason is that at low APRs little interest is received, at higher APRs more interest is earned with the probability of default rising. At still higher APRs the probability of default is so high that expected profit falls. Thus:

$$Expected\ Profit\{given\ take\ A\} < Expected\ Profit\{given\ take\ B\}$$

and

$$Expected\ Profit\{given\ take\ C\} < Expected\ Profit\{given\ take\ B\}$$

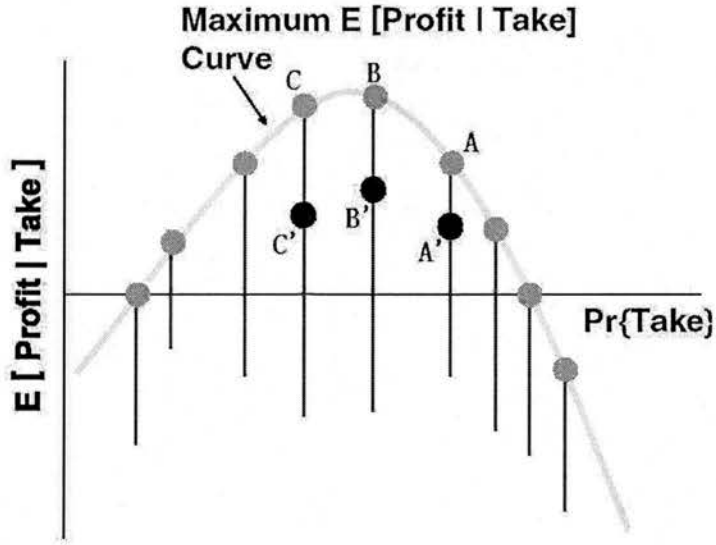

Figure 2.4: Maximum expected profit at each probability of take. Figure based on Keeney and Oliver (2005)

Since all points on each iso-preference curve indicate the same probability of taking the offer for the consumer, Fig 2.3 could be mapped into the Fig 2.4, which shows the maximum expected profit at each probability of take. Note that:

$$Prob\{take\ A\} = Prob\{take\ A'\} > Prob\{take\ B\} = Prob\{take\ B'\}$$

The y-axis shows the expected profit conditional on taking the offer

$$Expected\ Profit\{given\ take\ A\} > Expected\ Profit\{given\ take\ A'\}$$

$$Expected\ Profit\{given\ take\ B\} > Expected\ Profit\{given\ take\ B'\}$$

By multiplying the probability of take by the conditional maximum expected profit given the probability of take, unconditional expected profit can be derived, as shown in the Fig 2.5. Keeney and Oliver have assumed the profit is zero when the offer is not taken and are not explicitly taking the costs of acquisition into account. In Fig 2.5, the curve where offer R,S and T reside is the unconditional expected profit, while the other curve is for the conditional expected profit given that the offer has been taken. Any points under the unconditional expected profit curve are contributing less profit to the lender and therefore not desirable to the lender in terms of profitability.

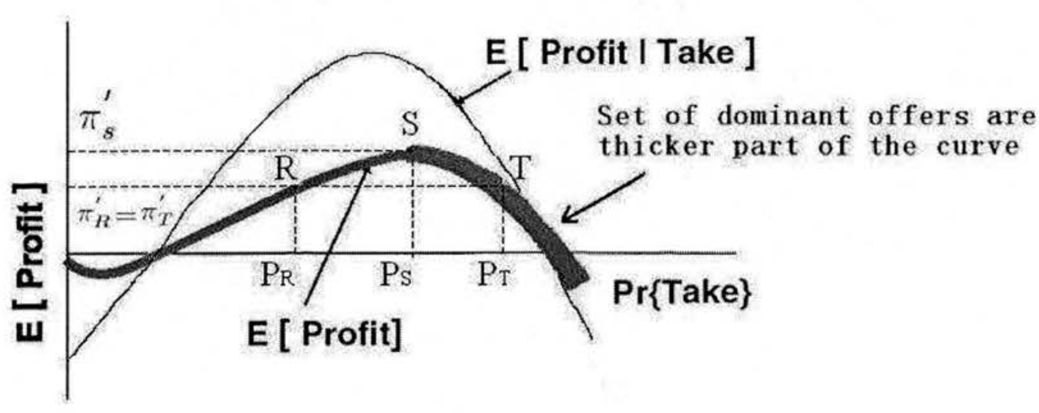

Figure 2.5: Unconditional expected profit. Figure based on Keeney and Oliver (2005)

The lender's utility is assumed to depend partially on market share and revenue. The dominant set of offers is shown in Fig 2.5 and consists of those on the thicker part of the curve. For each point not in the dominant set, we can always find an offer in the dominant set to beat it with higher probability of take and getting similar expected profit for the lender. Thus point T is preferred by the lender to point R.

Combining the utility function of the consumer with the profit function for the lender in a CreditLine-APR space, Figure 2.6 shows the iso-profit contours for the lender as well as the iso-perference curves for the customer. Note the expected profits for the lender at point Q,R,S,T,U,V are $\pi_Q^{'}, \pi_R^{'}, \pi_S^{'}, \pi_T^{'}, \pi_U^{'}, \pi_V^{'}$. The offer S will bring the lender the highest expected profit $\pi_s^{'}$ while the offer V, as it is located further away from the zero-profit-contour (where offer U and Q reside), will bring the lender negative expected profit. Offer T and R, residing on the same iso-profit contour, will bring the same amount of expected profit for the lender. The size of those expected profits is compared below

$$\pi_s^{'} > \pi_T^{'} = \pi_R^{'} > \pi_U^{'} = \pi_Q^{'} = 0 > \pi_V^{'}$$

Also note that the acceptance probabilities for the customer at point Q, R, S, T, U are P_Q, P_R, P_S, P_T, P_U. Since it has been assumed that the customer will prefer a lower rate and a higher credit line, the acceptance probabilities are increasing from Q to U

$$P_Q < P_R < P_S < P_T < P_U$$

The set of points that start from S linking T U and V is the set of dominant offers for the lender. Those are called dominant offers because for whatever offer that is not residing on this line of dominant offers, an equally profitable offer can be found on the line of dominant offers by moving the offer along the iso-profit contour towards the point which is tangent to the iso-preference curve that has the highest probability of acceptance.

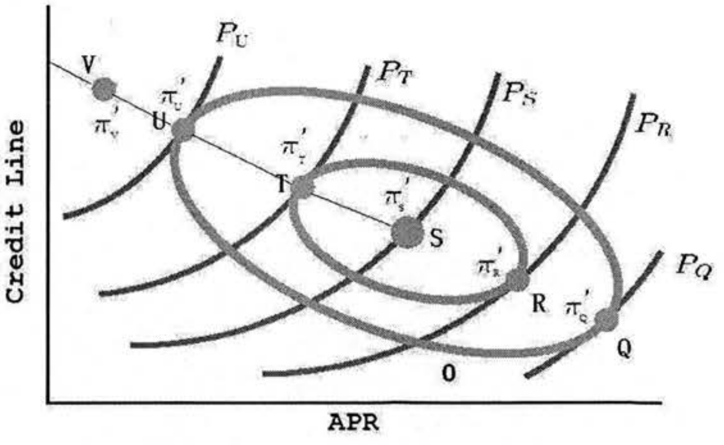

Figure 2.6: Iso-profit contours for the lender in the consumer iso-preference map. Figure based on Keeney and Oliver (2005)

For example, to the lender the offer R is equally profitable as the offer T since they are both on the same iso-profit contour. However, T is at the tangency point between this given iso-profit contour and the iso-preference curve with highest probability of acceptance ($P_T > P_R$). Provided the lender gains utility from both greater market share and more profit (assuming the lender is not extremely risk averse), the lender will always favour the offer T than offer R because the latter one means lower market share and equal profits.

2.1.3 Set of win-win offers

As discussed previously, any offers, like point M in Figure 2.7 below sitting on the iso-preference curves below where point S resides, are less desirable to the consumer than point S.

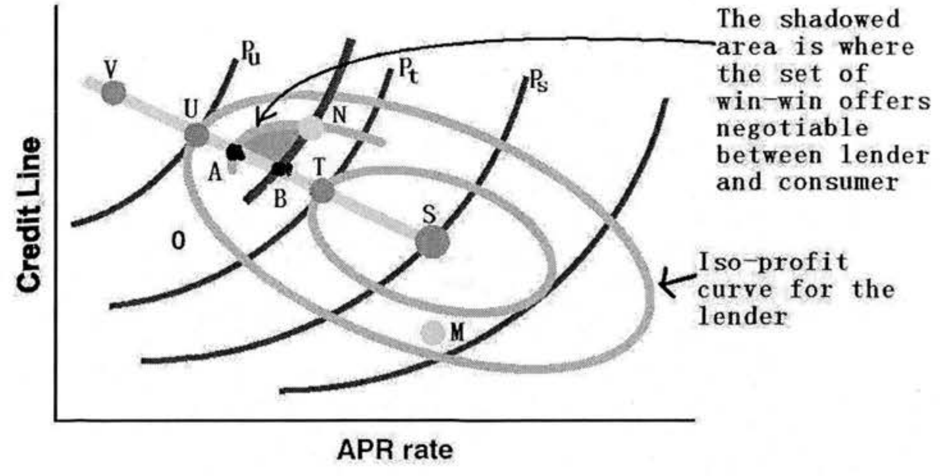

Figure 2.7: Set of win win offers. Figure based on Keeney and Oliver (2005)

The case for offer N, is different. From the lender's perspective, N is not profit maximizing. The shadowed area between A, B and N is the place where the lender and the consumer can negotiate an offer price of credit line and rate that is more preferable to both parties. Moving along the lender's iso-profit curve from N to A, the lender keeps the same amount of profit and the consumer receives an offer more acceptable. Following a different path along the consumer's iso-preference line from N to B, the consumer is indifferent to the changes but the lender will see an increasing profit until the arrival of offer B.

The lender wishes to be at the points which are on the tangency between lender's iso-profit contors and consumer's iso-preference curves. For any given profit, such combinations maximise probability of take. In the Fig 2.7 above, these tangencies form the line S-T-U-V. On the other side, the consumer wishes to be at the points that are as top-left as possible in the iso-preference space.

Since the lender is in the position of making the offers, the offers made to the consumer from a profit maximizing lender are the points along line S-T-U. The consumer wishes to move top-left towards point V while the lender prefers the other direction towards point S. An agreed deal is likely to lie somewhere on that line. The offers that are on the line from U to V are not likely to be considered by a profit-maximizing lender as they are making a loss. The offers on the line between U and S have the lender's expected market volume maximized (maximized probability of take by a typical applicant), conditional on not making a loss or making a given amount of profits. In general the expected profit earned by the lender can be expressed as

$$Expected\,Profit = p(accept|offer)\,\{L(offer)p(G|offer) - D(offer)p(B|offer)\}$$

where $L(offer)$ is the profit for the lender when the consumer that takes the offer is a good customer in the sense of not defaulting [2] and $D(offer)$ is the loss for the lender should the consumer take the offer and then default. $p(G|offer)$ and $p(B|offer)$ are the corresponding conditional probabilities of these good and bad cases.

2.1.4 Lender's preferences

The lender makes his decision by valuing trade-offs between profits and market share, selecting an optimal offer from the set of win-win offers outlined in the sections above. Keeney and Oliver described this situation using utility functions and letting a lender's objective be to maximize the expected utility of the business $u(\pi, s)$, π for profits and s for market share. Keeping u constant and letting π and s vary, the iso-preference curves for the lender can be plotted, as in the below figure.

[2] A customer not defaulting may also pay back the loan early and therefore not a good one in terms of profitability for the lender

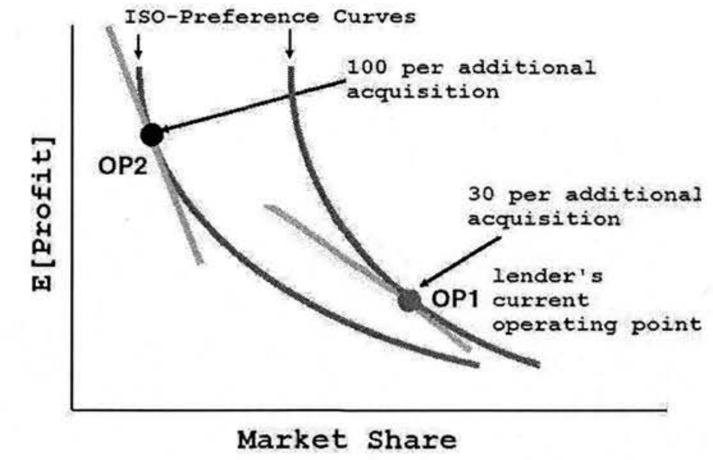

Figure 2.8: Lender's iso-preference curves. Figure based on Keeney and Oliver (2005)

The OP1 point in the figure above shows the lender's current operating point, indicated by its current market share and expected profits. The lender's current trade off value of market share and profit can be found by calculating the slope of the tangent line to the lender's iso-preference curve at point OP1. OP2, compared to OP1, on another lender iso-preference curve, implies higher expected profit and a much higher trade off value (in terms of expected profit) for each additional customer.

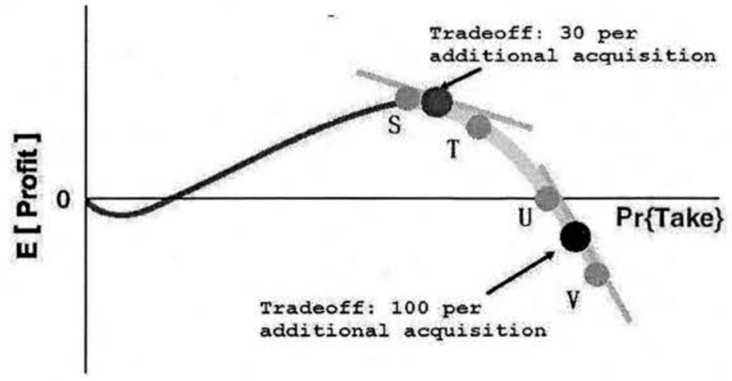

Figure 2.9: Trade off value per acquisition. Figure based on Keeney and Oliver (2005)

For a specific consumer, the optimal offer for the lender to make depends on its current operating point and current preference. If profit is the only target, the value trade off between profit and customer number is 0 dollars for each additional customer. Then the offer S in Figure 2.9 is the optimal choice. If the lender is operating at OP2 and working on increasing its market share then the offer at U or between U and V in Figure 2.9 is the optimal offer it can make.

A weakness of the model of Keeney and Oliver is that it does not take the competition between lenders into consideration. Blochlinger and Leippold (2006) simulated the competition between lenders but many assumptions they made results in oversimplification. A lender's best strategy depends on other lenders' market position and business objectives. The lenders may seek a Nash equilibrium in a mature market. If a lender wishes to maximize the profit only without considering itself and other lenders' market position, it will see itself squeezed out of the market due to relatively high price. If a lender is eager to enlarge market share without considering itself and other lenders' profitability, even at a risk of accepting zero or negative profit loan requests since profitable customers are hard to attract, it will see itself accumulating too much risk, placing itself in an adverse position in a downward economic cycle. Finally this model assumes lenders are risk neutral.

2.2 Previous research in acceptance modelling

Casual observations shows that current competition between lenders is intense. Efforts have to be made to attract new customers and retain them afterwards. Lenders are building various models to predict customers' acceptance behaviours such as whether to respond to marketing mail, whether to accept an offer of a credit product or whether to switch to other lenders.

Jung et al. (2003) modelled the likelihood of consumers accepting student bank accounts when being given different offers. Those offers had six features, including 5 choices of overdraft limit, 4 choices of credit card options, fee for foreign currency, discounts on insurance, interest paid on account surplus and 10 choices of free gifts. Their data set, named the Fantasy Student Current Account(FSCA), was gathered from a dedicated website, which was (and is) widely publicized to first year students at the University of Southampton with prize winning draws as enticement.

Using those hypothetical six offers and 18 applicant characteristics of 331 web participants, they estimated the probability of acceptance for each offer characteristic using three different modelling approaches, logistic regression, linear programming and an accelerated life model approach. Because of the particular nature of the samples of undergraduate students and the possibility of the 'testing effect' [3] of the data collected, the results may not be generalizable. In addition, the consumer's iso-preferences are not explicitly estimated.

Seow and Thomas (2005) modelled the probabilities of an applicant taking different offers using decision trees. Their analysis was based on the same data set used by Jung et al. (2003). A two layered decision tree structure was used whereby the enforced upper layer used applicant characteristics only and the lower layer used only offer characteristics. This structure offers the convenience for the lender to build an adap-

[3] The students were making choices towards hypothetical products rather than real ones.

tive application process by asking a customer about the applicant's characteristics first and afterwards providing the offer that is the mostly likely to be taken by this customer.

Different tree settings were examined with and analysed in their paper. These were an applicants characteristics only tree, an offer characteristics only tree, both types of characteristics and even with more flexibility allowed in the tree structure (so called alternate best tree) to generate a better fit to the data. They also explored the situation when imposing a limited number of questions asked as a restriction on the tree building process.

2.3 Previous research in profit scoring

Many issues arise when implementing profit scoring systems. The first one is how to build a fully integrated information system to identify and capture profit related information such as transactions, the merchant service charge for each account and how to aggregate them together. Other decisions are

- Should profit be measured for each product individually or calculated in total for all the products put together? Counting all products' profit considers the cross selling marketing opportunities that could be neglected when measuring individual product profit.

- Economic conditions. Crook et al. (1992) explored the differences observed in the cut-off scores and functions estimated when using data for different years. Their results showed the importance of economic condition changes over a business cycle and called for careful attention from the credit grantors. Bellotti and Crook (2007b) demonstrated that including macroeconomic variables in survival analysis models as time-varying covariates significantly boosted the predictive performance of the default compared with logistic regression.

- How to maximize profit using default based score? Subsection 2.3.1 will discuss that in detail.
- How much to charge to maximize profit without losing the customer by charging too much? Subsection 2.3.3 will discuss that in detail.
- The implication of the timing of the profits. To calculate the exact amount of expected profit, not only the propensity for each customer to default will be needed, but the timing of the defaults will also be important. Apart from defaults, the timing and likelihood of early repayment behaviour is also crucial in the profitability calculation. Subsection 2.3.4.2 will discuss the survival analysis in detail.

2.3.1 Maximize profit using default based score

Following the approach of Marshall and Oliver (1995) and Oliver (1993), Thomas et al. (2002) described how to make accept and reject decisions to maximize profit based on a traditional default credit scoring system and how to maximize the expected profit for a portfolio. They assumed that the profit from a consumer R is 0 when he is rejected. If the account is accepted and becomes good, a fixed amount of profit L is gained for the lender. A fixed amount of loss D is incurred for the lender when the customer defaults after being accepted. The expected profit $E(R|s)$ for a customer with score s is then:

$$E(R|s) = Lp(G|s) - Dp(B|s) - Cost = (L+D)p(G|s) - D - Cost \tag{2.1}$$

where $Cost$ is fixed cost per customer. $p(G|s)$ is the conditional probability that a customer with credit score s will be good. $p(B|s)$ is the conditional probability that customer with score s will be bad. Therefore $p(B|s) = 1 - p(G|s)$ The profit maximization decision to accept this customer can be derived through from equation 2.1 by setting $E(R|s) >= 0$, implying a customer is $p(G|s) >= \frac{D+Cost}{L+D}$

The total profit expected for the whole customer population that was accepted is

$$E^*(R) = \sum_{s \geq c} (Lp_G p(s|G) - Dp_B p(s|B)) \tag{2.2}$$

where p_G and p_B are the probabilities of good and bad respectively. Here the fixed cost is ignored. Assume $p(G|s)$ is monotonically increasing with s. The cut off value c is the score where for all the scores $s >= c$, $p(G|s) >= \frac{D}{D+L}$

2.3.2 Relationship between Profit, Volume and Loss in a portfolio

Oliver and Wells (2001) have discussed the effect of different cut off policies on the expected profit and volume as well as on loss. If all applicants with a score above s_C are accepted we can write:

$$\begin{aligned}
Expected\ fractional\ Volume\ E[V(s_C)] &= \int_{s_C}^{\infty} f(s)ds = 1 - F(s_C) \\
Expected\ Loss\ E[L(s_C)] &= \int_{s_C}^{\infty} Dp(B|s)f(s)ds = Dp_B(1 - F_B(s_C)) \\
Expected\ Profit\ E[P(s_C)] &= \int_{s_C}^{\infty} (Lp(G|s) - Dp(B|s))f(s)ds \\
&= Lp_G(1 - F_G(s_C)) - Dp_B(1 - F_B(s_C))
\end{aligned}$$

where $f(s)$ is the density function of score and $F(s_C)$ is the proportion of scores below the cut off score s_C. With no other constraints, the expected profit can be maximized

$$\begin{aligned}
Max_{s_C} E[P(s_C)] &= Max_{s_C} \int_{s_C}^{\infty} (Lp(G|s) - Dp(B|s))f(s)ds \\
&= Max_{s_C} \int_{s_C}^{\infty} Lp(B|s)(w(s) - \frac{D}{L})f(s)ds
\end{aligned} \tag{2.3}$$

where $w(s)$ is the odds for score s. $p(B|s)$ is monotonically increasing with s. L and $f(s)$ are both positive. The unconstrained optimal cut off score s_C is found when

$$w(s_C) = \frac{D}{L}$$

When the lender wants to minimize expected losses with a certain amount of expected profit as the lower bound, the problem can be written as:

$$Min_s\ E[L] = Min_s\ Dp_B(1 - F_B(s))$$

Subject to

$$\lambda : Lp_G(1 - F_G(s)) - Dp_B(1 - F_B(s)) \geq P_0$$

where λ is positive and called the shadow price. Using non-linear programming (Kuhn-Tucker conditions) to solve the optimality equations, the shadow price λ is

$$\lambda^* = \frac{1}{\frac{w^*}{w} - 1} \geq 0$$

where w^* is the optimal cutoff odds for a constrained problem and w is the optimal cutoff odds for an unconstrained profit maximizing problem. This can be illustrated in Figure 2.10, where the efficient frontier is the set of points forming the solid line. Moving along this efficient frontier in an anti-clockwise direction both expected profit and expected losses go up until reaching the intersection point with the dotted line. This intersection point is where the cutoff score equals $\frac{D}{L}$. After that point, the lowered cutoff scores bring more bads which means more loss. The expected profit decreases while the expected losses continue to increase. Therefore, the operating points on the solid line make up the efficient frontier which is optimal.

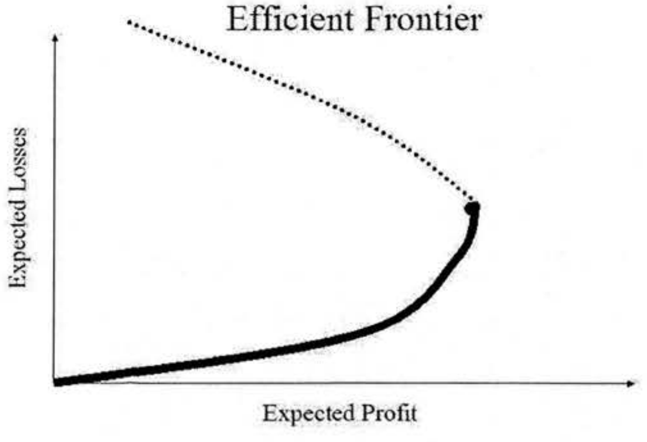

Figure 2.10: Efficient Frontier when the objective is to minimize expected loss subject to given expected profit, Figure based on Oliver and Wells (2001)

Similarly, Oliver and Wells show that non-linear programming can be used to solve the optimality equations when the lender has an optimizing target to maximize the expected profit subject to a lower bound on the expected volume.

$$Max_s\, E[P(s)] = Max_s\, Lp_G(1-F_G(s)) - Dp_B(1-F_B(s))$$

Subject to:

$$\mu:\ (1-F(s)) \geq V_0$$

This can be illustrated in Figure 2.11, where the efficient frontier is the set of points forming the solid line. Moving along this efficient frontier in clockwise direction expected profit decreases while expected volume increase. The operating points on this efficient frontier are always satisfying the minimum volume constraint set above when maximising the expected profit.

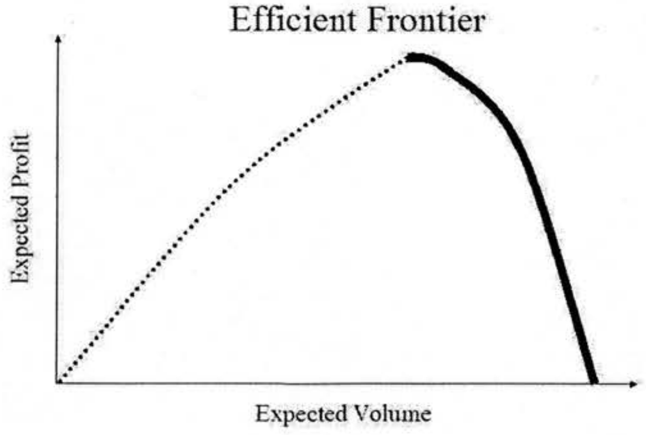

Figure 2.11: Efficient Frontier when the objective is to maximize expected profit subject to given expected volume, Figure based on Oliver and Wells (2001)

Using non-linear programming techniques, other business objectives can be incorporated in as extra constraints. For example, if the lender wants to add another constraint of minimum market volume V_0 when minimizing expected loss subject to minimum expected profit P_0.

$$Min_s\, E[L] = Min_s\, Dp_B(1 - F_B(s))$$

Subject to

$$\lambda : Lp_G(1 - F_G(s)) - Dp_B(1 - F_B(s)) \geq P_0$$

$$\mu :\ (1 - F(s)) \geq V_0$$

Similar Kuhn Tucker conditions are

$$-Dp_B f_B(s) - \lambda(-Lp_G f_G(s) + Dp_B f_B(s)) - \mu(-f(s)) = 0$$

$$\lambda(Lp_G(1 - F_G(s)) - Dp_B(1 - F_B(s)) - P_0) = 0$$

$$\mu(1 - F(s) - V_0) = 0$$

$$\lambda \geq 0$$

$$\mu \geq 0$$

If $\mu = 0$ and $\lambda = 0$, $Dp_B f_B(s)$ has to be 0, which cannot be true. If $\mu = 0$ and $\lambda > 0$, $\lambda = \frac{Dp_B f_B(s)}{Lp_G f_G(s) - Dp_B f_B(s)}$ If $\mu > 0$ and $\lambda = 0$, $\mu = \frac{Dp_B f_B(s)}{f(s)}$. If $\mu > 0$ and $\lambda > 0$, the conditions will be invalid unless both equations $Lp_G(1 - F_G(s)) - Dp_B(1 - F_B(s)) - P_0 = 0$ and $(1 - F(s) - V_0) = 0$ are satisfied. Under such conditions, the cut off that yields minimal expected profit P_0 will also bring minimal market volume V_0.

Beling et al. (2005a) continued the discussion of the optimal scoring cut off policies based on the trade-off between the lender's multiple business objectives, with whom the relationship to the receiver operating characteristic (ROC) curves have been illustrated. After presenting the policies to adopt for a single scorecard or dominant scorecards, they showed policies for those with two scorecards, none of which is dominating. In the presentation given by Beling et al. (2005b), the risk-neutral assumption had been replaced with various risk-averse assumptions in the study of optimal portfolio selection policies.

All the derivations depend on the assumption that we have exact information on the profit L for a good account and the loss D when the account is bad. Also they are not assumed to change with the score s, which is more likely to happen when risk based pricing is applied. Generally in risk based pricing, higher scores will be given lower interest rate charges and lower scores (more risky customers) will be charged with higher interest rates. The implication under such situations will be discussed in later subsection 2.3.3.

2.3.3 Risk based pricing

Often a lender charges customers the same fixed interest rate and rejects customers with poor credit scores that are below a cut off value to control the risk. This is commonly used for credit cards. Recently lenders have been separating customers into different groups using credit scoring techniques. The customers having the highest credit score are thought by the lender to be the lowest risk and are offered the lowest interest rates. Those customers without good scores are accepted anyway if the scores are higher than the cut off value. But they are not given such a low interest rate.

Thomas et al. (2002) used the example below to show how to set risk based interest rates according to the credit score. Assume in a scoring system the application characteristics x will be given a score s. p_G and p_B are the proportion of goods and bads in the whole population. $p(s)$ is the proportion of the population that has score s. $p(s|G), p(s|B)$ are the conditional probabilities. Now

$$p(s) = p(s|G)p_G + p(s|B)p_B$$

and the probability of being good at score s

$$p(G|s) = \frac{p(s|G)p(G)}{p(s)}$$

by Bayes Theorem. Likewise,

$$p(B|s) = \frac{p(s|B)p(B)}{p(s)}$$

Also assume that the interest rate i charged is a function of the credit score s, noted as $i(s)$. Assume the cost will be a constant D when the customer defaulted no matter what the interest rate charged. The profit from a good customer, $L(i)$, depends monotonically on the interest rate i charged. The lender just needs to decide whether to accept customers at each score s and what interest rate i to charge.

Denote the probability that customers with credit score s takes the offer with interest rate i as $a_s(i)$. $a_s(i)$ is decreasing in i. For a score s, highest expected profit can be obtained by solving the equation below

$$\max_i \{(L(i)p(G|s,i) - Dp(B|s,i))a_s(i), 0\} \tag{2.4}$$

The optimal interest rate i for a score s yielding the maximum expected profit can be found by differentiating equation 2.4 with respect to i and setting the derivative to 0.

$$-L^{'}(i)p(G|s,i)a_s(i) - (L(i)p^{'}(G|s,i) - Dp^{'}(B|s,i))a_s(i) = (L(i)p(G|s,i) - Dp(B|s,i))a_s^{'}(i) \tag{2.5}$$

As a special case with many simplified assumptions, assume interest rate i is the only factor that influences a customer's acceptance behaviour and $a_s(i) = e^{-\alpha(s)(i-i^*)}$. The customer with score s will be exponentially less likely to accept an offer of interest rate i when it is larger than i^*. Also assume $L(i) = \frac{R}{(1+i)^T} - \frac{R}{(1+i^*)^T}$. The payment is R during time T charged interest rate i, at the cost of funds interest rate i^*. Further assume $p(G|s) = p(G|s,i)$. Interest rate i has no effect on the odds of goods given the score. Then equation 2.5 becomes

$$\frac{p(G|s)}{(1+i)^T}\left(\alpha(s) + \frac{1}{1+i}\right) = Dp(B|s)\alpha(s) \tag{2.6}$$

Solving equation 2.6 will give the optimal interest rate i to charge for customers with credit score s.

To get to this result, Thomas et al. (2002) implicitly made a lot of assumptions, which, may not necessarily hold. The first assumption is that the loss D, which is assumed constant across all considerations. Banasik et al. (1999) have estimated expected time to default as well as the expected time to pay off early using survival analysis techniques. The earlier a customer of a fixed term loan product defaults, the less payments will be received by the lender, therefore the greater the losses incurred for the lender. Their estimation of pay off early time also invalidates the assumed form of profit term $L(i)$, which not only increases with the interest rate charged but also varies with the time for the customer to pay off early. The earlier the customer pays off or switches to other lender, the less profit for the lender. Considering the high attrition rates due to the current extremely competitive market, the inclusion of time to pay off early in the function of profit $L(i)$ cannot be omitted.

A practical argument against the implementation of risk based prices according strictly to the score is that ill-intentioned fraudsters may be able to work out the mappings between interest rates and credit scores. The latter are and should be kept secret during operations.

In a simulation study Blochlinger and Leippold (2006) compared three different lending strategies. The first one was a policy which selects a threshold cut off point on the ROC curve and striking a zero profit. The second was a risk based pricing strategy where the risk premium was linked with the credit score rather than a constant. The third one was called the mixture regime but in fact was largely a risk based pricing policy with the risk premium rounded towards the next quarter of a percentage point. They simulated the competition in the loan market where 3 lenders fight for profit and market shares when employing different lending strategies. The difference between them is the predictive ability (quantified using AuROC) of the credit rating methods the lenders are using. The results were not surprising. The better the scoring model the more economic benefit and market share for the lender. The significance of the benefit was more evidenced when risk based pricing oriented strategies were used across lenders rather than a cut off strategy.

2.3.4 Other profit scoring approaches

Four different approaches to profit scoring have been considered in the literature. The first one is to build indirect score cards separately for each profit related variable such as default, acceptance, attrition and usage. Then combine the intermediate information together to determine a final decision (see Li and Hand (2002)). But this approach is open to the criticism that indirect scoring may propagate errors from the estimation of the intermediate stage model to the final decision. Compared with such indirect approaches, a reversed approach is to directly regress the profit on explanatory application variables.

The second one is to directly regress the profit on a linear function of the predictor variables. One recent example is Somers and Whittaker (2007), which used linear and kernel smoothed quantile regressions to model the revenue on a credit card portfolio and loss given default on a mortgage portfolio.

The third approach, the Markov chain approach will be discussed in section 2.3.4.1. The fourth approach, survival analysis will be discussed later in section 2.3.4.2 .

2.3.4.1 Markov Chain approaches

Cyert et al. (1962) developed a Markov chain process model to describe the behaviour of current accounts to estimate profit related variables like loss expectancy rates and allowances for doubtful accounts. Liebman (1972) formulated the credit control problem using an infinite horizon Markov decision model to model the transition probability between customer states. The customer state model is optimized in terms of minimizing total credit costs using linear programming techniques after the definition of the cost matrix is formulated.

Frydman (1984) argues that a mover-stayer model, a special mixture of two independent Markov chains, one for the 'stayer' in which the transition probability matrix is equal to the identity matrix and the other for the 'mover', in which the transition probability matrix is a normal one, describes the dynamics in payment states. She presented a maximum likelihood procedure to estimate the parameters of the mover-stayer model. Then, Frydman et al. (1985) compared stationary and non-stationary Markov chain models with the mover-stayer model. They applied those models to data of retail revolving credit accounts and found that the mover-stayer model provided a better description of the data when dealing with a heterogeneous population of credit accounts.

Till and Hand (2001) modelled repayment behaviours of credit card customers using two kinds of Markov chains, stationary model and mover-stayer model. They found most accounts stay in the state of being up to date from period t to $t+1$. For those who do not stay, the chance that they miss a further payment in $t+1$ goes up and not levels off until state 5. They also showed that although mover-stayer model describes data better than the stationary model, first order MCs may not be appropriate.

Trench et al. (2003) designed and implemented a system using Markov decision processes to make decisions on whether to grant offers to reduce APR or to increase credit lines. Their model utilized account level historical information on credit card customers and estimated probabilities of the customer transferring from his/her current state to other states using a transition matrix. After the estimation of the transition matrix, a set of actions that will maximize the expected future profits in the future 36 months are selected using a recursive calculation method.

Ho et al. (2004) applied Markov chain models on a large sample of current account data. They found that a first-order Markov chain is not appropriate to fit the data and describe the customer behaviour. Instead, they applied higher order Markov chains on the individual segments to address the non-homogeneity in the data yielding scorecards that perform better than normal application based scorecards.

2.3.4.2 Survival analysis

Survival analysis is one of the statistical techniques widely used in medical research and also in analysing system reliability. Survival analysis answers the question of when certain events occur rather than just how likely they are to happen, which has traditionally been the aim of credit scoring.

For each individual case we will record its time to the event happening(a 'failure' or 'default' for example) or no such events because of censoring. When the observation is censored, the only information we can infer is that the time to such an event is greater than our observation time period. Denote T as the time of the event, the survival function $S(t)$ can be expressed as

$$S(t) = p(t \leq T)$$

A hazard function $h(t)$ is defined as the event rate at time t conditional on that the event has not happened until time t or later.

$$h(t) = \frac{p(t \leq T < t + dt | t \leq T)}{dt} = \frac{f(t)}{S(t)} = \frac{S'(t)}{S(t)}$$

where $f(t)$ is the density function $f(t) = S'(t) = \frac{p(t \leq T < t + dt)}{dt}$. The three functions, density function, hazard function and survival function are interchangeable in describing the time distribution in survival analysis.

The survival function $S(t)$ can be modelled parametrically using an Exponential distribution $S(t) = e^t$ or Weibull distribution $S(t) = e^{-(\lambda t)^k}$. Log-normal or log-logistic models have also been tried. Kaplan and Meier (1958) suggested a non-parametric approach (K-M estimator) to estimate the survival function.

Cox (1972) proposed regression models to analyse the relationship between survival time and explanatory variables $x = (x_1, x_2, \ldots x_p)$. The hazard function is

$$h(t) = f(x)h_0(t) \tag{2.7}$$

where $f(x) = e^{w \cdot x}$ and w is a corresponding vector of parameters to be estimated. $h_0(t)$ is called the baseline function and takes the form a certain time distribution. In Accelerated Life models, the hazard function is in the form of

$$h(t) = f(x)h_0(f(x) * t) \tag{2.8}$$

where $f(x) = e^{w \cdot x}$. The difference between the two models is that in accelerated life models explanatory coefficients w and variables x together are interacting with the time variable t in the baseline function h_0 so that explanatory variables x can accelerate or decelerate the ageing of the subject studied. For the Proportional Hazard model, the ratio of hazard for i and hazard for j is independent of h_0

$$\frac{h_i(t)}{h_j(t)} = \frac{e^{w \cdot x_i}}{e^{w \cdot x_j}}$$

Narain (1992) applied the Accelerated Life Exponential Model to loan data and showed that estimated survival time could be be used to support a better credit granting decision. Banasik et al. (1999) applied three types of Proportional Hazards models and an Accelerated Life model to data for a personal loan and compared their results with logistic regression approaches. The results suggest PH models are competitive against logistic regression in predicting default probabilities. They also showed how competing risks (propensity to pay off early and propensity to default) can be accommodated in credit scoring systems.

Bellotti and Crook (2007a) introduced macroeconomic variables in survival analysis as time-varying covariates. Their results confirmed the influence of macroeconomic factors on the probability of default and showed that their inclusion did improve the predictive performance of default.

Andreeva et al. (2005) applied a Proportional Hazard Cox model to data for a revolving store card product. When the customers made further purchases, this additional information was taken into account to enhance the models. They also reported the different behaviour patterns observed between the Good and the Bad segments and within the Bad segments as well.

On the data of the revolving credit card across three European countries, Andreeva

(2006) compared survival analysis models (parametric and Proportional Hazard Cox models) with the widely used Logistic Regression. She observed a similar predictive performance across countries from those models.

Andreeva et al. (2007) combined a survival probability of default and the survival probability of a second purchase using data relating to the store card in Germany to form a survival combination model using OLS regression in a second stage to fit the net revenue. This profitability oriented approach was shown to accrue more profit than a logistic regression score optimized to minimize default risk only. However, it also came at a price of accepting more defaults.

Stepanova and Thomas (1999) made improvements in the application of Cox's Proportional Hazards model to build credit-scoring models. They used a coarse-classifying approach for characteristics, explained how residual tools can be used to check model fitness, expanded the Cox PH model by including time-by-covariate interaction. Stepanova and Thomas (2001) furthered the modelling with their application of survival analysis in behaviour scoring. Their results showed the scores from their model are competitive compared with logistic regression and yet provided more information crucial to calculate expected profit.

Stepanova and Thomas (2001) gave example equations for calculating expected profit at the application time and month K, assuming the products sold were personal loans and the survival probability estimated to month i are S_i.

$$Profit(Application\ Time) = \sum_{i=3}^{T+2} S_i \frac{a}{(1+r)^{i-2}} - L$$

where a is the monthly repayment amount or instalment, L is the amount of the loan and T is the term of the loan and r is the monthly interest rate for interbank lending. Similarly the expected profit at month K is calculated as

$$\begin{aligned} Profit(Month\,K) &= \sum_{i=1}^{T+2-K} S_i \frac{a}{(1+r)^i} - (1+r)^K * \sum_{j=K}^{T} \frac{a}{(1+r')^j} \\ &- (B_K - (T-K)a) \end{aligned}$$

where r' is the monthly interest paid by the consumer and B_K is the actual balance at month K. This is quite a step forward towards profit scoring, although more complicated formulae should be considered to account for the estimation of both time to default and time to pay off early. The two events is quite different, in a competitive personal loan market, early repayments can be 10 to 20 times more likely to happen than defaults. On the other hand, the potential loss from default for the lender is much bigger in amounts than the potential loss of revenue due to early repayments.

When considering models of the more than one type of event of interest, the approach of modelling them in a competing risks context has been tried. Lambrecht et al. (2003) studied a special UK mortgage data set and built a bivariate competing hazards duration model to analyse the time to voluntary possessions or forced processions during a number of years when economic conditions were changing. Although claiming the model to be a competing risk model, their main assumption was that the two random time events are independent and therefore wrote the joint density as the product of marginal densities in their parametric formulation. By investigating the results, they identified the variables that are most important to the lenders and borrowers when making their own foreclosure decisions accordingly.

Statistical methods are not the only approaches that can be used to estimate the timing of the events, Baesens et al. (2005) have investigated neural network models as alternatives for survival analysis. Based on their analysis comparing the predictive performance (on defaults as well as on early repayments) of a neural network and that of other predictors including a Cox proportional hazards model and a logistic regression model, they concluded that the improvement achieved through the neural network model is marginal.

2.3.5 Profit affected by Basel II

Since 1988, the Basel Committee on Banking Supervision at the Bank for International Settlements introduced a capital measurement system (Basel I) that required banks to hold a fixed percentage of capital for their loans against possible loss. The limitations of Basel I led to the drafting of Basel II. The Basel II Accord mandates that the minimal capital required for a loan is a function of the LGD (loss give default), PD (probability of default) and EAD (exposure of default). Oliver and Thomas (2005) discussed what implications the introduction of regulatory requirements in Basel I and II will bring and compared the optimal profit-maximizing cut-off scores under the requirements of Basel I, II and before the Accords. Their model assumed the lender borrowed all the funds and all equity is the shareholder capital. C_B stands for per unit borrowing cost and C_Q is for the equity capital cost. The expected profit for a single account can be written as

$$E[P] = E[R] - E[L] - E[C_B] - E[C_Q] - C_F$$

where R is for revenue, L for the loss when default. C_F is fixed costs.

The expected profit for the portfolio of accounts with scores larger than the cut-off score S_C is

$$E[P_P|S_C] = r_L p_G F^c(S_C|G) - f_D p_B F^c(S_C|B) - r_B F^c(S_C) - r_Q f_D \int_{S_C}^{\infty} K(p(s))dF(S) - C_F$$

where r_L and r_Q are the interest rate charged for lending and capital correspondingly. p_G and p_B are the probability of good and bad. The optimal cut-off is obtained when

$$o_c = \frac{p_G f(s_c^*|G)}{p_B f(s_c^*|B)} = \frac{(f_D + r_B) + r_Q f_D K(s_c^*)}{(r_L - r_B) - r_Q f_D K(s_c^*)}$$

Using a numerical example they showed that under Basel I a higher cut-off score is needed compared with Basel 0. The optimal cut-off score under Basel II may be higher than under Basel I when the lender is taking high risk applicants by charging very high rates, otherwise Basel I is more restrictive than Basel II.

2.4 Conclusion

This chapter has reviewed previous research relevant to profit scoring. Following Keeney and Oliver (2005), we presented the preferences of the consumers in a two dimensional space, the consequence to the lender in terms of expected profit and how to locate the set of win-win offers. Next we described recent practical and theoretical work modelling acceptance probabilities, maximizing expected profit using default-risk-based scores and how risk based pricing may be implemented using default based scores. Afterwards brief introductions were given to different profit scoring approaches.

Chapter 3

Default Risk Modelling

3.1 Introduction

This chapter is dedicated to the modelling of the default risk. Section 3.2 reviews previous research work in default modelling methods and compares their performance. In section 3.3 and 3.4 modelling details of logistic regression and support vector machines (SVMs) are described. Section 3.5 explains why the area under ROC curves is used as our performance measure instead of the accuracy ratio. Section 3.6, which describes how the data is prepared is followed by section 3.7, reporting and comparing performance of logistic regression and SVMs on the default data. Section 3.8 investigates the bidirectional relationship between the probability of default and the interest rate. Finally, section 3.9 summarizes the findings in this chapter.

3.2 Previous research in default risk modelling

To facilitate faster and safer lending practice, lenders build credit scoring models to assess the risk of default (non-repayments). These models are designed and trained to discriminate between future applicants based on the observed performance of existing customers together with their application characteristics and the bureau data shared between lenders. Predicting a default or no default outcome is a typical binary classification problem.

Many statistical and non-statistical classification methods have been proposed and applied in the credit scoring literature. Those methods include discriminant analysis, logistic regression, mathematical programming, decision trees, neural networks, genetic algorithms, genetic programming, support vector machines and nearest neighbour methods. Combinations of different classifiers have also been tried. Comparative studies of those classification methods have been carried out in various papers (see Srinivasan and Kim (1987), Yobas et al. (2000), Baesens (2003), Ong et al. (2005) and Lia et al. (2006)), with the classification accuracy rate used as the performance indicator. Table 3.1 compares the performance of different classification methods using accuracy rate.

Newer and more complex methods, however, bring diminishing improvements, as Hand (2006) observed. Hand suggested some reasons why little improvement happens using much more complex models. The first is the flat maximum effect (Winterfeldt and Edwards (1982) Hand (1997)), where by adding additional variables little can be gained after equal weights of a linear predictor are carefully optimized. The second reason is population drift (Kelly et al. (1999)). One of the fundamental assumptions of credit scoring models is that the customer population distribution with respect to the risk of default is supposed to be stationary over time. That is without doubt unrealistic in practice. Customer behaviour will change because of constant changes in the exter-

Table 3.1: Compare the performance of different classifiers using accuracy, enlarged based on Crook et al. (2007)

Authors	Linear Regression	Logistic Regression	Classification Trees	LP	Neural Nets	GA/GP	SVM	Rough Set
Srinivasan and Kim (1987)	87.5	89.3	93.2	86.1				
Boyle (1992)	77.5		75	74.7				
Henley (1995)	43.4	43.3	43.8					
Yobas et al. (2000)	68.4		62.3		62.0	64.5		
Desai (1997)	66.5	67.3			66.4			
Baesens (2003)		79.2	79.9					
Ong et al. (2005)		80.795	78.2		81.72	82.805		79.145
Lia et al. (2006)					73.17		84.83	

nal economic environment. Customers acquired in this advertising campaign will also be different to those attracted in a next one due to dynamic competition in the market between the lenders.

Hand also pointed out some common questionable practices during the modelling process by the researchers. One of the problems is the mismatch between the optimization criteria and the performance assessment methods. A common example is the using of likelihood to select a model, followed by using the accuracy or misclassification rate to evaluate the model performance, and finally reporting cost weighted misclassification rate in practice.

Using the accuracy percentage alone is not a good indication of predictive performance as it fails to reflect the difference in the predictive ability towards different classes. Area under ROC curve (AuROC) is a more appropriate choice. Detailed discussion will be given in subsection 3.5.2. For a comparison of the credit scoring models whose performance is measured in AuROC, please see Baesens et al. (2003) and Bellotti and Crook (2007c). Besides, the datasets used in the some of the comparative studies may not reflect the real consumer credit data distribution faced by today's major lenders in the UK. Take the German Credit Data used in Ong et al. (2005) and Baesens (2003) for example, thirty percent of them are defined as 'bad' and others are 'good', which contrasts with typical UK consumer credit data with default rates at around or less than 5 percent.

In this chapter we will compare the performance of logistic regression and support vector machines in predicting default. Logistic regression is the classifier that is most commonly used by lenders and it has yielded consistently good predictive performance compared with other classifiers (see Table 3.1). Support vector machines are one of the most promising newer methods. The following two sections will introduce the modelling detail of logistic regression and support vector machines.

3.3 Logistic regression[1]

When the dependent variable to predict is binary or dichotomous , logistic regression may be a more appropriate model than linear regression models. Linear regression relates the explanatory (or predictor) variables to the dependent(or outcome) variable by the formula

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_n x_n$$

The dependent variable y could range from $-\infty$ to $+\infty$ if the values of x_n varies from $-\infty$ to $+\infty$. This cannot accommodate the data when the dependent variable is taking values between 0 and 1 only, although for application scoring this may not matter since only a ranking is required. However for the calculation of PDs (Probability of Default) for regulatory capital purposes, predictions outside the $[0,1]$ interval would be problematic. The logistic transformation solves that problem so that the dependent variable, p, ranges from 0 to 1 and so can be interpreted as the estimated probability.

$$g(x) = ln\left[\frac{\pi(x)}{1-\pi(x)}\right] = \beta_0 + \beta_1 x_1 + \ldots + \beta_n x_n$$

And the probability p is

$$p = \pi(x) = \frac{e^{g(x)}}{1+e^{g(x)}}$$

Maximum Likelihood Estimation methods are normally used to find the estimates of the parameters that will maximize the likelihood, the probability of observed data. The

[1]The introduction of logistic regression models follows Hosmer and Lemeshow (2000)

likelihood function for one observed data point (x_i, y_i), is

$$\pi(x)^{y_i}\,[1-\pi(x)]^{1-y_i}$$

Since the observed data are assumed to be independently collected, their joint probability can be written as

$$L = \prod \pi(x)^{y_i}\,[1-\pi(x)]^{1-y_i}$$

The parameters to be found to maximize L are also going to maximize $ln(L)$ as the log of L is monotonically related to L. In practice, it is much easier to find the maximum of $ln(L)$ or the minimum of $-ln(L)$.

Unlike linear discriminant analysis, logistic regression does not require that the covariates are normally distributed or that the covariates have an identical covariance matrix. When dealing with real data in which normality conditions are sometimes not met, logistic regression can cope well without the restriction of these assumptions.

3.3.1 Variable selection using step-wise selection

Hand (2006) shows that in a linear model, introducing additional variables may result in diminishing improvement in explanatory power. Adding very large numbers of variables may also result in over-fitting of the model. So we have run step-wise variable selection routines.

There are two directions to select variables, Backward selection and Forward selection. Backward selection starts from a complete set of variables and then tries to remove those insignificant ones step by step. Forward selection is working in the opposite direction by adding the most significant variables step by step. At each step of a Backward stepwise selection, an attempt is made to remove any insignificant (when their p value is smaller than a pre-set threshold p_{out}) variables from the model before adding a significant variable (when their p value is larger than a pre-set threshold p_{in}) to the model.

3.3.2 Existence of MLE(maximum likelihood estimation)

Albert and Anderson (1984) discussed 3 different situations under which the existence of a MLE solution of the logistic regression model depends. These 3 different data configurations are listed below.

- Complete Separation

  The data can be completely separated. There exists a vector b that can classify all cases into the observed classes correctly.
  $$\begin{cases} bx_j > 0 & Y_j = 1 \\ bx_j < 0 & Y_j = 2 \end{cases}$$
  In this situation, non-unique infinite estimates are given by the SAS logistic procedure.

- Quasicomplete Separation

  The data are not completely separable but there is a vector b such that
  $$\begin{cases} bx_j \geq 0 & Y_j = 1 \\ bx_j \leq 0 & Y_j = 2 \end{cases}$$
  In this situation, non-unique infinite estimates are also given by the SAS logistic

procedure.

- Overlapped

  Overlapped data configuration means there is no complete or quasi-complete separation existing in the sample points. In this situation, unique maximum likelihood estimates exist.

The problems of quasi-complete or complete separation of data points are typically encountered when the sample size of data is small. This is later evidenced in the logistic regression results for a subsample of our data.

3.4 Support vector machines

Support vector machines (SVMs) are becoming one of the most promising learning methods used for classification and regression. Numerous applications have been suggested, such as hand written digit recognition, object recognition, text categorization etc. In many problems SVMs have been found to perform extremely well compared to other classifiers.

Not surprisingly, SVMs have been tried on credit scoring problems with encouraging results reported. Baesens (2003) compared the classification performance of Least Squares SVM against other commonly used techniques including decision trees, logistic regression, naive Bayes, linear and quadratic discriminant analysis and k-nearest neighbours. He concluded that SVMs achieved very good test set classification performances in terms of accuracy rate. However, his analysis was carried out on publicly available UCI benchmark datasets (Statlog Australian Credit, Bupa Liver Disorders, The statlog German Credit, The Statlog heart disease, The Johns Hopkins University Ionosphere, The Pima Indians Diabetes, the Sona, The Tic-Tac-Toe Endgame, the Wisconsin Breast Cancer and The Adult), none of which resembles a typical low default-rate dataset a major UK bank now faces. Therefore some questions still remain unanswered on how SVMs will perform on an extremely imbalanced dataset, which presents a difficult learning task.

Schebesch and Stecking (2005a) also applied SVMs to classification using loan data from a bank. However they conducted their analysis and reported findings based on re-sampled data so that the good-bad ratio was equalized rather than the original 6.7 percent. They have reported the performance of the classifiers using leave-one-out-error rate, which is around 25%, which is slightly better(but not significant) than logistic regression. In the results of Schebesch and Stecking (2005b), they have observed that SVM outperformed logistic regression with unequal sizes of good and bad in the sample as well as asymmetric costs of misclassification.

3.4.1 Formulation[2]

In a typical classification scenario, we have a set of training data to split into two classes. The data are in the form of pairs (x_i, y_i) where i is case i. In a normal binary classification problem, y_i is a class label of the data vector x_i taking value of either -1 or $+1$. For a linear machine on separable data all the training data should satisfy the constraints below:

$$\begin{cases} w \cdot x_i + b & \geq +1 \quad if \quad y_i = +1 \\ w \cdot x_i + b & \leq -1 \quad if \quad y_i = -1 \end{cases}$$

[2]The introduction of SVM formulation follows Burges (1998).

Where w and b are the weights and constant accordingly. The two inequalities can be combined to be written as

$$y_i(w \cdot x_i + b) - 1 \geq 0 \quad \forall i \tag{3.1}$$

The best classifier is selected when the pair of parallel hyper-planes $H_1 : w \cdot x_i + b = -1$ and $H_2 : w \cdot x_i + b = +1$ is found to have the largest orthogonal distance, or the maximum margin ($\frac{2}{||w||^2}$), as illustrated in Figure 3.1 where none of the data points fall in the region between the hyper-planes. Mathematically, finding the maximum margin is equivalent to minimising $||w||^2$ subject to the constraint in equation 3.1.

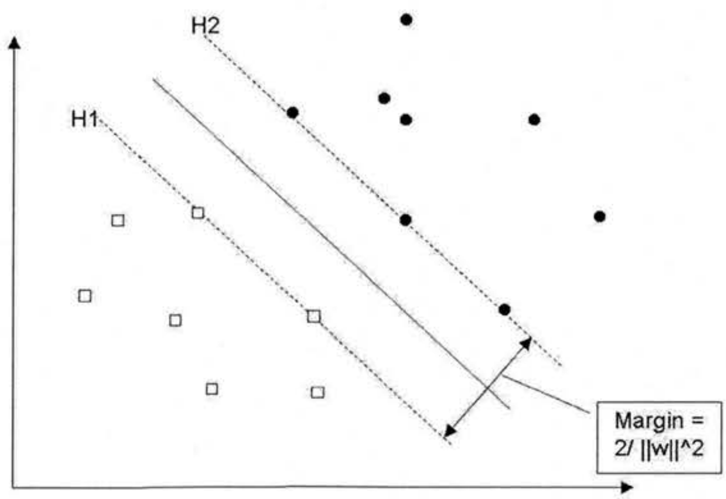

Figure 3.1: SVM on two-class linearly separable data

In most cases the data are not linearly separable. We can relax the inequalities by introducing positive slack variables ξ_i (Vapnik (1995)) in the inequalities to allow for training errors .

$$\begin{cases} w \cdot x_i + b & \geq +1 - \xi_i \quad if \;\; y_i = +1 \\ w \cdot x_i + b & \leq -1 + \xi_i \quad if \;\; y_i = -1 \\ \xi_i \geq 0 & \quad \forall i \end{cases}$$

The objective function to minimize is now

$$\frac{||w||^2}{2} + C(\sum_i \xi_i)$$

where C is a cost parameter controlling how much we can tolerate training errors. A larger C means a higher penalty assigned to the objective function during optimization. The Lagrangian L in primal form is:

$$L = \frac{||w||^2}{2} + C(\sum_i \xi_i) - \sum_i a_i(y_i(w \cdot x_i + b) - 1 + \xi_i) - \sum_i \lambda_i \xi_i \tag{3.2}$$

where a_i and ξ_i are the Lagrange multipliers introduced. Using Karush-Kuhn-Tucker conditions,

$$\frac{\partial L}{\partial w} = w - \sum_i a_i y_i x_i = 0$$

$$\frac{\partial L}{\partial b} = -\sum_i a_i y_i = 0$$

$$\frac{\partial L}{\partial \xi_i} = C - a_i - \lambda_i = 0$$

$$a_i \geq 0$$

$$\lambda_i \geq 0$$

$$y_i(w \cdot x_i + b) - 1 + \xi_i \geq 0$$

$$\xi_i \geq 0$$

$$a_i(y_i(w \cdot x_i + b) - 1 + \xi_i) = 0$$

$$\lambda_i \xi_i = 0$$

Note that if we choose i so that $\lambda_i = C - a_i > 0$, then we can have $\xi_i = 0$ because $\lambda_i \xi_i = 0$. Also if $a_i \neq 0$, $y_i(w \cdot x_i + b) - 1 + \xi_i = 0$ because $a_i(y_i(w \cdot x_i + b) - 1 + \xi_i) = 0$. Those points (x_i, y_i) that are called support vectors form the decision boundaries that are the two hyper-planes that separate the two classes. Substituting the equations above back into Equation 3.2, the problem of minimizing L becomes the problem of maximizing L_{Dual} in dual form:

$$\max L_{Dual} = \sum_{i=1}^{S} a_i - \frac{1}{2} \sum_{i=1}^{S} \sum_{i=j}^{S} a_i a_j y_i y_j (x_i \cdot x_j) \tag{3.3}$$

subject to

$$\begin{cases} 0 \leq a_i \leq C \\ \sum_{i=1}^{S} a_i y_i = 0 \end{cases}$$

where S is the number of support vectors.

When the decision function is better described using a non-linear function, better performance can be achieved using a non-linear support vector machine, which projects input data to higher or even infinite dimensional feature space where a linear classifier can separate mapped data much more easily. Such a method is called the 'kernel trick'. Note the mapping ϕ from lower l dimensional space to higher h dimensional space

$$\phi : R^l \mapsto R^h$$

Therefore in higher feature space, similar to maximising the margin, we need to solve an equation similar to equation 3.3

$$\max L_{Dual} = \sum_{i=1}^{S} a_i - \frac{1}{2} \sum_{i=1}^{S} \sum_{i=j}^{S} a_i a_j y_i y_j (\phi(x_i) \cdot \phi(x_j)) \tag{3.4}$$

subject to similar constraints. Note that components of $\phi(x_i) \cdot \phi(x_j)$ always appear together, and is replaced with the kernel function $K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j)$ when Mercer's Condition is met. The cunning bit here is that $\phi(x)$ does not need to be computed explicitly because only the values of $K(x,y) = \phi(x_i) \cdot \phi(x_j)$ are needed to be evaluated. By choosing an appropriate functional form of $K(x,y)$, computational complexity does not increase exponentially with the mapping from lower dimensional space to higher dimensional space.

A classification result for a test point T can be computed based on the sign of function

$$f(x) = \sum_{i=1}^{S} a_i y_i K(T, s_i) + b$$

where s_i is the ith point in the set of support vectors and S is the number of support vectors. Only those support vectors forming the decision hyper-planes are needed to give predictions. After the training of the SVM is finished, only these support vectors instead of the whole set of data points need to be retained in the model. That saves a lot of runtime memory and increases the speed of generating predictions.

3.4.2 Practical concerns

3.4.2.1 Kernel Choices

One of the most important factors that affects the performance of SVMs is the choice of kernels. Three very commonly used kernels will be used.

- linear kernel: $K(x,y) = x \cdot y$
- polynomial kernel : $K(x,y) = (\gamma \cdot x \cdot y + r)^d$
- RBF kernel: $K(x,y) = e^{-\|x-y\|^2/2\sigma^2}$
- Sigmoid kernel: $K(x,y) = tanh(\gamma \cdot x \cdot y + r)$

The kernel parameters like d r and σ have to be determined using some model chosen methods. Cross-validation is a common choice to select the best model in predicting unseen holdout data.

3.4.2.2 Different Weights on Different Classes to Handle Imbalanced Data

The original formulation in the SVMs aims to maximize the margin ($\frac{2}{||w||^2}$) subject to constraints. This implicitly assumes that the importance of classes are equal when trying to separate them. However, this maximizing objective is questionable when dealing with imbalanced data in which the minority class happens to be more important. SVMs trained with equalized weight on both classes will be rewarded (quite rightly according to the formulation) to allocate all cases into the majority class.

To handle an imbalanced data set, instead of minimising the equation treating good and bad cases at the same cost C which gives

$$\frac{||w||^2}{2} + C(\sum_i \xi_i)$$

Osuna et al. (1997) suggested the extension of introducing separate cost parameters C_+ and C_- to form a new objective equation to minimize.

$$\min \frac{||w||^2}{2} + C_+(\sum_{i:y_i=+1} \xi_i) + C_-(\sum_{i:y_i=-1} \xi_i)$$

subject to

$$\begin{cases} y_i(w \cdot x_i + b) & \geq +1 - \xi_i \\ \xi_i \geq 0 & \forall i \end{cases}$$

Here C_+ stands for the cost of misclassifying the positive class (minority and more important) and C_- is for misclassifying the majority negative class. Similar to previous treatment, Lagrangian multipliers can be introduced in the L as below,

$$\min L = \frac{||w||^2}{2} + C_+(\sum_{i:y_i=+1} \xi_i) + C_-(\sum_{i:y_i=-1} \xi_i) - \sum_i a_i(y_i(w \cdot x_i + b) - 1 + \xi_i) - \sum_i \lambda_i \xi_i$$

using KKT conditions, dual form L_{Dual} can be written as:

$$L_{Dual} = \sum_{i=1}^{S} a_i - \frac{1}{2} \sum_{i=1}^{S} \sum_{i=j}^{S} a_i a_j y_i y_j (x_i \cdot x_j) \tag{3.5}$$

subject to

$$\begin{cases} 0 \le a_i \le C_+ & if y_i = +1 \\ 0 \le a_i \le C_- & if y_i = -1 \\ \sum_{i=1}^{S} a_i y_i = 0 \end{cases}$$

After solving this optimization problem, the prediction can be made in a similar fashion as before, $sgn\{\sum_{i=1}^{S} a_i y_i K(T, s_i) + b\}$. Our estimations have been carried out on Libsvm Chang and Lin (2001) with the setting of $\frac{C_+}{C_-} = \frac{C*W_+}{C*W_-} = 25$. During the optimization, two model parameters are being searched, a common cost C and weighted odds $W = \frac{W_+}{W_-}$.

3.5 Performance measures

This section will define Receiver Operating Characteristics (ROC) curves and the area under the ROC curve and compares the advantages and disadvantages of alternative measures of predictive performance.

3.5.1 Definition of the ROC and area under ROC curve

The ROC curve was introduced to measure the ability to detect signals from noises. It is widely used in medical research to describe the detection ability of classifiers. For a dichotomous outcome problem, each instance case belongs to a positive or negative class label. The objective of the classifier is to label the cases into positive or negative classes. Given the classifier and existing known class labels, there are four possible outcomes of the prediction.

1. True Positive: Positive instances correctly labelled as positive

2. False Positive: Negative instances incorrectly labelled as positive

3. False Negative: Positive instances incorrectly labelled as negative

4. True Negative: Negative instances correctly labelled as negative

A 2x2 confusion matrix can describe this

		Actually Observed Classes	
		Positive	Negative
Predicted Classes	Positive	True Positives	False Positives
	Negative	False Negative	True Negatives

Table 3.2: Confusion Matrix

The True Positive Rate, also called sensitivity or recall, can be expressed as

$$True\ Positive\ Rate = \frac{TruePositives}{TotalActualPositives}$$

And the False Positive Rate,

$$False\ Positive\ Rate = \frac{FalsePositives}{TotalActualNegatives}$$

Specificity is defined as:

$$False\ Positive\ Rate = 1 - Specificity$$

A discrete two-class classifier outputting some probabilistic results predicts class labels given a cut-off or threshold value. The instances where the attached probabilities are higher than the cut-off value are then classified as positive and the rest as negative. For each cut-off value we can calculate the corresponding True Positive Rate and False Positive Rate. Then in a two dimensional True Positive vs. False Positive space, connecting all those (TP Rate, FP Rate) points will give us a ROC curve as in Figure 3.2, in which sensitivity equals to the true positive rate while false positive rate equals to 1 - specificity.

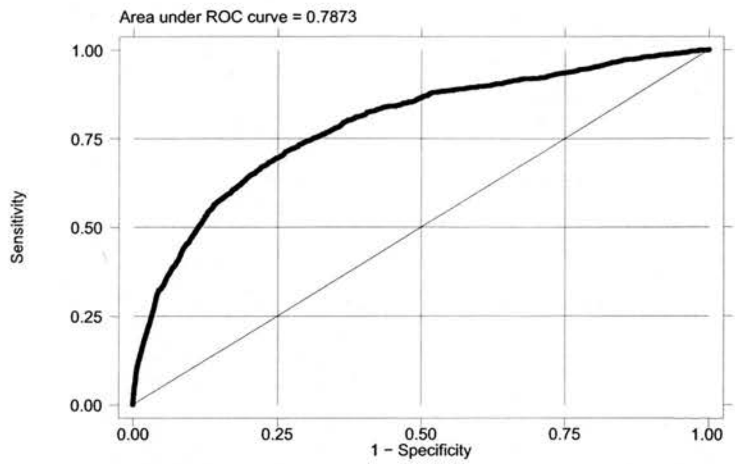

Figure 3.2: Sample ROC curve

One important value to read from a ROC curve is the area under the ROC curve (AuROC or AUC) value. Because both the X axis (FP rate or 1-specificity) and Y axis (TP rate or sensitivity) range from 0 to 1, the Area under a ROC curve is between 0 and 1. Its value is widely used to measure the ability of the model to correctly discriminate 'good'(positive) cases from 'bad'(negative) cases . The bigger and nearer to 1 the AuROC is, the better the model is considered to be in its classification performance. The AuROC is also related to the Gini Coefficient, which can be calculated as $2 * AuROC - 1$.

A random classifier doing nothing but wild guessing should get a straight ROC curve directly connecting points (0,0) and (1,1). Therefore its AuROC is 0.5. Any correctly discriminating classifier should do better than that. If not, reversing all the predictions from negatives to positives and positives to negatives should increase the AuROC.

Calculation of the AuROC can be done by counting the true positives and false positives under each cut-off setting and then summarizing them. A much more efficient way of calculating the AuROC value is to firstly sort the prediction results together with the actual class labels by the predicted probabilities(or scores). Recognizing that previously counted true positives and false positives can be re-used to calculate true positives and false positives under lower cut-offs. The computational complexity is therefore reduced to just one linear scan.

3.5.2 Area under ROC curve compared with Accuracy Ratio as a performance measure

The Accuracy Rate is widely used to measure the predictive performance of classifiers. The Accuracy rate is calculated as

$$Accuracy\ rate = \frac{Correctly\ Predicted\ Instances}{Total\ Instances}$$

The problem with Accuracy Rate is that per se it cannot correctly reflect the differences in the prediction ability on difference classes.

Many binary classification problems involve very skewed class distributions where as low as 1 in 1000 cases are positives that need to be detected. Contrary to the relative size in the distribution, those minority class members are more likely to receive special attention in the investigation. This is normally because when those minority class members are misclassified the costs incurred will be much higher than those from the majority class. A classifier making all of its mistakes on one important class can achieve the same Accuracy Rate as another classifier making mistakes on both classes or on the class that is less important.

If the cost of mis-classification can be determined by the researchers before modelling, a 2 by 2 cost matrix can be created (assuming the problem is a binary classification) as shown in Table 3.3. Each cell of the cost matrix represents the cost of mis-classification for that corresponding cell in the confusion matrix. For example, based on the cost matrix, a cost sensitive classifier can be calculated either re-weighting the training data according to the cost matrix or predict the class by minimizing expected misclassification cost instead of Accuracy Rate.

1	20
2	1

Table 3.3: Example of a Cost Matrix

The AuROC does not account for different costs of misclassification. The benefit of AuROC is that it gives an indication of predictive performance over all cut-off values. Its weakness is that for practical purposes we may be interested in only a narrow range of cut-offs, to be precise, the slope of the tangent line to the ROC curve at the cut-off point. Such optimal cut-offs can be found using ROC curve when loss and gain have been quantified, as shown in the chapter 7.6 of Thomas et al. (2002). Blochlinger and Leippold (2006) continued this discussion of optimal pricing strategies using the ROC curves and related those to the profitabilities. They simulated the competition between (an assumed) 3 lenders' loan market where the lenders took different rating models (with predictive abilities quantified by the AuROC) and various pricing strategies. The results are not surprising, the better the scoring model the more economic benefit.

3.6 Data preparation

3.6.1 Definition of default

Given the performance data we have access to, and depending on the type and length of time period we are researching ('the whole life of an account, the first 12 months or the most recent 12 months), we can classify the status of an account into one of 3 categories

1. Good. The observation never missed more than one or two payments during the time period we set.

2. Paying back early. The customer paid back early and settled the loan before the end of the observation period of the data .

3. Default. The customer missed 2 or more than 2 payments during the observation time period.

When modelling the default probability using binary logistic regression, the dependent variable takes only 2 possible values, 0 for no default or 1 for default. Therefore the paying back early cases are assumed to be good and marked as 0. The default is marked as 1 on those who have missed 2 or more than 2 payments in the first 12 months of their account histories.

3.6.2 Bands separation

The data we are working on were collected by a financial institution selling fixed term loan products. Based on the credit scores, a customer is allocated to one of 7 different bands numbered from 0, 10 20, ... 60. The number of cases in each band varies. The

Interest Rates	Mean	Std. Err.	[95% Conf.	Interval]	N
BAND0	17.60092	0.430556	16.75702	18.44481	251
BAND10	21.35344	0.086132	21.18462	21.52226	2240
BAND20	17.54549	0.055612	17.43649	17.65449	2915
BAND30	12.94531	0.033621	12.87942	13.01121	5078
BAND40	10.06131	0.031029	10.00049	10.12213	4604
BAND50	7.396764	0.021013	7.355579	7.437948	6915
BAND60	7.215612	0.010638	7.194761	7.236463	31347
Total	9.236919	0.0190147	9.19965	9.274188	53350

Table 3.4: Average interest rates across bands

smallest band 0, has only 253 cases while the most frequently populated band 60, contains 21840 cases.

Customers in each band were offered different interest rates to test their acceptance propensities. Across the bands, the average interest rate applied to each band reflects the level of risk the bank attached to the cases in that band. The lower the band, the higher the average interest rate will be charged. Within each band, the exact interest rate each applicant was charged was the average band rate plus or minus a random adjustment.

Table 3.4 reports the mean interest rate with standard deviations across different bands. Generally, the higher the band number, the lower the average interest rates that was charged with a smaller standard deviation observed. BAND0 is a very small band within which the institution has offered a relatively wide range of interest rates. This is evidenced by a much higher standard error of 0.43 within the band compared to those of other bands (0.01 to 0.08).

This analysis estimated a logistic regression model using all the data in aggregate and logistic regressions for each band separately.

3.6.3 Data transformation

Continuous variables were categorized into 7 equal sized bins by choosing 6 splitting points so that the number of cases falling between every two splitting points are the same (or as similar as possible). By doing so each bin has a similar and large enough number of cases. The traditional method of "coarse classifying", grouping cases based on the similarity of odds, is not adopted. This is because we have to predict at least two binary dependent variables, default and acceptance, both of which can provide conflicting odds (please see next chapter for acceptance modelling).

For categorical variables, some very rare levels cannot be divided evenly into separate training and holdout sample sets. If by chance all such cases fall into the holdout sample set, the model trained on the training set will have difficulties in predicting cases with such 'novel' values. As a remedy, their values are assigned to the nearest levels. These rare cases should have little impact on the predictive ability of the trained model.

Two methods are available for the coding of categorical variables. The first one is the so called Weights of Evidence. These can be calculated based on the log odds information so that categorical variables with a lot of levels can be easily transformed into just 1 dimensional numerical values, saving a lot of extra dimensionalities that would otherwise be required. This method is efficient and much faster for parameter estimation when we have only one targeting dependent variable to model.

However, since we have to model the probability of acceptance and default and also time to default in the next stage we chose dummy variables encoding in the later analysis (the only exception is the interest rate, retaining its original numerical values ranging from 5% to 32%). For a categorical variable with k multiple levels, $k-1$ dummy variables are created to replace the original variable. Each dummy variable takes a binary value of 1 or 0, corresponding to the presence of each level in the original categorical variable. The level left not coded is represented when all other dummy variables take the value 0.

3.6.4 Training and holdout sample separation

A model trained and tested with the same set of data cannot be used convincingly. This is because the model may be over fitted with the training data and so performs extremely well in classifying every case in the training data correctly but performs much less well in an independent holdout data set, which is representative of the population of all applicants (Thomas et al. (2002)).

By comparing the difference in classification performance between training set and randomly selected holdout set we can examine whether the current model is overfitting. Our data are separated into a randomly selected 70% training set and 30% holdout set.

3.7 Results

3.7.1 Logistic regression results

Logistic regression models were parameterised for each individual band to predict the default probabilities within that band. The BAND0 data are so small that the SAS logistic procedure reports finding quasicomplete data separation. The MLE estimates reported for BAND0 are therefore questionable. To get convergence in the maximum likelihood estimation, the sample size must be increased. We combined BAND0 and BAND 20 to form a larger set and ran logistic regression on it as well. The reason for choosing BAND0 and BAND20 is because the two bands have similar average interest rates, as shown in Table 3.4.

Table 3.5 reports the performance of the classifier across bands using the AuROC. From BAND0 to BAND60, the size of holdout samples is increasing. Band0 is so small that the likelihood estimation routine cannot converge. The AuROC of the model based on all bands data put together is much higher than those reported from other individual bands. This can be explained by the fact that the average interest rates offered to different bands are varied and combining bands together increases the variance of interest rates, one of the most predictive independent variables. In some bands, some holdout samples are omitted because they have some dummy variables not appearing in the training set due to the random training-holdout-split procedure.

On the training data with all bands combined together, stepwise routines were used to select the variables for the logistic regression. For the meaning of the names of the

BAND	Converged?	AuROC	Holdout Sample Size
0	N	0.8571	26
10	Y	0.7623	85
20	Y	0.6803	163
30	Y	0.6934	445
40	Y	0.6129	591
50	Y	0.6514	1430
60	Y	0.7063	6649
0&20	Y	0.6199	193
ALL	Y	0.7907	9397

Table 3.5: The predictive performance measured in AuROC across bands.

variables selected, please see Table 3.6. Appendix A: Tables A.1 and A.2 list the maximum likelihood estimates of all parameters. The coefficient of the interest rate APR is 12.2737, which means a percent unit increase in APR will lead the odds of default increasing by a factor of $exp(12.2737/100) = 1.1306$ when other variables are fixed at the same value. For example, the coefficient of 'Insurance=N' is -0.5967, which means if the customer does not take the insurance, the odds that he/she is to default will decrease by a factor of $1 - exp(-0.5967) = 0.4493$.

The MLE parameter estimates for each of individual band from 10 to 60 can be found in Appendix A: Tables A.3 A.4 A.5 A.6 A.7 A.8 A.10. Bands 0, 10 and 20 have few observations so the validity of the model estimated is questionable. In the models estimated for those bands, the Rate variable was even excluded after stepwise selection. One possible explanation could be that those bands are previously screened as high

Table 3.6: Variable name explanation

Variable Name	Explanation
Insurance	Insurance indicator
LOANAPR1	apr rate
LOAN_AMT	requested loan amount
TERM	requested loan term
internet	internet indicator
newbus	new bussiness indicator
ALCIFDET	any surname detected by CIFAS fraud prevention service
CCJGT500	Number of unsatisfied County Court Judgements which have a value greater than 500, recorded against the applicants name
LOANBAL1-4	Loan or card balance 1 to 4
NETINCM	net monthly income
SEARCHES	Total searches
SMO89	number of CAIS accounts in status category 8 or 9, mail order, same person
SNW12TV	worst status in the last 12 monthes, tv rental, same person
SOCNOACT	number of own company accounts
SOCSETT	number of settled CAIS account with status 0s, own company, same person
SPL6M4	Worst account status in the last 6 months of active CAIS accounts opened in the last 4 - 12 months matched to the applicant(s) as Same Person associated.
SPL6MACT	Worst account status in the last 6 months of all active CAIS accounts matched to the applicant(s) as Same Person.
SPSETLD	The number of settled Non Mail order CAIS with current status 0 matched to the applicant(s) as Same Person.
SPVALDEL	Total value of delinquent accounts - same person
SSRC4TO6	number of searches last 4-6 monthes, same person
SVALCAIS	The total of default balances of CAIS accounts with current status 8 or 9 matched to the applicant(s) as Same Person. This is grouped in bands of 10.
SWRSTCUR	Worst current status of all active CAIS accounts matched to the applicant(s) as Same Person.
TOSETTL6	if the credit commitment is being settled with the loan funds applied for.
WORST12	Worst status in last 3 months for any active CAIS accounts opened in the last 12 months matched to the applicant(s) as Same Person.
WRST46AL	Worst status in the period 4 to 6 months ago for any active CAIS accounts matched to the applicant(s) as Same Person.
AGE	age of the applicant(s)
INC_SURP	income surplus
MOR_RENT	Monthly mortgage or rent

risk bands and allocated higher interest rate so their default probabilities are relatively unaffected by the rate imposed on them.

3.7.2 Unweighted SVM results

This section reports the performance of SVMs on holdout data using different kernels and kernel parameters. The cost parameter C on both positive and negative classes is kept the same(un-weighted).

3.7.2.1 Data Sampling, Scaling and Variable Selection

Due to the limitation of computation resources required for massive model space searches for SVMs, 5000 cases were randomly drawn from the data and model searches were conducted on this subset of the data. This sample size is large enough to draw generalizable conclusions without being too slow to run on SVMs.

Some trial analysis showed that scaling numerical variables into values ranging from 0 to 1 improved the classifier's performance. The only variable not coded as dummy variable in the data, the interest rates, ranging from 5% to 32%. After scaling the values into ranges [0,1], the best AuROC value reached a maximum value of 0.7234, compared to the previous AuROC value of 0.70 from unscaled data. This can be explained by the fact that SVM is more sensitive to numerical values. Too large a value will unfortunately dominate the classifier and lead to sub optimal performance.

Although large dimensionality is not a problem for SVMs to handle, too many unpredictive independent variables may not be desirable and result in over fitting. Logistic regression procedures address this issue by variable selection mechanisms such as step-wise selection. For comparison of the performance based on the same set of data and variables, the same set of variables selected by step-wise selection in the logistic regression are used in the SVM classification.

Most of our variables are categorical and coded as dummy variables. That means for a three level variable, (0,0,1), (0,1,0) and (1,0,0) are used to represent each level within the variable. This is necessary as SVMs are good at handling numerical data problems. Many fields in which SVMs claimed the crown for the best predictor are those dealing with lots of numerically measured data, such as hand written characters or image recognition (see Chapelle et al. (1999)).

3.7.2.2 Linear kernel results

The Linear kernel $K(x,y) = x \cdot y$ performed rather poorly. The Fig 3.3 reports a ROC curve for a SVM using a Linear kernel with cost parameter $c = 1$. The Area under the ROC curve is only around 0.6.

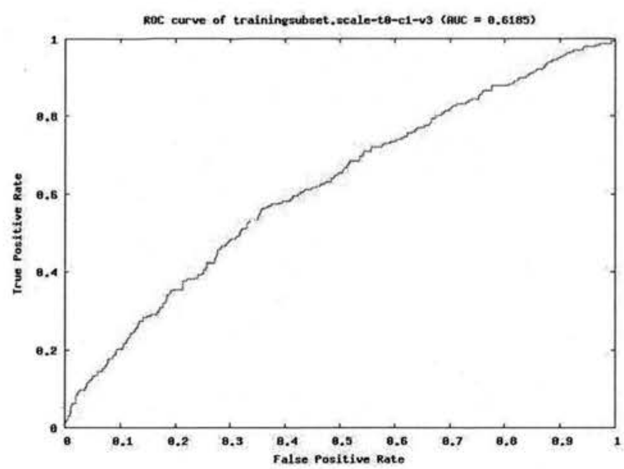

Figure 3.3: AuROC of SVM using Linear kernel

3.7.2.3 Polynomial kernel results

The results reported below are based on the Polynomial kernel $K(x,y) = (\gamma \cdot x \cdot y + r)^3$. ($r$ has been assumed to be a constant 0.) The Area under the ROC curve reaches 0.69, better than a Linear kernel but worse than a RBF kernel. The cost parameter ranges from 0.1 to 5. The gamma parameter ranges from 0.01 to 1. It seems the Polynomial kernel SVM built on this data is more sensitive to the choice of gamma parameter, where the best AuROC value 0.69 is achieved around $\gamma = 0.01$.

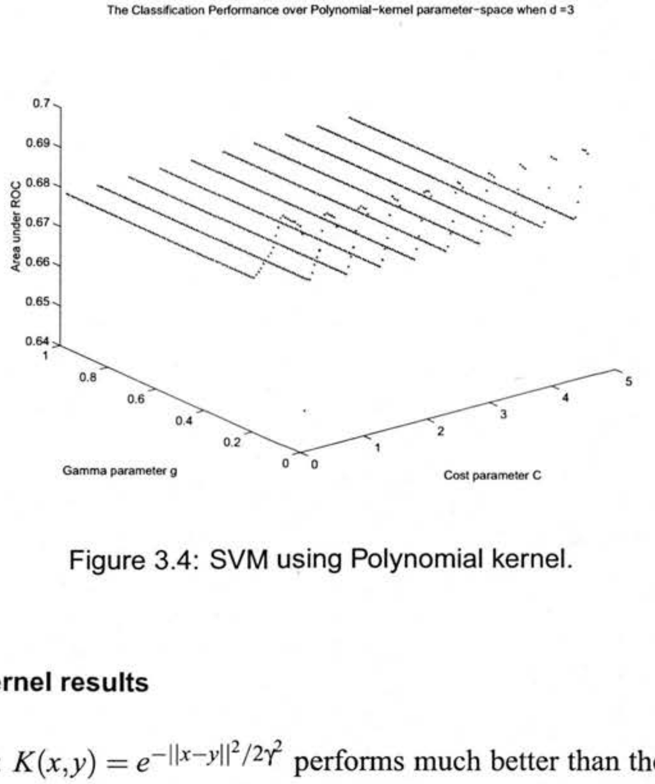

Figure 3.4: SVM using Polynomial kernel.

3.7.2.4 RBF kernel results

The RBF kernel: $K(x,y) = e^{-||x-y||^2/2\gamma^2}$ performs much better than the Linear kernel. A grid-fashioned search results in model parameter space of cost vs. gamma is shown in Figure 3.5. The cost parameter ranges from 0.1 to 5. The gamma parameter ranges from 0.01 to 1. It seems the RBF kernel SVM bult on this data is more sensitive to the choice of gamma parameter, where the best AuROC is achieved around $\gamma = 0.16$.

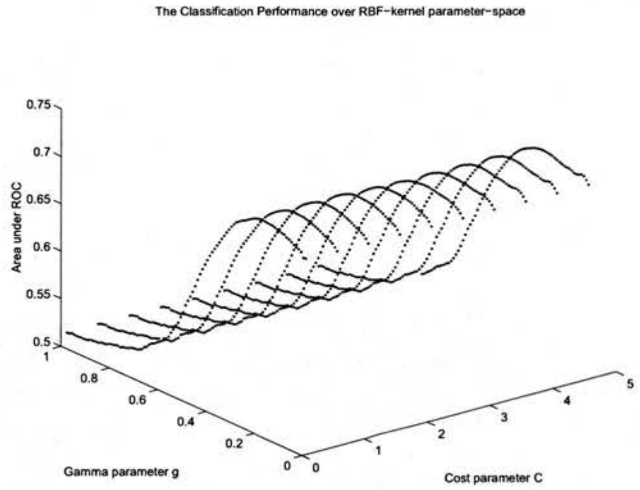

Figure 3.5: SVM using RBF kernel.

The best area under the ROC curve value of 0.7234 can be found when the cost is 0.05 and gamma is 0.16, as shown in Fig 3.6. Similar performance can be found with other cost parameters when the gamma parameter is around 0.16. This indicates the importance of the gamma parameter to the RBF kernel SVM classifier.

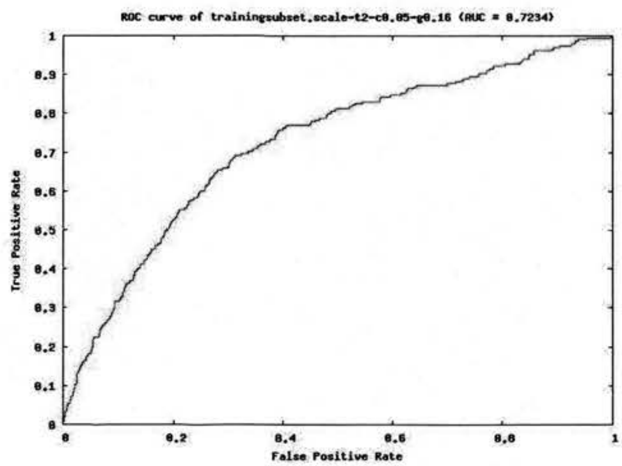

Figure 3.6: AuROC of SVM using RBF kernel

3.7.2.5 Sigmoid kernel results

The results for the Sigmoid kernel $K(x,y) = tanh(\gamma \cdot x \cdot y + r)$ (where r has been set as a constant 0) are shown below. The Sigmoid kernel achieved performance in terms of Area under the ROC curve (0.7172) similar to that of RBF kernel. The cost parameter ranges from 0.1 to 1.1. The gamma parameter ranges from 0.01 to 0.7. It seems the Sigmoid kernel SVM bult on this data is also sensitive to the choice of gamma parameter, where the best AuROC 0.7172 is achieved around $\gamma = 0.01$.

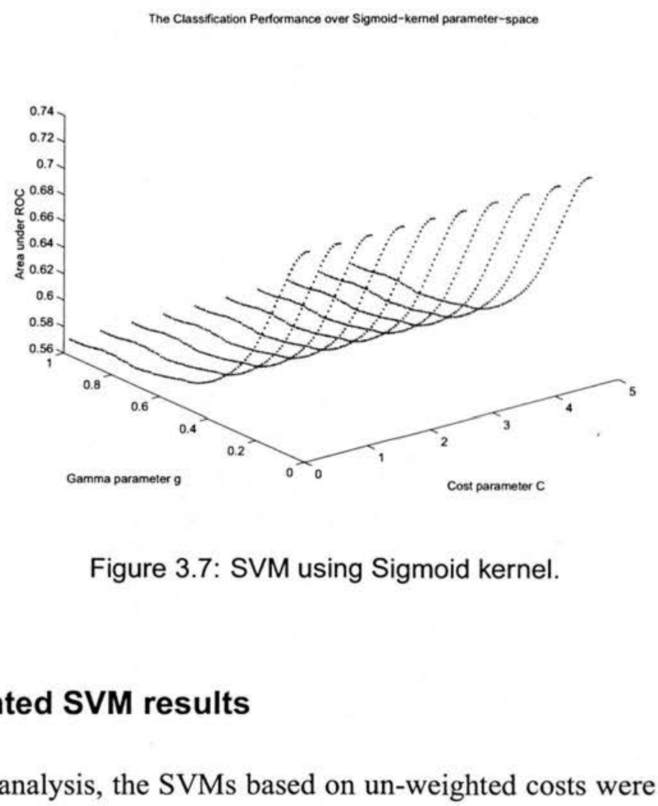

Figure 3.7: SVM using Sigmoid kernel.

3.7.3 Weighted SVM results

In our previous analysis, the SVMs based on un-weighted costs were not performing as competitively as reported in previous research papers. To see if the weighted SVM methods as described in section 3.4.2.2 can improve the predictive performance, we conducted the analysis as followed. We ranked cases by the probabilities predicted by LR and removed those between the 20th and 80th deciles, that is the middle 60% of the cases. Later analysis on the weighted SVMs are carried out on this set of data.

The weighted SVM was estimated on this subsample after a logistic regression. Unweighted SVM can achieve Area under ROC value of 0.7677 using an RBF kernel. The best AuROC value of 0.8037 can be found at when the cost parameter is 0.1 and the gamma parameter is 0.01 .

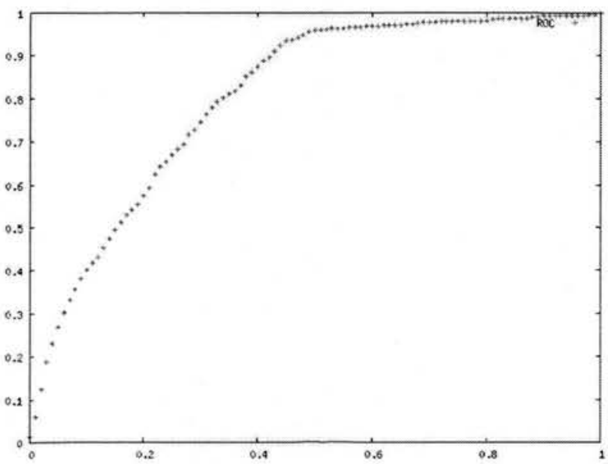

Figure 3.8: Area under the ROC of Weighted SVM using RBF kernel.

3.7.4 Summary

The LR model for the whole sample gives an AuROC of 0.7907 whereas the highest AuROC for unweighted SVM model, can only reach 0.7234, clearly being outperformed. The superiority of a weighted SVM using RBF kernel over an unweighted SVM using RBF kernel is also demonstrated by the improvement of AuROC from 0.7677 to 0.8037 on a set of subsample.

3.8 Simultaneous Equations Model

The financial institution providing the data carried out their analysis by first credit scoring the customers and separating them into different bands according to their risk of default. Each customer was then offered an interest rate based on the assigned risk band and some other random adjustments. Apparently the interest rate of the loan can be explained using the risk of default together with other exogenous demographic and bureau data variables. On the other hand, it is also possible that the probability of default is also affected by the interest rate of the loan simultaneously.

This section studies this simultaneous relationship hypothesis. Firstly we used a Bayesian Network classifier to search for the most predictive Bayesian networks hoping the structure of the network might reveal some conditional dependences. Next we fitted the data using a Simultaneous Equations Model and we examined the parameters estimated. We used a simultaneous equations model because we required unbiased estimates of the parameters of the default equation.

3.8.1 Investigation of the relationships between default, rate and score using Bayesian network structure search

Firstly, a logistic regression model was fitted where the dependent variable was the probability of default while the interest rate variable was excluded as a covariate. Secondly, using the predicted probabilities of Default as the score, together with the interest rate variable and the observed default indicator, a Bayesian network was searched and its parameters estimated.

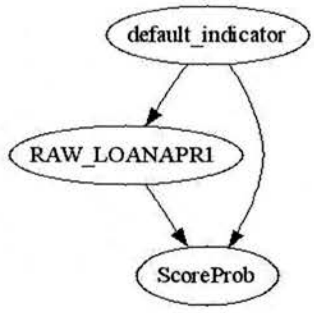

Figure 3.9: The relationships between 3 variables in a graph found after a global search

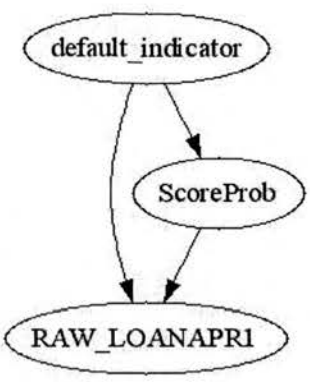

Figure 3.10: The relationships between 3 variables in a graph found after a local search

Figures 3.9 and 3.10 present the results of the search. The two different graphs are found using the same search mechanism TAN (Tree augmented network) but Figure 3.9 was searched globally while Figure 3.10 searched locally. TAN (Tree augmented network) was proposed in Friedman et al. (1997). A TAN is similar to a Naive Bayes network in which the class variable C is the root of the network and has no parent. Unlike a Naive Bayes network, the attributes of a TAN not only have the class variable as their parent, but also allow augmenting edges between the attributes.

A TAN is built by firstly constructing a complete undirected graph in which the vertices are the attributes and the weight of the edge is calculated from the conditional mutual information. Then use the procedures proposed by Chow and Liu (1968) to find a maximum weighted spanning tree. The resulting undirected tree can then be transformed to a directed one by choosing a root variable and setting the direction of all edges to be outward from it. The final step is to add the class variable C to the tree by adding an edge from C to each attribute.

Due to the nature of the Bayesian Network structure search algorithm, it always assumes the node of the default indicator is the root and parent of other explanatory variables (because it is the default probability to be finally predicted). Both structures achieved an Area under the ROC curve of 0.803. However, the two networks as shown in Figure 3.9 and Figure 3.10 differ in the way the dependence relationship between rate and score as presented. In Figure 3.9 there is a link pointing from LOANAPR to Score while in Figure 3.10 the direction of the dependency link is just the opposite. Both networks provide a good fit to the data and both conditional dependent relationships cannot be ruled out. For this reason we estimated a Simultaneous Equation Model based on the assumption that both Probability of Default (ScoreProb) and APR (RAW_LoanAPR1) are simultaneously affecting each other.

3.8.2 Simultaneous equations model

The CDSIMEQ package written in Stata by Keshk (2003) was applied to fit our data using a Simultaneous Equations Model. This package is well suited in the situation where one continuous (the rate) variable and one dichotomous (the default behaviour) variable are believed to simultaneously determine each other. The structure of the model is:

$$y_1 = \gamma_1 y_2^* + \beta_1 X_1 + \varepsilon_1 \tag{3.6}$$

$$y_2 = \gamma_2 y_1 + \beta_2 X_2 + \varepsilon_2 \tag{3.7}$$

y_1 is treated as an observed continuous endogenous variable and y_2 is a dichotomous endogenous variable and observed as

$$y_2 = 1 \quad if \quad y_2^* > 0$$
$$y_2 = 0 \quad if \quad y_2^* <= 0$$

where y_2^* is a latent continuous variable. X_1 and X_2 are matrices of exogenous variables in equation (3.6) and equation (3.7) accordingly. The exogenous variables selected into X_1 are chosen using forward stepwise linear regression and the exogenous variables selected into X_2 are chosen using forward stepwise probit regression. Both sets of variables contain at least one variable that resides in only one equation but not the other. Therefore the rank condition for identification is satisfied.

Because endogenous variables appear in the right hand side of the equation , the standard ordinary least square (OLS) estimates are inconsistent and biased. To address this problem, one of two methods are normally used. The first is to use indirect least squares (ILS) by solving the structural equations through reduced-form equations. The estimates of reduced-form equations by OLS are consistent and lead to consistent structural parameter estimates. The second method, the method we use, is using two-stage least squares procedures.

3.8.3 Two-stage least squares method

The two-stage Least Squares method works through two stages. The first stage is to create instrumental variables for the endogenous variables. The second stage replaces the endogenous variables in the structural equations with those instrumental variables.

In the first stage, the endogenous variables are regressed with all of the exogenous variables in the structural equations (3.6) and (3.7), noted as X, the matrix of all exogenous variables.

$$y_1 = \Pi_1 X + u_1^* \tag{3.8}$$

$$y_2^{**} = \Pi_2 X + u_2^* \tag{3.9}$$

where y_1 and y_2^{**} are instrumental variables. X are exogenous variables and not correlated with error terms u_1^* and u_2^*, both of which are assumed to be normally distributed. So consistent estimates can be obtained using OLS for equation (3.8) and probit for equation (3.9). Those estimated parameters are used to predict the instrumental variables as below

$$\hat{y_1} = \hat{\Pi}_1 X \tag{3.10}$$

$$\hat{y_2}^{**} = \hat{\Pi}_2 X \tag{3.11}$$

In the second stage, the original endogenous variables with their predicted values from equations (3.10) and (3.11) are substituted as:

$$y_1 = \gamma_1 \hat{y_2}^{**} + \beta_1 X_1 + \varepsilon_1 \tag{3.12}$$

$$y_2^{**} = \gamma_2 \hat{y_1} + \beta_2 X_2 + \varepsilon_2 \tag{3.13}$$

The final step is to correct standard errors generated in the second stage estimation, which are based on $\hat{y_2}^{**}$ and $\hat{y_1}$ not on the original variables. Maddala (1983), page

244-245, has given the corrected covariance matrix for the sets of estimates (γ_1, β_1) and (γ_2, β_2) and they were used here.

3.8.4 Estimates of the simultaneous equations model and prediction results

3.8.4.1 Parameter estimates

The final estimates are

$$APR = 0.1013 * Default + \beta_1 * X_1 \tag{3.14}$$

$$Default = 0.0272 * APR + \beta_2 * X_2 \tag{3.15}$$

where the β_1 and β_2 are the vector of coefficients of those exogenous variables X_1 and X_2. For detail please see Appendix A: Tables A.11 and A.12.

Table A.11 reports the parameter estimates for the first equation above. The instrumental variable for Default is denoted as I_Default. The coefficient of I_Default is a positive 0.1013, which indicates that higher probability of Default may contribute to higher interest rates. Its numerical value is small compared with some other coefficients attached to variables such as 'APR Adjustment'(1.0362). That is plausible as the APR adjustment variable is the random rate adjustment within the band. The APR adjustment plus band average rate will give the exact APR.

Table A.12 reports the parameter estimates for the second equation above. The instrumental variable is denoted as I_APR. The coefficient of I_APR is also a positive number 0.0272. That may be interpreted as when everything else being equal, charging interest with 1 unit higher rate will lead to 0.0272 standard deviation increase in the predicted probability of Default.

It is worth noting that the coefficient on APR is 0.0272 with p value of 0.151. The coefficient on Default also has a big p value of 0.132. Both are not significant. The numerical values of the equation coefficients may be interpreted as showing the propensity to default is affecting the interest rate much more than the way the default is influenced by the interest rate.

Another way to interpret this could be that other assumed exogenous variables are explaining most of the variance and therefore dwarfed the coefficients of the endogenous variable in the RHS of the equation. APR adjustment might be such a factor. After removing this variable from equations, the estimates turn out to be

$$APR = 2.0669 * Default + \beta_1 * X_1 \tag{3.16}$$

$$Default = 0.1097 * APR + \beta_2 * X_2 \tag{3.17}$$

This set of estimates can be interpreted as everything else held equal, 1 unit increase in APR will lead to nearly 0.11 increase in the predicted probability of Default. 1 percentage higher probability of Default means 2.07 unit increase in the APR if everything else is the same. The details can be found in Appendix A: Tables A.13 and table A.14. This set of estimates now has a coefficient value with p values near 0, although the Area under ROC is decreased to 0.7622.

3.8.4.2 Modifications of the CDSIMEQ package

The Stata package CDSIMEQ written by Keshk (2003) handled estimation and covariance correction well. However its post estimation routine has difficulties in dealing with out-of-sample estimation. This is because when we are to use the estimated Probit model to predict default for future samples, we need the values of the instrumental variable $\hat{y_1}$ in equation (3.13). CDSIMEQ overlooks this situation and only works when doing post-estimation on the data set on which the model was fitted.

To discover an unbiased estimate of the predictive ability of models we carried out 10 fold cross-validation. The original package of CDSIMEQ is not capable of that. For this reason, our own version of 2SLS estimation was written. In our own package, the data values that are used for the prediction of instrumental variable values are predicted based on the estimated parameters in the first stage, as below.

$$\hat{y_1} = \hat{\Pi}_1 X$$

The predicted instrumental variable $\hat{y_1}$ is then plugged into the formula below to generate the prediction for y_2

$$y_2 = \gamma_2 \hat{y_1} + \beta_2 X_2$$

3.8.4.3 Prediction results

The ROC curve based on the predictions on the training set is shown in Figure 3.11. The area under the ROC value is 0.7873. Using our modified 2SLS routine to run 10 fold cross validation, the area under the ROC achieved is slightly decreased to 0.7837 (when predicted using Instrumental variable values predicted in the first stage). The AuROC is increased a little bit to 0.7862 when using observed interest rate instead of using predicted instrumental variable values. As a comparsion, a 10 fold cross validated Probit achieves a similar Area under the ROC value of 0.7854 and a 10 fold cross validated Logistic regression model achieves an Area under the ROC value of 0.7819.

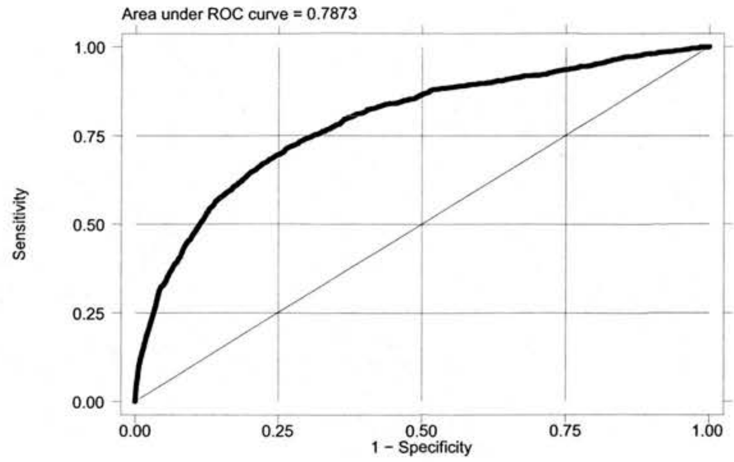

Figure 3.11: Area under ROC given by CDSIMEQ

3.9 Conclusion

This section presented predictive performance results using logistic regression, SVM and Simultaneous Equations Probit Model. Building models based on the data, the logistic regression predictor (achieving an AuROC of 0.79) is the most competitive compared with SVMs (using various kernels achieving the best AuROC of 0.72) and Simultaneous Equations Probit Models (achieving AuROC around 0.78).

It has to be noted that the difference between model performance results can be accounted for by the difference in my relative familiarity or expertise with those models. Different preprocessing methods used in different models might also lead to biased comparison results. Potentially either of this happening could endanger the validity of the model comparison conclusion, as observed by Hand (2006). Despite this, best efforts have been made to remove bias as much as possible during the model building.

The conclusion presented here is the best I can draw given my existing knowledge. Therefore it is not useless unless future investigations prove the other way.

Considering simplicity and computational resources requirement, logistic regression is no doubt the best choice for predicting defaults. However, when the inter-relationships between Rate and Default is of more interest to the lenders that are offering rate-varied products, the simultaneous equations can provide more insights of the dynamics between those factors.

The poor performance of SVMs can be accounted for with two reasons. First, the data is converted into 0-1 dummy variables, while SVMs normally excel when dealing with continuous variables. The second, extremely skewed class distribution makes it difficult for the SVMs, which are originally formulated to minimize the classification rate.

Chapter 4

Acceptance Modelling

4.1 Introduction

This chapter is dedicated to the modelling of consumer acceptance behaviour. Section 4.2 describes previous research on modelling consumer acceptance behaviour. The next sections will describe the model design, the data and present the data preparation procedures. Section 4.5 reports the results of fitting logistic regression Models. Section 4.7 will report our investigations into the modelling of acceptance elasticity with respect to interest rates. In section 4.8 the possibility of improving the default probability estimation by applying bivariate Probit Sample Selection Models is examined. Section 4.9 plots the indifference curves following Keeney and Oliver (2005). In the final appendix section the tables of estimated MLE results given by the models are listed.

4.2 Previous research in acceptance modelling

Thomas et al. (2006) and Jung et al. (2003) have suggested that significant changes are happening in the evermore competitive consumer lending market . The first one is the need for tailoring varied-features financial products to improve the likelihood of a consumer accepting an offer of the product made to him. The second is the requirement for building interactive application processes in the newer communication and marketing channels like the Internet or telephone so that during the application process the lender can adjust their offers to make acceptance more likely.

Both changes necessitate the ability of the lender to infer the probability of a particular consumer accepting a specific offer during the interactive application process. Some recently published papers have presented how this issue may be addressed. Rossi et al. (1996) investigated various forms of purchase history data of Chicago households. They employed multinomial Probit models to predict the price sensitivities and household preferences in terms of 'target couponing'. They used their model to explain the heterogeneity across households using a hierarchical Bayesian model. The inference was conducted in a Bayesian way and posteriors were acquired using Gibbs Samplers through Markov chain simulation. By offering a customized coupon strategy to attract different customers, they estimated that a seller could have a potentially substantial gain in revenue than if a blanket coupon strategy in which all coupons have the same value is offered.

Montgomery (2001) discussed many applications of quantitative marketing techniques on the Internet when consumers are 'addressable' thanks to advances in information technology. In one of the examples given, a multinomial logit model was fitted to the data featuring factors that affect consumers' purchase choices. Those factors included item price, shipping price, tax, delivery charges as well as the brand names of sellers. The parameter estimates of the coefficients on those factors implied the feature importance. The price sensitivities were then quantified.

Jung et al. (2003) investigated and compared three different methods (logistic regression (LR), linear programming (LP) and an accelerated life (AL) model) to model the likelihood of consumers accepting student bank accounts when being given different offers. Those offers have six features, including 5 choices of overdraft limit, 4 choices of credit card options, fee for foreign currency, discounts on insurance, interest paid on account surplus and 10 choices of free gifts. Their data set, named the Fantasy Student Current Account (FSCA), was gathered from a dedicated website, which was widely publicized to first year students at the University of Southampton with prize winning draws as enticement.

Seow and Thomas (2005) not only investigated the effects of specific features on acceptance behaviour, but also tested the influence the number of questions could have on the consumer acceptance behaviour. They modelled the probabilities of an applicant taking different offers using decision trees and based their analysis on the same data set as used by Jung et al. (2003). A two layered decision tree structure is used whereby the enforced upper layer uses applicant characteristics only and the lower layer uses only offer characteristics. This structure offers the convenience for the lender to build an adaptive application process by asking customers about applicant characteristics first and afterwords providing the offer that is the mostly likely to be taken by this customer.

Different tree settings were tested and analysed in their paper. Those trees included an applicant characteristics only tree, offer characteristics only tree, trees with both types of characteristics and even with more flexibility allowed in the tree structure (so called alternate best tree) to generate a better fit to the data. They also explored the situation when imposing limits on the number of questions asked as a restriction on the tree building process. Pruning this tree can reduce the number of questions asked and hence potentially increase the probability of acceptance by the customers.

Because of the particular nature of the sample and the possibility of the 'testing effect' of data collected, the results obtained in Jung et al. (2003) and Seow and Thomas (2005) may not be generalizable. Thomas et al. (2006) mentioned that once the likelihood of acceptance is estimated for a customer, the lender can make the offer based on the optimality of profitability. However, they did not give comments as to how the optimality of profitability can be achieved. Besides, there is no research looking into how the consumer behaviour of accepting the offer may affect their risk of default or vice versa.

The contribution of this chapter is to model the probability of acceptance using data relating to the actual acceptance or rejection of the offers made to the applicants for a fixed term loan product. Since the data recorded the acceptance decisions of those applicants, the results are not subject to a "testing effect".

4.3 Model design

As shown in Figure 4.1, the whole potential customer population can be partitioned into those people who did not apply (*NA*) for the credit product, those who did apply but got rejected (*AR*), those who applied, received an offer but refused to take it (*CRO*), those who applied, received and took offers and being good customers (*G*) or bad customers (*B*). The whole population can be expressed as

$$Whole = G \bigcup B \bigcup CRO \bigcup AR \bigcup NA$$

while the intersections between sets *G*,*B*,*CRO*,*AR*,*NA* are all empty sets.

$$G \bigcap B = B \bigcap CRO = CRO \bigcap AR = AR \bigcap NA = NA \bigcap G = \emptyset$$

We now consider customers who applied, passed their credit check, and received an offer. That is the set $G \bigcup B \bigcup CRO$. We observe performance information for those who have applied, passed the credit check and then took the offer: $Accept = G \bigcup B$.

We assume a case makes a choice between defaulting and not defaulting, and between accepting a credit offer and rejecting it. In each case we assume the consumer makes the choice which maximises his utility. In each case we model the utility of default (acceptance) as an unobserved continuous variable D^* (A^*) such that

$$D^* = \quad (Default)^* \quad = \beta_1 X_1 + \varepsilon_1 \tag{4.1}$$

$$A^* = \quad (Accept)^* \quad = \beta_2 X_2 + \varepsilon_2 \tag{4.2}$$

We do not observe the utilities underlying the chosen action: default or none default, acceptance or rejection. But we do observe the binary situation of default ($P(Default = 1)$) or non-default ($P(Default) = 0$) , acceptance ($P(Accept) = 1$)or non-acceptance ($P(Accept) = 0$). The observational regime is therefore:

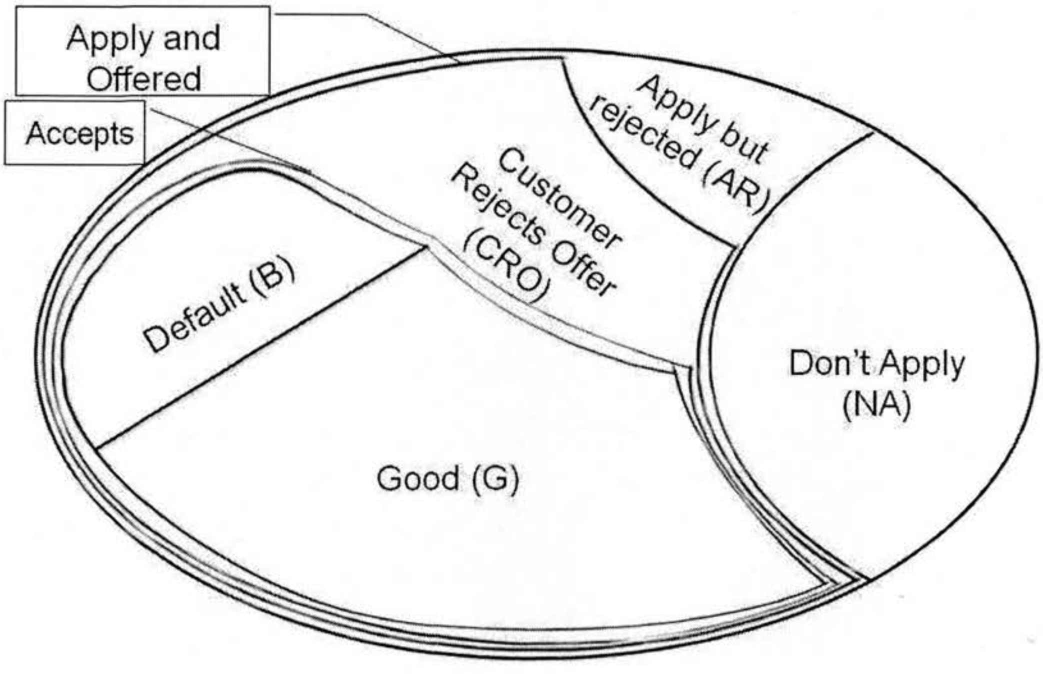

Figure 4.1: Our data samples sets

$$\begin{aligned} P(Default) &= 1 \; if \; (Default)^* > 0 \\ P(Default) &= 0 \; if \; (Default)^* <= 0 \\ P(Accept) &= 1 \; if \; (Accept)^* > 0 \\ P(Accept) &= 0 \; if \; (Accept)^* <= 0 \end{aligned}$$

Notice that we can observe $P(Default)$ only if $P(Acceptance) = 1$. [1]

Our research strategy is firstly to model the probability of acceptance directly using logistic regression, assuming any correlation between ε_1 and ε_1 in equation 4.1 and 4.2 is zero. Second, in section 4.8 we will drop this assumption and estimate the bivariate

[1] A further expansion of the model could be utilizing the data in set AR and using a doubled selection model to fit the whole data.

Probit sample selection models, allowing for the possibility that the errors in the two equations are correlated.

4.4 Data preparation

Before presenting the results, it is necessary to report how the acceptance behaviour is defined, why there is band separation, the transformation of the encoding of data variables and why the data have to split into training and holdout sets.

After the customer completed an application form, requested an amount of the loan, chose whether or not to request insurance with the loan, and passed credit check, he/she may be given an offer with a specific interest rate to accept or reject. This offer will consist of a loan of a given loan amount, usually the amount requested, sometimes adjusted by the lender (this does not happen often, however). The lender will allocate the customer into a certain band reflecting the risk of default. Most of the interest rates offered within a given band are the same but small variations exist within the band in some cases because of the lender's adjustments.

The customers who accepted the offer and took the loan are marked with 1 in a binary 'paid' indicator. Only the customer who accepted the loan will have performance data recorded and subsequently can be classified as 'good' or 'bad' customers depending on the definition of default. Each applicant who received an offer was given an interest rate from one of the seven bands as described in section 3.6.2.

We chose to use Dummy Variables [2]instead of Weights of Evidence [3] for consistency with the data used for modelling Default as well as Acceptance.

The dataset has over 53,000 cases of applicants applying for a fixed term loan product. The length of the term ranges from 24 months up to 84 months. To test the predictive performance of our models, we trained the model using the training set consisting of 70% randomly selected cases from all the samples and tested it on a holdout set consisting of the remaining 30% of the total sample.

4.5 Logistic regression results

We fitted the logistic regression model on all the data put together as well as data in each individual band. The performances of the models' predictive abilities are compared using area under ROC values on the holdout sample data set.

4.5.1 Performance across bands

From Table 4.1 , we can see the performance measured by area under ROC values is increasing with the size of holdout samples across different bands. The area under ROC of the model based on all bands put together is much higher than that of other individual bands. This may be accounted for by the doubled sample size. Or this can be explained because the average interest rates offered to different bands are varied, com-

[2]For a categorical variable with k multiple levels, $k-1$ dummy variables are created to replace the original variable. Each dummy variable takes a binary value 1 or 0, corresponding to the presence of each level in the original categorical variable. The level left not coded is represented when all other dummy variables take the value 0.

[3]Please refer to the explanations in previous chapter

bining bands together increase the variance of interest rates, one of the most predictive independent variables.

BAND	Converged?	AuROC	Holdout Sample Size
0	N	0.9167	67
10	Y	0.6956	683
20	Y	0.6730	894
30	Y	0.6797	1585
40	Y	0.6680	1447
50	Y	0.6854	2146
60	Y	0.7187	9362
0&20	Y	0.7168	965
ALL	Y	0.7832	16193

Table 4.1: Comparison of the predictive performance for different risk bands

The sample size for BAND 0 is so small that SAS logistic procedure reports finding quasi complete data separation. The MLE estimates reported for BAND 0 are therefore questionable. To get convergence in the maximum likelihood estimation, the sample size must be increased. We combined BAND 0 and BAND 20 to form a larger set (because the two sets have similar acceptance percentages) and reported results.

4.5.2 Features selected from stepwise selection

Compared with the 12 features selected using stepwise selection in modelling Default, more (37 in total) features were selected in modelling Acceptance. Similar to the features selected when modelling Default, the first 2 features selected are the Interest Rate variable (R1) and Insurance-take-up indicator (CPI). The features of new customer indicator (newbus), loan amount requested, as well as the length of the loan (TERM) entered the model at an early stage and stayed there. That is sensible as those variables are very likely to influence customers acceptance behaviour.

4.5.3 ROC curve and MLE results

Figure 4.2 shows the ROC curve on the holdout sample using the logistic regression model fitted on the data with all bands combined together. The area under ROC value is 0.7832. The table of estimated MLE results given by logistic regression can be found in Table B.1 in the appendix section. Please note that many dummy variables do not appear to be significant in the MLE results, even those variables were previously selected from a stepwise selection routine. This happens because of the way the SAS package conducts the stepwise selection on categorical variables by adding or removing each categorical variable as a whole. Therefore even some dummy variables that were created out of a categorical variable are not significant in the model, they still enter the final model because some other levels within the categorical variable are highly significant that they cannot be removed. (The stepwise routine from Stata package, on the other hand, can evaluate the dummy variables individually for each level of the category, and therefore yields slightly different models.)

4.6 SVMs results

In previous chapter we have modelled SVMs on default and found their predictive performances not as good as logistic regression. The acceptance data is different from the default data. A big difference is the class distribution. In the acceptance data over sixty

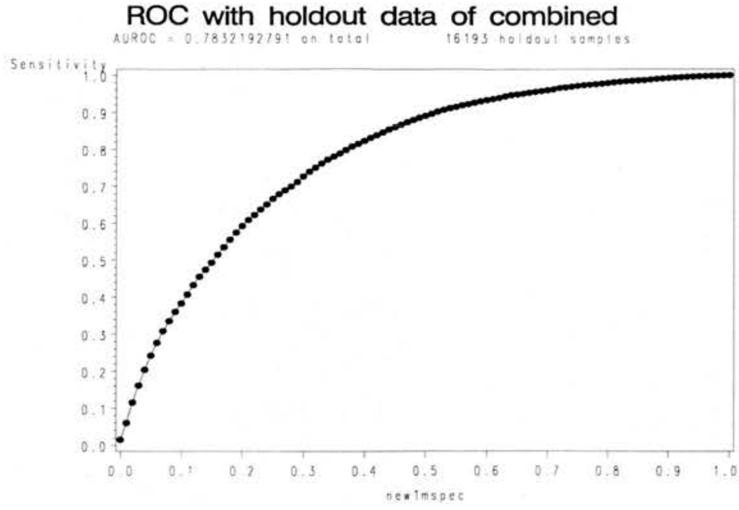

Figure 4.2: The ROC curve for the Acceptance model all bands combined

percent are the positive outcome (offer accepted) while in the default data only around four percent of cases are observed to default.

One of the most important factors that affects the performance of SVMs is the choice of kernels. Two very commonly used kernels listed below will be used and shown to be very competitive compared to the logistic regression: the polynomial kernel and the RBF kernel. Their kernel parameters have to be determined using some model selection methods. Cross-validation is a common choice to select the best model in predicting unseen holdout data. As our data is very large, the SVMs can be very slow to train. A two-fold cross validation has been used in the grid searches of the models that will yield the highest AuROC values.

To make the model comparison between logistic regression on the same ground without being affected by the choices of the feature selection routines, the SVM used for

Acceptance modelling will be using the same set of variables selected from the step-wise selection routine in the logistic regression.

4.6.1 RBF kernel

The RBF kernel uses the kernel function like $K(x,y) = e^{-\gamma * ||x-y||^2}$. A grid-fashioned search results in model parameter space of cost C vs. gamma γ is shown in Figure 4.3. The cost parameter ranges from 0.1 to 5. The gamma parameter ranges from 0.01 to 1. The third axis is the area under ROC achieved through a two fold cross validation on the holdout set. It seems the RBF kernel SVM built on this data is more sensitive to the choice of gamma parameter, where the best AuROC so far is 0.7905, achieved at around gamma $\gamma = 0.04$ and cost $C = 1.1$.

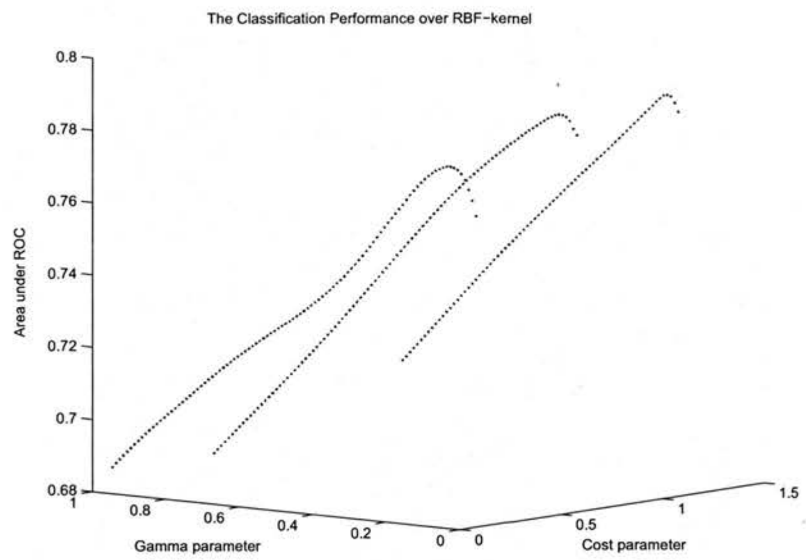

Figure 4.3: Grid search of the best predictive model parameters for SVM RBF kernel

4.6.2 Polynomial kernel

The polynomial kernel uses the kernel function $K(x,y) = (\gamma \cdot x \cdot y)^d$. When the dimension parameter d is assumed to be one, the SVM is generally a linear kernel SVM. A reasonable large enough dimension parameter $d = 3$ [4] was chosen and analysed with results shown in Figure 4.4. The best AuROC value is 0.7937, where cost $C = 0.3$ and gamma $\gamma = 0.0100$. Compared with the Figure 4.3, where a lot more parameter combinations have been searched, the range of the grid search for this polynomial kernel is much smaller. This is because the polynomial kernel with a higher dimension parameter is very slow to run on a larger data set. Restriction of the computation resources limited the range of the parameters search.

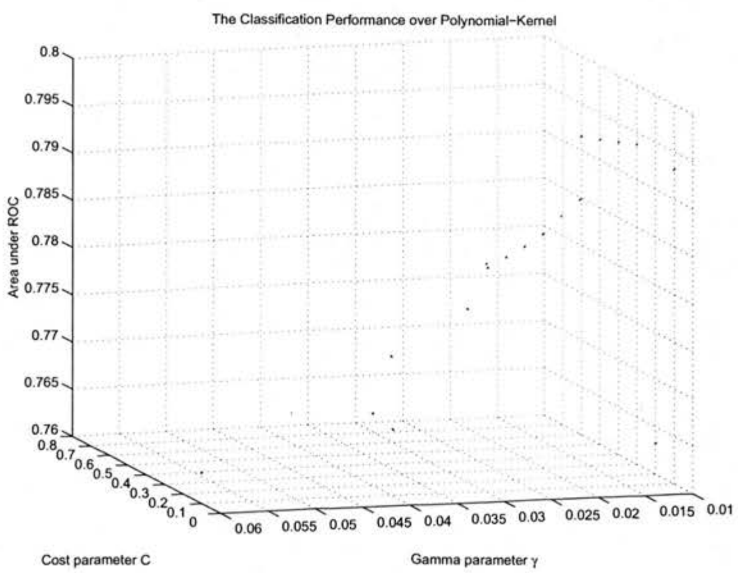

Figure 4.4: Grid search of the best predictive model parameters for SVM Polynomial kernel with dimension parameter $d = 3$

[4] Other dimension parameters can also be tested but the limitations of the computation resources prevented us from doing so.

4.7 Modelling acceptance elasticity of interest rate

Elasticity could be defined as the proportional change in one variable divided by the proportional change in another variable.

A general formula for the elasticity (the "x-elasticity of y") is:

$$E_{x,y} = \left|\frac{percent\ change\ in\ y}{percent\ change\ in\ x}\right| = \left|\frac{\partial lny}{\partial lnx}\right| = \left|\frac{\partial y}{\partial x} * \frac{x}{y}\right|$$

Previous logistic regression results [5] on the acceptance modelling have shown that the variable having the most influence over customers' decisions to take or reject the offer is the interest rate charged. To analyse the price elasticity of the propensity of customers to take loan product offers, we calculate $\frac{\partial P(A)}{\partial i}\frac{i}{P(A)}$

The functional form of acceptance probability is assumed to be logit as below. x is a vector for the independent variables and β is the vector of parameters.

$$logit(p(Accept)) = log(\frac{p(Accept)}{1-p(Accept)}) = w = \beta x$$

$$p(Accept) = \frac{e^w}{1+e^w}$$

The partial derivative on one of the independent variable x_j (with β_j as the corresponding parameter)is

$$\frac{\partial P(Accept)}{\partial x_j} = \frac{e^w}{(1+e^w)^2}\frac{\partial w}{\partial x_j} = \frac{e^w}{(1+e^w)^2}\beta_j$$

So the price(interest rate i, as x_j) elasticity of the Acceptance Probability is

$$El_{i,P(Accept)} = \frac{\partial P(Accept)}{\partial i} * \frac{i}{P(Accept)} = \frac{\beta_i * i}{1+e^w}$$

[5] Please see previous section 4.5.3 and the estimated coefficients in Table B.1

4.7.1 Elasticities for bands

As previously given, the elasticity can be written as a function of price when other independent variables are assumed to be constant values. Mean values of those variables are used as the constant values when calculating the elasticities. Following previous data transformation, all those variables except the interest rate charged were converted to dummy variables. For each dummy variable, the relative frequency of each dummy variable is used as the mean value.

As the interest rates offered range from 4.99% to 32.99%, the price elasticities of probability of acceptance are calculated by fitting the interest rate value into the previous elasticity equation. The table below lists the average elasticities within each band and all bands combined together. The charts of price elasticities of acceptance in each individual band can be found in Table 4.2 and in Figures 4.5, 4.6 , 4.7 , 4.8 , 4.9 , 4.10 and 4.11. The price elasticities of acceptance for others variables can be found in Table 4.3.

Notice from Table 4.2:

- The first column indicates on which data set the elasticities are calculated. The results for band 0 are questionable because of its very small sized sample leading to quasi complete separation in the data during maximum likelihood estimation. Band 0 and Band 20 are combined to get reliable estimates.
- The second column presents the average values of acceptance elasticities of interest rate.
- The third column and fourth column shows at which the point the elasticity is the biggest.

Data	average elasticity	interest rate with max abs(elasticity)	maximum abs(elasticity)
*band0	-1.054006	- 9 %	- 3.140571
band10	-0.546448	- 13%	- 0.714744
band20	-0.541663	- 12%	- 0.74064
band30	- 0.59789	- 13%	- 0.804827
band40	- 0.65196	- 14%	- 0.834453
band50	-1.000923	- 17%	- 1.231596
band60	-1.547289	- 29%	- 2.15465
band0 And 20	-0.683178	- 12%	1.005482
combined	-1.369484	- 14%	1.997809

Table 4.2: The price elasticities across different bands

	Elasticity at the Mean Values	Mean Values of Variable
APR	-1.34056	9.23288
Insurance	-6.57422	0.371393
Loan Amount	-3.63484	9612.67
Term	9.382553	52.249
Internet	-23.9071	0.407225
New Business	-6.6639	0.848887

Table 4.3: The elasticities of other variables on all bands combined

- We observed that average elasticity grows steadily from band 20 up to band 60, that makes sense by realizing the fact that higher band generally has been regarded as lower risk and been charged lower price. They shall find themselves having more financial alternatives and therefore more likely to be put off at higher rates.

- Interestingly, the points where maximum absolute elasticity is observed are also shifting from lower band to higher band, as shown in Figures 4.5 to 4.10. This may happen because of the logistic function form we have chosen. Recall that the elasticity function takes the form of

$$El = \frac{\beta_i * i}{1 + e^w}$$

Differentiate El with respect to interest rate i

$$\frac{\partial El}{\partial i} = \frac{\beta_i}{1 + e^w} - \beta_i * i \frac{e^w * \beta_i}{(1 + e^w)^2}$$

The maximum or minimum point resides where

$$\frac{\partial El}{\partial i} = 0$$

So solving the equation below we can get the point.

$$e^{\beta * x}(i * \beta_i - 1) = 1$$

Notice that i and β_i are always appearing together in equation and both $e^{\beta x}$ and $i * \beta_i - 1$ are monotonically changing with $i * \beta_i$, so a unique answer to the equation above shall exist. Assume we have found the $i * \beta_i$ satisfying the equation. A smaller β_i means the interest rate i has to be larger. Looking at the coefficient in our logistic regression estimates corresponding to the interest rate confirms that.

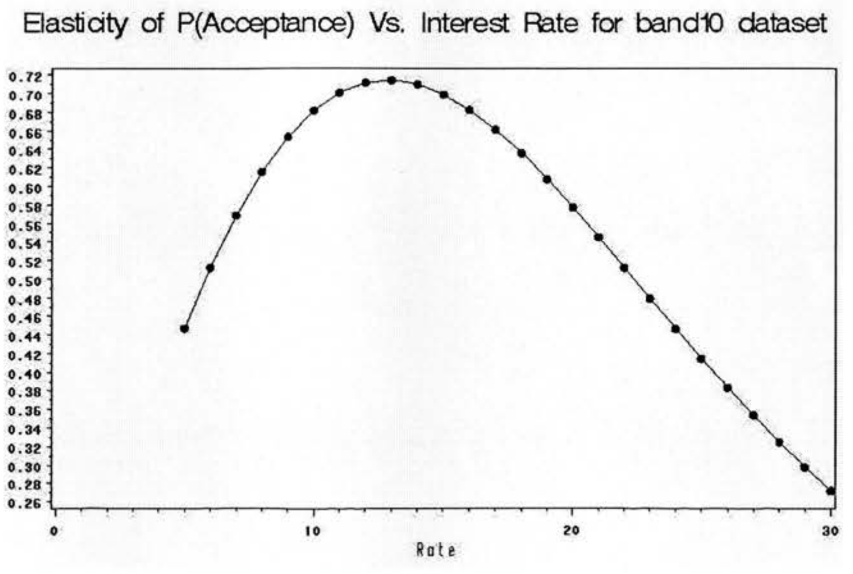

Figure 4.5: Band 10 has shown the most elasticity at interest rate of 13%

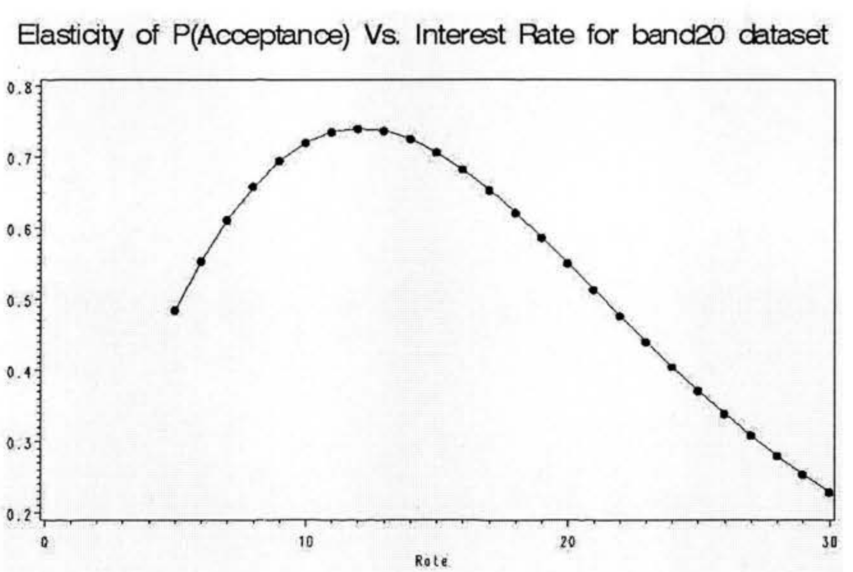

Figure 4.6: Band 20 has shown the most elasticity at interest rate of 12%

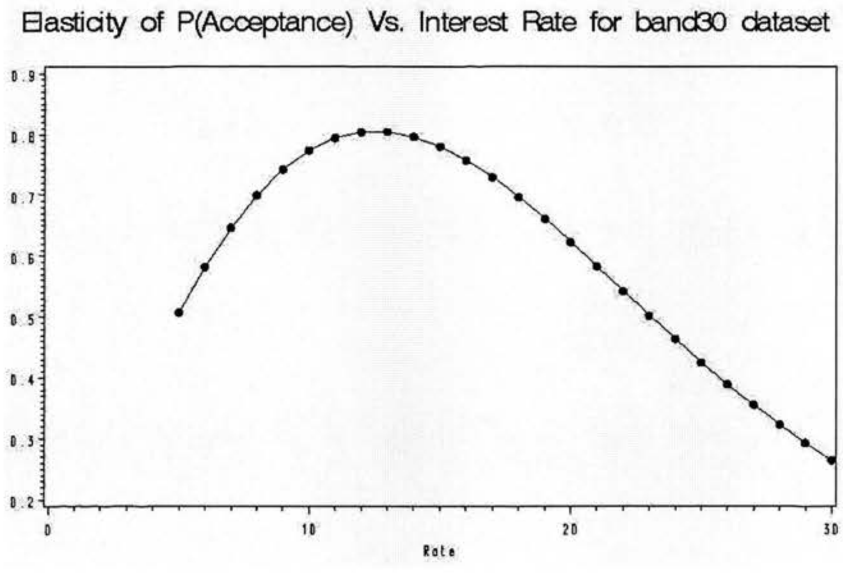

Figure 4.7: Band 30 has shown the most elasticity at interest rate of 13%

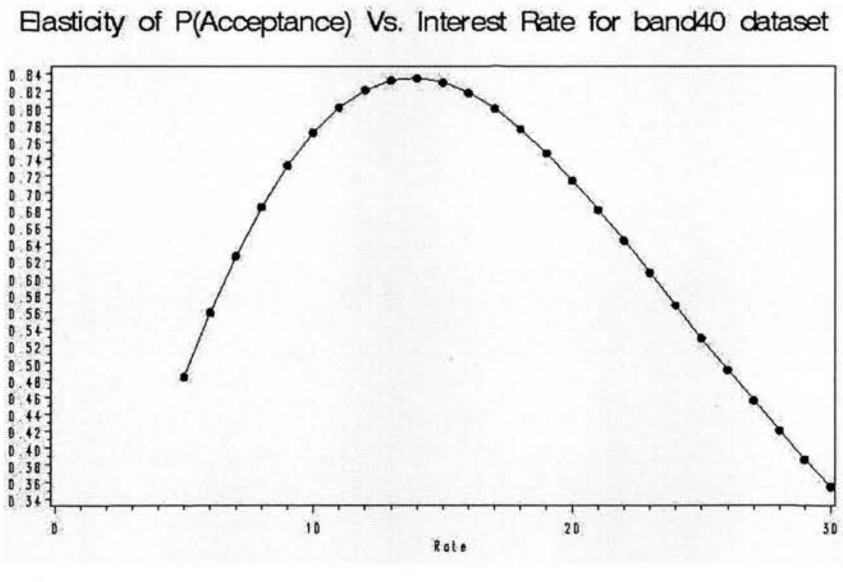

Figure 4.8: Band 40 has shown the most elasticity at interest rate of 14%

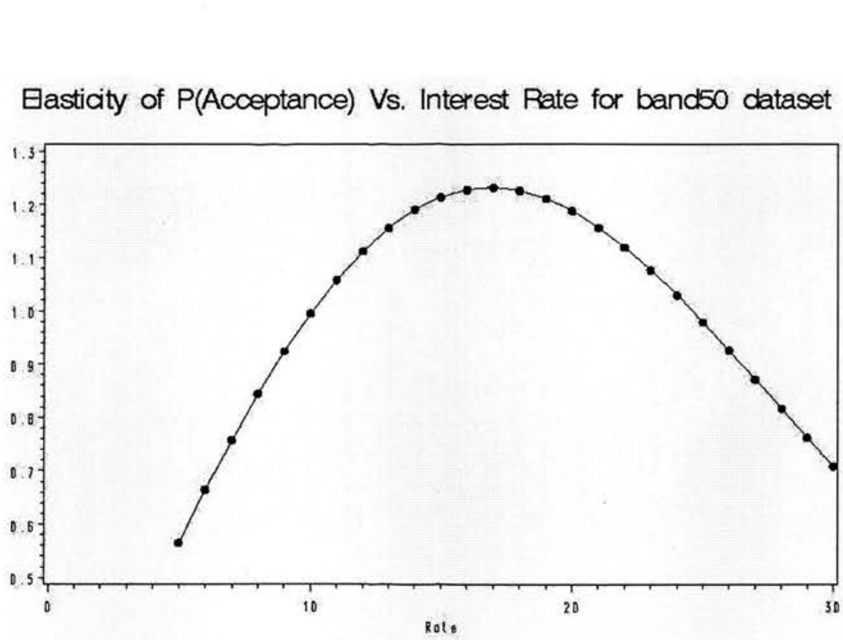

Figure 4.9: Band 50 has shown the most elasticity at interest rate of 17%

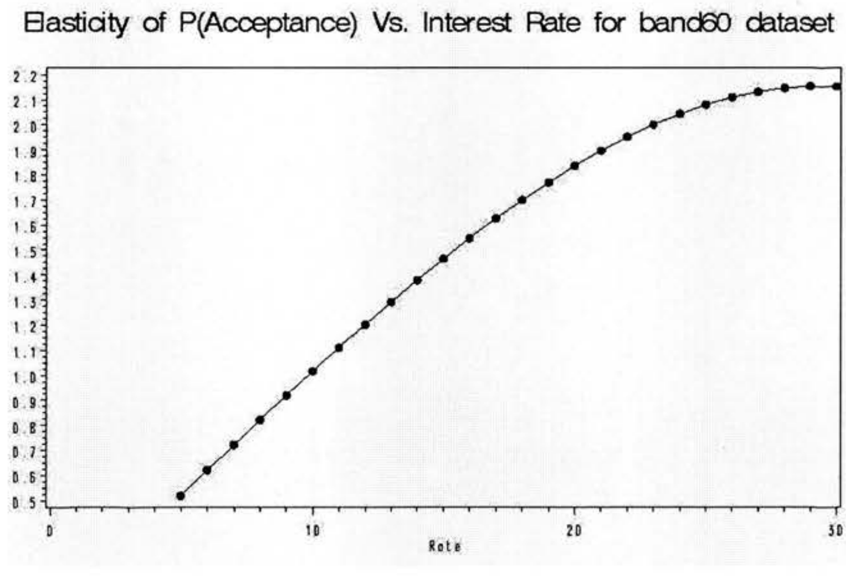

Figure 4.10: Band 60 has shown the most elasticity at interest rate of 29%

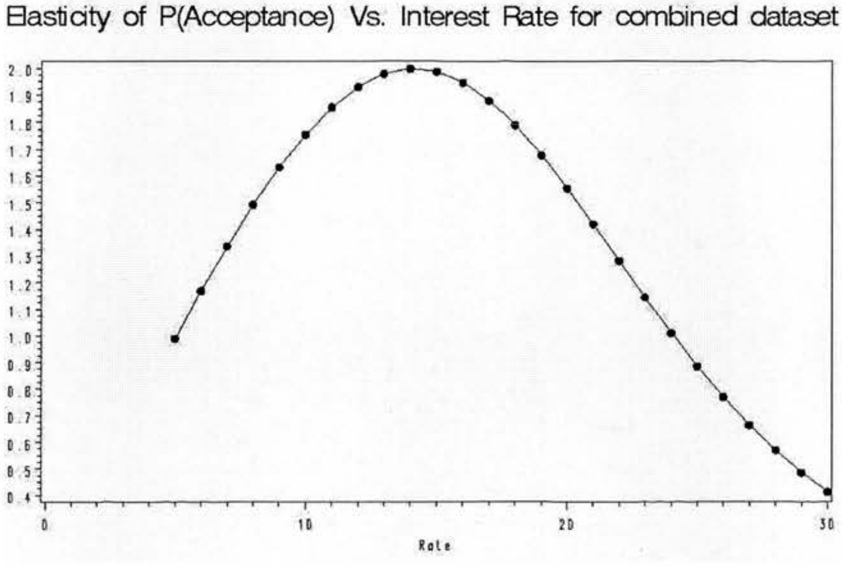

Figure 4.11: Combining all bands together has seen the most elasticity at interest rate around 14%

4.8 Bivariate Probit sample selection model

In this section the possibility of improving default probability estimation by applying bivariate Probit sample selection models is examined. We estimated with various model settings and found that only when using a lean model with less variables would the estimated correlations between the error terms in the bivariate Probit sample selection model become significant. However, the predictive ability is still slightly worse than the usual Probit model applied using area under ROC curve values as the performance indicator.

4.8.1 Background and previous research

When doing Credit Scoring, the data collected for analysis are normally pre-screened subject to previous scoring practices that have eliminated a substantial portion of applicants who were regarded as most likely to default or not profitable enough to keep as customers. In the sets of data collected for analysis, only the performance information of those who have been accepted is available, leaving the performance of those rejected unobservable (and those refusing to take the offer are also missing in the records). Based on these non-randomly selected sample data only, a traditional predictor trying to predict probability to default $p(D|x)$ (where D denotes default and x is the vector of a set of explanatory variables) is in fact modelling $p(D|A,x)$ (where A denotes accepts), and is assuming that $p(D|A,x)$ equals to $p(D|x)$.

When the assumption above is under question, using only the observed-accepts to represent the whole set (including accepts and rejects), results given by the maximum likelihood estimation will lead to bias, known as 'sample selection bias'. Heckman (1979) studied this selection bias in the model structured as below:

$$\begin{aligned} Y_1 &= \beta_1 x_1 + \varepsilon_1 \\ Y_2 &= \beta_2 x_2 + \varepsilon_2 \end{aligned}$$

where Y_1 and Y_2 are continuous random variables. x_1 and x_2 are vectors of independent variables. ε_1 and ε_2 are the errors. Y_1 is only observed when $Y_2 >= 0$

The dependent variables in Heckman's model are continuous. When outcomes are observed as binary results, a bivariate Probit model is more appropriate. Meng and Schmidt (1985) discussed the bivariate Probit models under various levels of observability of the dependent variables. Their model can be written as

$$\begin{aligned} Y_1^* &= \beta_1 x_1 + \varepsilon_1 \\ Y_2^* &= \beta_2 x_2 + \varepsilon_2 \end{aligned}$$

where Y_1^* and Y_2^* are continuous random variables that are not observable directly. The binary outcomes that are observable are Y_1 and Y_2

$$\begin{aligned} Y_1 &= 1 \ \ if \ \ Y_1^* &>& \ 0 \\ Y_1 &= 0 \ \ if \ \ Y_1^* &<=& \ 0 \\ Y_2 &= 1 \ \ if \ \ Y_2^* &>& \ 0 \\ Y_2 &= 0 \ \ if \ \ Y_2^* &<=& \ 0 \end{aligned}$$

The errors ε_1 and ε_2 are assumed to be normally distributed $N(0,0,S1,S2,\rho)$, where $S1$ and $S2$ are the variances of ε_1 and ε_2 respectively. Meng and Schmidt discussed different cases where the observability differs. In their case three (the 'censored Probit or partial partial observability'), Y_1 is observed if and only if $Y_2 = 1$ (Y_2 is observed for all cases). This case is similar to our credit scoring problem and therefore of special interest to us.

Greene (1992) and Boyes et al. (1989) both used a bivariate Probit model with sample selection to predict the probability of default and estimated card expenditure so that a profit oriented scoring approach is possible based on these estimates. Greene (1998) presented three statistical models to predict the default, expenditure and the number of derogatory reports in credit history and showed that results were quite different when sample selection factors were included in the models. However, their results did not provide indications of the models' *predictive performance* on the cross-validation sets.

Banasik et al. (2003) compared the prediction results in terms of the classification accuracy and area under ROC values from the bivariate Probit model with sample selection and those from original models based on accepted applicants only. They observed that small improvements with bivariate Probit model can sometimes be achieved, depending on the choice of risk bands and cut-off values selected.

Hand and Henley (1993) reviewed the methods of reject inference and claimed that reliable reject inference based on rejected applicants is not possible without additional assumptions being made. Banasik and Crook (2005) conducted analysis on a rare data set where almost all applicants were granted credit. They found that both the scope and effectiveness of reject inference is unaffected by the model leanness while still some benefits are possible with high rejecting rate.

4.8.2 Estimation results

The data used here is encoded as continuous variables using weights of evidence based on odds of acceptance. The weights of evidence is used because it yields far fewer dimensions than dummy variables and is quite helpful for a faster and easier convergence in the maximum likelihood estimation in the Heckprob routine in Stata.

Stepwise Probit models are fitted using two different selection criteria, one is p=0.002 and the other is p=0.05. When p value is 0.002, less variables were selected and therefore we gained a leaner model. Probit models were estimated on the training data then the predictive performance was evaluated using area under ROC based on the holdout data .

Our results shown in Table 4.4 and 4.5 [6] showed that the correlation parameter (between ε_1 and ε_2) in the bivariate Probit model is only significant when using a lean model. The significance value of the correlation coefficient is 0.015, when $p < 0.002$

[6]The dictionary of variables can be found in Table 3.6.

is the selection criteria. The result of the likelihood-ratio test (if the two equations in the model are independent) in the lean model is $\chi^2 = 6.94$ $Prob > \chi^2 = 0.0084$. We can reject the null hypothesis that the two equations are independent. This indicates that ignoring the selection process will give biased estimates in a lean model. The larger model (variables selected using $p < 0.05$) did not show a similar pattern. The result of the likelihood-ratio test (if the two equations in the model are independent) in the larger model is $\chi^2 = 0.00$ $Prob > \chi^2 = 0.9608$. The null hypothesis that the two equations are independent cannot be rejected.

Predictive performance on an independent holdout data set shows that the bivariate Probit model and Probit models are almost equally predictive in terms of area under ROC values. Nevertheless, the more complex model (variables selected with $p < 0.05$) is significantly more predictive (AuROC=0.7979) than the lean model ($p < 0.002$) (AuROC=0.7925).

Table 4.4: Bivariate Probit model with variables stepwise selected with significance value of 0.002

	Coef.	Std. Err.	z	$P > \lvert z \rvert$	[95% Conf.	Interval]
default						
loanapr1	0.2191171	0.0414665	5.28	0.000	0.137844	0.300390
cpi	0.5184754	0.0474135	10.94	0.000	0.425547	0.611404
wrst46al	0.5377122	0.0792789	6.78	0.000	0.382328	0.693096
timebank	0.3059594	0.0504106	6.07	0.000	0.207157	0.404762
ssrc4to6	0.3118096	0.0688254	4.53	0.000	0.176914	0.446705
socworst	0.2098031	0.0657783	3.19	0.001	0.080880	0.338726
loanbal2	-0.7548026	0.1653233	-4.57	0.000	-1.078830	-0.430775

loanbal6	-1.5859630	0.3956968	-4.01	0.000	-2.361515	-0.810412
spsetld	0.3207764	0.0877495	3.66	0.000	0.148791	0.492762
term	0.8631102	0.2440975	3.54	0.000	0.384688	1.341533
netincm	-1.1657040	0.3475501	-3.35	0.001	-1.846890	-0.484519
_cons	-1.8472910	0.0527410	-35.03	0.000	-1.950661	-1.743920
paid						
cpi	-0.6039442	0.0175323	-34.45	0.000	-0.638307	-0.569582
loanapr1	-0.5469236	0.0099439	-55.00	0.000	-0.566413	-0.527434
newbus	-0.0607446	0.0576190	-1.05	0.292	-0.173676	0.052187
loan_amt	-1.4419610	0.0661858	-21.79	0.000	-1.571683	-1.312239
tosettl1	-0.4955130	0.0513494	-9.65	0.000	-0.596156	-0.394870
snball6m	-0.4451250	0.0372410	-11.95	0.000	-0.518116	-0.372134
loanbal3	-0.5439364	0.1099655	-4.95	0.000	-0.759465	-0.328408
timadd1	-0.2130155	0.0275202	-7.74	0.000	-0.266954	-0.159077
gdscde2	-0.5306470	0.0522068	-10.16	0.000	-0.632971	-0.428324
internet	-1.3356410	0.1164642	-11.47	0.000	-1.563907	-1.107375
socsett	-0.5165921	0.0605816	-8.53	0.000	-0.635330	-0.397854
swrstcur	-0.2122401	0.0422736	-5.02	0.000	-0.295095	-0.129385
brand	-0.7150864	0.1072868	-6.67	0.000	-0.925365	-0.504808
age	-0.1093330	0.0247086	-4.42	0.000	-0.157761	-0.060905
loanbal2	-0.6094566	0.0918317	-6.64	0.000	-0.789444	-0.429470
mortbal	-1.6385940	0.2980171	-5.50	0.000	-2.222696	-1.054491
tosettl4	-0.7459363	0.2545549	-2.93	0.003	-1.244855	-0.247018
socworst	-0.2300384	0.0486660	-4.73	0.000	-0.325422	-0.134655
noopen6	-0.1552179	0.0379108	-4.09	0.000	-0.229522	-0.080914

gdscde3	-0.4151366	0.0817705	-5.08	0.000	-0.575404	-0.254869
timebank	-0.1029272	0.0240946	-4.27	0.000	-0.150152	-0.055703
no_store	-0.9392338	0.2438892	-3.85	0.000	-1.417248	-0.461220
ccjgt500	-0.8238198	0.2257602	-3.65	0.000	-1.266302	-0.381338
wrst46al	0.2929748	0.0590219	4.96	0.000	0.177294	0.408656
spl6mact	-0.3240852	0.0480901	-6.74	0.000	-0.418340	-0.229830
loanbal4	-0.5339032	0.1310558	-4.07	0.000	-0.790768	-0.277039
spl6m12	0.2024613	0.0564088	3.59	0.000	0.091902	0.313021
alcifdet	-0.3752600	0.0978557	-3.83	0.000	-0.567054	-0.183466
tosettl3	-0.6354947	0.1807832	-3.52	0.000	-0.989823	-0.281166
mor_rent	-0.7817131	0.2424378	-3.22	0.001	-1.256883	-0.306544
_cons	0.2124088	0.0071413	29.74	0.000	0.198412	0.226406
/athrho	0.2961050	0.1222285	2.42	0.015	0.056542	0.535668
rho	0.2877441	0.1121084			0.056481	0.489702

Table 4.5: Bivariate Probit model with variables stepwise selected with significance value of 0.05

	Coef.	Std. Err.	z	$P > \|z\|$	[95% Conf.	Interval]
default						
loanapr1	0.2983323	0.04767	6.26	0.000	0.204900	0.391764
cpi	0.6101280	0.050713	12.03	0.000	0.510733	0.709523
wrst46al	0.3580448	0.10866	3.30	0.001	0.145075	0.571014
timebank	0.2676062	0.055129	4.85	0.000	0.159555	0.375657
ssrc4to6	0.3148383	0.071293	4.42	0.000	0.175108	0.454569
socworst	0.2548701	0.070986	3.59	0.000	0.115740	0.394000

loanbal2	-0.7924271	0.196375	-4.04	0.000	-1.177314	-0.407540
loanbal6	-1.4821650	0.417788	-3.55	0.000	-2.301015	-0.663314
spsetld	0.3099023	0.090616	3.42	0.001	0.132298	0.487506
term	0.8055490	0.257671	3.13	0.002	0.300523	1.310575
netincm	-1.3158470	0.358635	-3.67	0.000	-2.018758	-0.612936
alcifdet	0.6051299	0.197823	3.06	0.002	0.217405	0.992855
age	0.1704406	0.054249	3.14	0.002	0.064115	0.276767
worst12	0.3693976	0.122369	3.02	0.003	0.129559	0.609237
spl6m12	0.2764318	0.105255	2.63	0.009	0.070137	0.482727
loan_amt	0.4010860	0.168954	2.37	0.018	0.069943	0.732229
tosettl2	0.7046708	0.321418	2.19	0.028	0.074704	1.334638
socsett	0.2552237	0.124509	2.05	0.040	0.011190	0.499258
ccjgt500	-1.1017680	0.5135	-2.15	0.032	-2.108209	-0.095328
_cons	-1.6901550	0.0872138	-19.38	0.000	-1.861091	-1.519219
paid						
cpi	-0.6089806	0.017601	-34.60	0.000	-0.643477	-0.574484
loanapr1	-0.5539081	0.010092	-54.89	0.000	-0.573688	-0.534129
newbus	-0.0784717	0.0581947	-1.35	0.178	-0.192531	0.035588
loan_amt	-1.4586650	0.067862	-21.49	0.000	-1.591672	-1.325657
tosettl1	-0.5169559	0.051777	-9.98	0.000	-0.618437	-0.415475
snball6m	-0.4497474	0.0373964	-12.03	0.000	-0.523043	-0.376452
loanbal3	-0.5403831	0.110517	-4.89	0.000	-0.756992	-0.323775
timadd1	-0.2093114	0.027634	-7.57	0.000	-0.263474	-0.155149
gdscde2	-0.5245726	0.05241	-10.01	0.000	-0.627294	-0.421852
internet	-1.3568380	0.117011	-11.60	0.000	-1.586175	-1.127501

socsett	-0.5419442	0.062302	-8.70	0.000	-0.664053	-0.419835
swrstcur	-0.2136315	0.042534	-5.02	0.000	-0.296996	-0.130267
brand	-0.7075670	0.107861	-6.56	0.000	-0.918970	-0.496164
age	-0.1156389	0.024993	-4.63	0.000	-0.164623	-0.066655
loanbal2	-0.6052619	0.092307	-6.56	0.000	-0.786181	-0.424343
mortbal	-1.6143770	0.299528	-5.39	0.000	-2.201441	-1.027312
tosettl4	-0.8178460	0.259038	-3.16	0.002	-1.325551	-0.310141
socworst	-0.2082265	0.049441	-4.21	0.000	-0.305129	-0.111324
noopen6	-0.1491900	0.038207	-3.90	0.000	-0.224074	-0.074306
gdscde3	-0.4184251	0.081961	-5.11	0.000	-0.579066	-0.257784
timebank	-0.1015448	0.024147	-4.21	0.000	-0.148871	-0.054219
no_store	-0.9209791	0.243301	-3.79	0.000	-1.397840	-0.444118
ccjgt500	-0.8412360	0.226	-3.72	0.000	-1.284187	-0.398285
wrst46al	0.2761811	0.0593490	4.65	0.000	0.159859	0.392503
spl6mact	-0.3143237	0.048462	-6.49	0.000	-0.409307	-0.219341
loanbal4	-0.4291300	0.13753	-3.12	0.002	-0.698683	-0.159577
spl6m12	0.2067295	0.056553	3.66	0.000	0.095887	0.317572
alcifdet	-0.3083031	0.099652	-3.09	0.002	-0.503617	-0.112989
tosettl3	-0.7089536	0.183059	-3.87	0.000	-1.067743	-0.350164
mor_rent	-0.7998439	0.243753	-3.28	0.001	-1.277591	-0.322097
loanbal6	-0.7482971	0.256081	-2.92	0.003	-1.250206	-0.246388
no_visa	-1.3502590	0.511138	-2.64	0.008	-2.352071	-0.348446
snw12tv	-0.5664918	0.220665	-2.57	0.010	-0.998988	-0.133996
no_deps	0.4072923	0.17275	2.36	0.018	0.068709	0.745876
term	0.2367741	0.108833	2.18	0.030	0.023465	0.450083

spsetld	0.0907106	0.042396	2.14	0.032	0.007616	0.173805
smo89	-0.3150311	0.150135	-2.10	0.036	-0.609291	-0.020772
_cons	0.2125689	0.007145	29.75	0.000	0.198565	0.226573
/athrho	0.0061964	0.126291	0.05	0.961	-0.241330	0.253723
rho	0.0061964	0.126287			-0.236752	0.248415

4.8.2.1 Comparisons of the predictive performance across different models

The results of the models are compared in Table 4.6. Probit002 was predicted using a Probit model with variables selected with $p < 0.002$. Heckprobit002 was predicted using a bivariate Probit model with variables selected with $p < 0.002$. Probit05 was predicted using Probit model with variables selected with $p < 0.05$. Heckprob05 was predicted using a bivariate Probit model with variables selected with $p < 0.05$

We conclude from Table 4.6 [7] that model Probit05 is slightly more predictive than model Probit002 when the performance is measured by area under ROC curves. (Ho: area(Probit002) = area(Probit05), $\chi^2(1) = 4.39$, Prob> $\chi^2 = 0.0362$) Comparing the results from bivariate Probit models and Probit models put together, we cannot reject the hypothesis that they are equally predictive. (Ho: area(Probit002) = area(Heckprob002) = area(Probit05) = area(Heckprob05), $\chi^2(3) = 5.66$, $Prob > \chi^2 = 0.1296$). From the Figure 4.12, plotting both ROC curves, we can hardly tell which model, the Probit002 model or Heckprob002 model, is more dominant.

[7] The standard errors for the area under ROC curves are calculated based on the nonparametric approach by DeLong et al. (1988). The asymptotic confidence intervals are calculated by assuming the distribution for the area under the ROC curve is normal.

Models	Obs	ROC Area	Std. Err.	-Asymptotic Normal– [95% Conf. Interval]	
Probit002	9471	0.7925	0.0115	0.77004	0.81502
Heckprob002	9471	0.7915	0.0114	0.76922	0.81386
Probit05	9471	0.7979	0.0114	0.77561	0.82026
Heckprob05	9471	0.7979	0.0114	0.77561	0.82025

Table 4.6: The predictive performance of lean model and a complex model

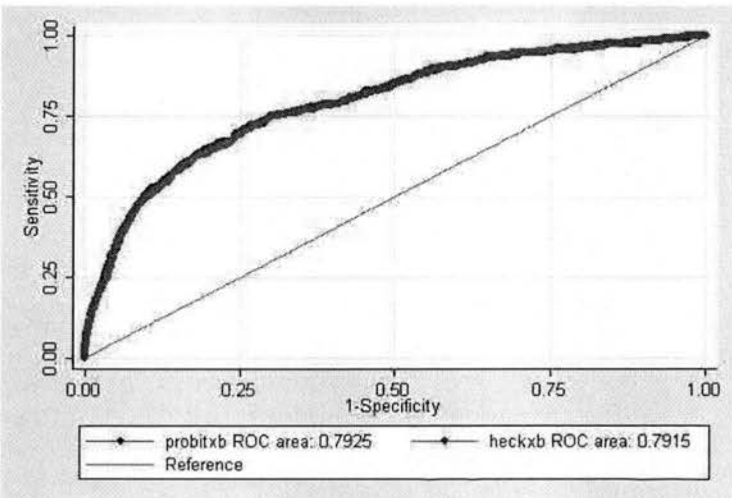

Figure 4.12: Compare ROC curves of Probit and Heckprob models

In conclusion, we do not find the default models to suffer from sample selection bias due the cases being included only if they accepted a loan offer. Therefore the acceptance inference may not be necessary for our data. On a reasonably large model, the sample selection bias is not significant. Comparisons of the predictive performance did not find significant improvement achieved through the Bivariate Sample Selection model than a normal Probit model. Note however that this test is subject to the weakness that we assumed ε_1 and ε_2 are normally distributed.

4.9 Indifference curves

As our data were collected from real customers, we can plot the mean indifference curves similar to those described by Keeney and Oliver (2005), in which indifference curves for individuals are described whereas our indifference curves are for the population [8] This section describes how the indifference curves are plotted on a 2 dimensional space(Rate vs Loan Amount). Also notice that in our dataset the customer chose the loan amount (albeit with some minor adjustment at some occasions by the lender) whereas in Keeney and Oliver (2005) the lender chose the credit line (limit). This however would not affect the validity of the construction of the indifference curves.

4.9.1 Indifference curves based on Logit model

The indifference curves for the customer can be plotted directly from the estimation results of a Logit model. Assuming the probability of acceptance p can be fitted using the functional form as:

$$\log \frac{p}{1-p} = \beta_0 + \beta_L * \log(L) + \beta_{APR} * \log(APR) + \beta_Z * Z \tag{4.3}$$

where Z is the vector of predictive variables other than the Loan Amount variable L and the Interest Rate variable APR. The set of variables in Z was selected using a stepwise

[8] Please note that the indifference curves drawn from each individual can be totally different from the curves drawn based on the population. Making inference based on population means and ignoring individual differences can lead to so called 'ecological fallacy' when the assumption of within group homogeneity does not hold. However, due to the nature of the way the data was collected, we cannot test each individual repeatedly to construct indifference curves for each individual. Mean indifference curves were used.

selection routine on the training set. For the values of the estimated coefficients β_0 β_L β_{APR} β_Z please see Appendix B: Table B.4. For each given probability of acceptance p, the indifference curve in the two dimensional L and APR space can be written as

$$1 = L^{\beta_L} * APR^{\beta_{APR}} * C$$

where $C = e^{\beta_0} * e^{\beta_Z * \overline{Z}} * \frac{1-p}{p}$. The $\overline{Z}$ is a vector in which the mean values of the variables Z have been used.

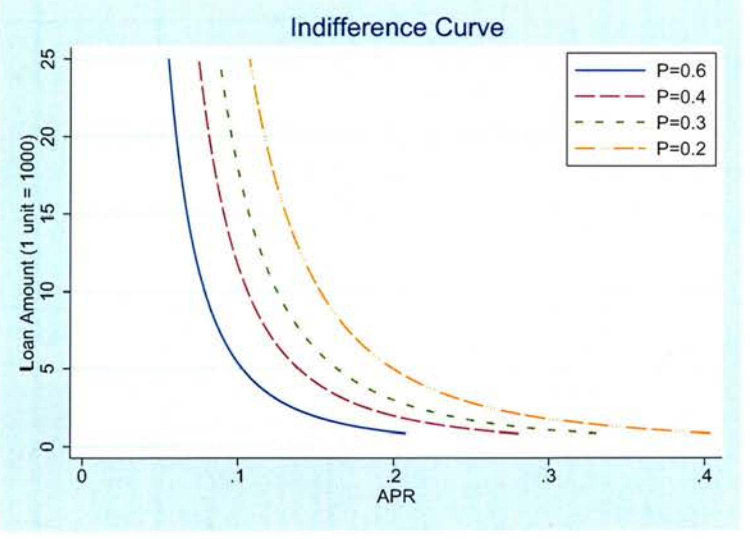

Figure 4.13: Indifference curves

All the points found on the same indifference curve in Figure 4.13 represent the equality in the attractiveness of the offers to the applicant. That means, given all other variables at their mean values, the average customer will accept the offer at the same probability if the combinations of loan amount and interest are on the same indifference curve. Different indifference curves indicate a different probability of acceptance. The curve which is closer to the origin point (0,0) has a higher probability of acceptance (The indifference curve with $p = 0.6$ shown in Figure 4.13, for example, has the highest probability of acceptance in the four indifference curves displayed). This shows that both lower loan amounts and lower interest rates increase the likelihood of the offer acceptance.

One of Keeney and Oliver's assumptions is that the interest rate has a negative impact on the probability of acceptance, which is consistent with our finding. For a given loan amount, the applicant will prefer being charged a lower interest rate. However, another assumption in the Keeney and Oliver model is that for any given interest rate the probability of acceptance will be lower for a lower credit line than a higher one. We have observed the opposite: that an acceptance of an offer is more likely if the applicant requested lower loan amount.

This result is interesting and merits some further discussion. To argue that on average applicants prefer to borrow less than more and are willing to pay a higher interest rate to be "able" to borrow less is inappropriate because each has the choice as to how much he wishes to borrow and can choose to borrow less if he wishes to. An appropriate explanation is perhaps more subtle. An individual may wish to buy a product now and has a choice as to how much to borrow. The more he borrows the lower his assets and the greater the chance he will be unable to finance emergency calls on his wealth. Borrowers must compare the marginal disbenefit from borrowing with the marginal benefit from consuming the good. In economic theory the more an individual borrows today the less he expects to consume tomorrow because of the repayments he must make tomorrow. Given a set of preferences between consumption today and consumption tomorrow there will be an optimum amount of borrowing which he desires (see Attanasio (1999)).

If an applicant increases the amount he wishes to borrow, a lower interest rate may be necessary to maintain the same probability of acceptance because the larger loan would, if the rate were constant, imply larger payments whereas these payments may be correspondingly lower if the rate is lower. In short, it would seem that applicants are making the choice to accept based on the cash outlays required to service the loan rather than the amount per se. In addition a larger loan may reduce utility by the borrower due to increased risk they will be unable to repay. To reduce this risk and maintain the same probability of acceptance a lower rate may be necessary.

4.9.2 Indifference curves using a different approach

As a check on the robustness onto our calculations concerning the shape of the indifference curve, we tested with a different functional form of the indifference curves. We assumed the form of the equation that describes the probability of acceptance as below

$$P = \alpha * L^{(-\beta)} * APR^{(-\gamma)} * Z^{(-\chi)}$$

where P is the probability of acceptance, L is the loan amount, APR is the interest rate charged and Z is a vector of principal components of other covariates retained. Taking log of both sides:

$$ln(P) = ln(\alpha) + (-\beta)ln(L) + (-\gamma)ln(APR) + (-\chi) * ln(Z)$$

To estimate the parameters α β γ χ, different options are available to treat variable P properly. The first is to assume P as another constant, which is very likely to be a wrong way. The second is importing the predicted values from a previous Probit model and use the predicted values as P, as will be implemented here. A third way is using Maximum Likelihood Estimation to estimate the parameters of the equation above. This has not been implemented.

Principal Component Analysis (PCA) is used to reduced the number of dimensions and to gain covariates that are orthogonal to each other. In this case we wished orthogonal covariates to reduce the chance of collinearity with APR and loan amount. The PCA is done by retaining the eigenvectors with the highest eigenvalues. All the eigenvalues are sorted and shown in Figure 4.14. In total 18 eigenvectors associated with eigenvalues higher than 1 were retained.

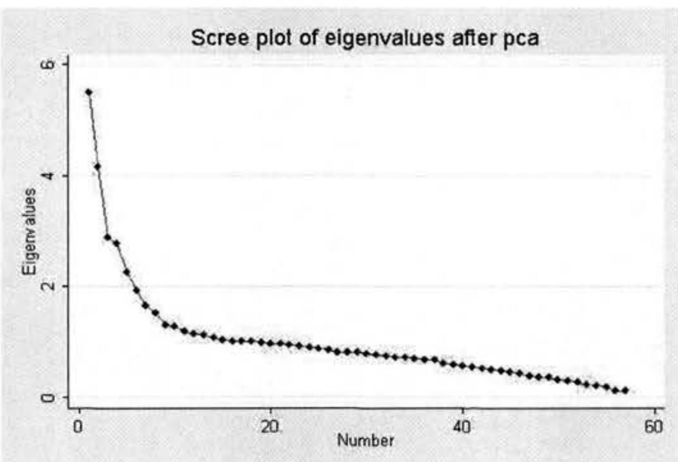

Figure 4.14: Eigenvalues after PCA

A Probit model was firstly called to generate the predicted probability values to be plugged into the variable P in equation below

$$ln(P) = b_0 + b_L ln(L) + b_{APR} ln(APR) + b_Z ln(Z)$$

Then APR can be expressed as

$$APR = L^{\frac{b_L}{-b_{APR}}} * Z^{\frac{b_Z}{-b_{APR}}} * e^{\frac{b_0 - ln(P)}{-b_{APR}}}$$

where P is the probability of acceptance (predicted by Probit model), L is the loan amount and Z is the vector of principal components. b_L, b_{APR},b_Z are the parameters estimated from OLS regression, as shown in Appendix B: Table B.5. The indifference

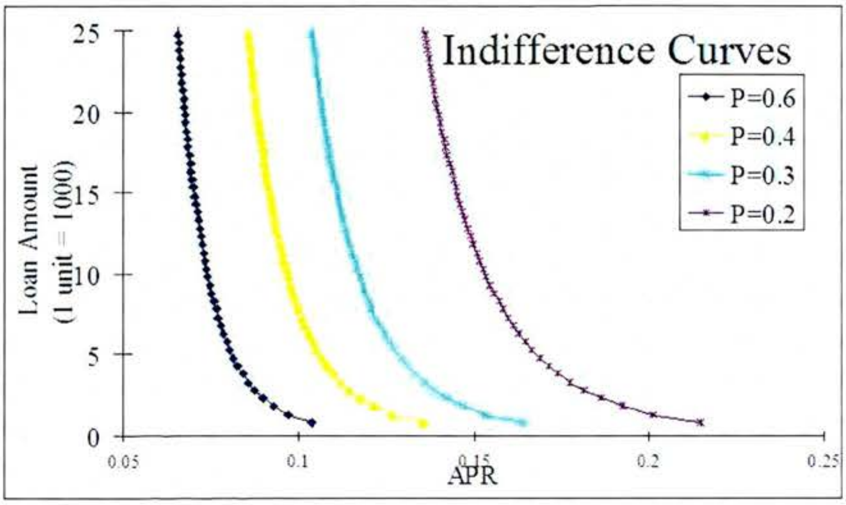

Figure 4.15: Indifference curves using a different approach

curves on a Loan Amount vs APR space are plotted in Figure 4.15 using the estimates of $ln(APR)$ and $ln(AMT)$ with Z and P variables treated as constant. The mean values of Z have been used in all indifference curves while each indifference curve is generated from a different P value. Similar to the indifference curves presented in previous subsection, the indifference curves in Figure 4.15 are in similar shapes. Same conclusion can be drawn that both lower interest rates and lower loan amounts increase the attractiveness of the offers.

4.10 Conclusion

This chapter reported the results of the modelling of consumer acceptance behaviour. Logistic regression was used to model the probability of acceptance on each band as well as for all bands combined together. Based on the estimates of the logistic regression, the acceptance elasticities with respect to interest rate were calculated. After modelling the P(Accept) directly, efforts were made to improve the prediction of default behaviour with the help of acceptance data using a bivariate Probit sample selection model. However, the predictive performance on the holdout sample is not improved using the bivariate Probit sample selection model. Finally, the indifference curves are plotted on the APR vs. Loan Amount space.

Chapter 5

Survival Analysis

5.1 Introduction

This chapter presents the results from survival analyses of default and paying back early. Section 5.2 describes the background of survival analysis followed by the structure of different survival models. Section 5.3 gives the details of the data used in the survival analysis. The following two sections are each devoted to the individual modelling of one of the two different types of customer behaviour, default and paying back early. Section 5.6 shows the results of modelling these two types of behaviour in a competing risks framework instead of separately. Finally, conclusions are drawn in section 5.7.

5.2 Introduction to survival analysis

Survival analysis deals with the modelling of time to event data. The event could be death in a biological study or a breakdown in an engineering problem. In an analysis of Credit Scoring, an event of interest could be the action of a customer to stop paying the monthly payment at a given month for some reason, either because of default or just switching to other lenders. One important advantage of survival analysis over static binary dependent variable models is that in survival analysis timing information has been utilized and modelled. This timing information can be very useful when the estimation of profit is needed, which we shall see in the next chapter.

When an event under study has occurred, we can call it a 'failure'. The probability that the failure occurs at a time T that is later than some arbitrary time t is called the survival function $S(t)$.

$$S(t) = Pr(t < T)$$

where t is the continuous duration time variable starting from $t = 0$. From the survival function $S(t)$ we can define the failure function $F(t) = 1 - S(t)$. The density function $f(t)$ of this failure function $F(t)$ can be expressed as

$$f(t) = \frac{\partial F(t)}{\partial t} = -\frac{\partial S(t)}{\partial t}$$

The conditional failure rate, or hazard function, defined as the event rate at time t conditional on that the subject having survived at least until time t, can be written as

$$h(t) = \lim_{\delta t \to 0} \frac{P(t \leq T \leq t + \delta t | T \geq t)}{\delta t} = \frac{f(t)}{S(t)} = -\frac{\partial S(t)}{\partial t}\frac{1}{S(t)} \tag{5.1}$$

The survival function $S(t)$ can also be derived from the hazard function as

$$S(t) = e^{[-\int_0^t h(u)du]}$$

5.2.1 Nonparametric model

Without making assumptions about the shapes of the hazard functions with respect to time, nonparametric models can be estimated to explore survival patterns. A Kaplan and Meier (1958) estimate of a survival function can be expressed as

$$S(t) = \prod_{j|t_j<t} (\frac{n_j - d_j}{n_j})$$

where $t_1 ... t_j$ are rank ordered survival times such that $t_1 < t_2 < .. < t_j$. n_j is the number at risk of the events of interest before time t_j. d_j is the number of observed events of interest at time t_j.

5.2.2 Parametric regression survival models

By assuming the form of parametric distributions of the hazard function, parameters can be estimated by maximum likelihood. Different distributions for the hazard function can be assumed, such as exponential, Weibull, gompertz, lognormal, loglogistic or gamma. For example, a Weibull proportional hazard model can have the hazard function as

$$h(t,X) = h_0(t)f(x)$$

where $f(x) = exp^{\beta * X}$ and the baseline hazard function is $h_0(t) = p * t^{p-1}$. p and β are the parameters to be estimated. When $p = 1$, $h_0(t) = 1$, the hazard rate is constant and the Weibull model becomes an exponential model where $h(t,X) = exp^{\beta * X}$. Figure 5.1 shows the baseline hazard function $h_0(t) = p * t^{p-1}$, (the Weibull function) with varying shape parameter p.

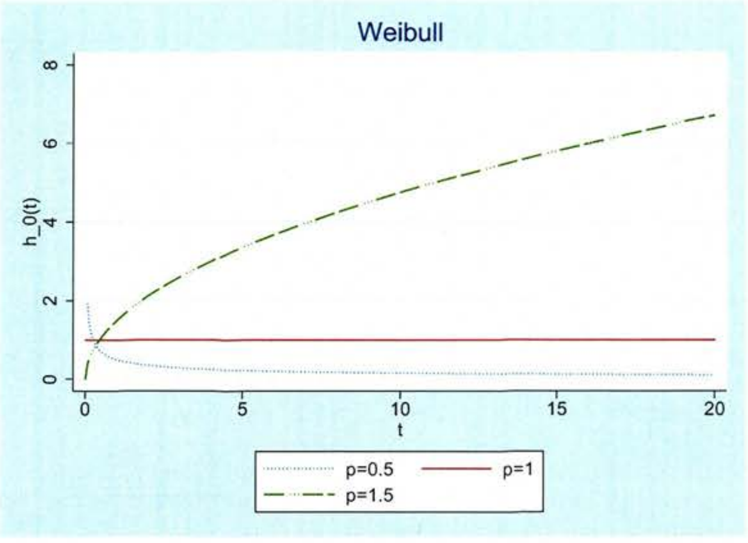

Figure 5.1: Example of hazard functions for Weibull models

5.2.3 Cox proportional hazards model

Forcing the hazard function to take a particular shape may be a disadvantage if it does not accurately represent the data. Placing no restrictions on the shape of the baseline hazard function $h_0(t)$, Cox (1972) argued that $h(t,x) = h_0(t) * f(x)$ and suggested that $f(x)$ ought to be modelled as $e^{\beta x}$. The hazard function can be written as

$$h(t,X) = h_0(t) * e^{\beta * X}$$

where $X = (x_1, x_2, ...x_K)$ is the time independent vector of K explanatory variables and the β is the vector of shape parameters to be estimated. Here the $h_0(t)$ is called the baseline hazard. The hazard ratio between x_i and x_j is irrelevant to the baseline hazard $h_0(t)$ as it can be cancelled as shown below

$$\frac{h(t,x_i)}{h(t,x_i)} = \frac{h_0(t) * e^{\beta x_i}}{h_0(t) * e^{\beta x_j}} = e^{\beta(x_i - x_j)}$$

Assuming j is the index of the ordered D distinct observed failure times from t_1 to t_D (for every i, $t_i < t_{i+1}$, and there are no ties) and R_j is defined as the risk set at time t_j, which equals the collection of the observations that are at risk of failure at time t_j, the likelihood can be written as the product of the conditional probabilities P_j, $L = \prod_{j=1}^{D} P_j$. The P_j is the conditional probability that for a particular failure at time t_j the failure is observed

$$P_j = \frac{e^{\beta x_j}}{\sum_{i \in R_j} e^{\beta x_i}}$$

where x_i is a vector of data with K variables for case i observed within risk set R_i. The estimate of β can be found by maximizing the natural logarithm of the partial likelihood function

$$L = \prod_{j=1}^{D} \frac{e^{\beta x_j}}{\sum_{i \in R_j} e^{\beta x_i}}$$

When there are tied failures to handle, Efron (1977) provided a closer approximation to the exact marginal likelihood which is computationally intensive. Breslow (1974) proposed a much faster approximation as below.

$$L_{Breslow} = \prod_{j=1}^{D} \frac{e^{\beta x_j}}{[\sum_{i \in R_j} e^{\beta x_i}]^{d_j}}$$

where d_j is the number of observations in the risk set R_j. Efron's approximation is a closer approximation than Breslow's method but at the price of higher computation demands.

In the data used for credit scoring the status of each account is usually recorded in a monthly fashion. A natural treatment would be to treat the time T as discrete and estimate the hazard function in a discrete logistic model (Cox (1972)).

$$\frac{h(t)}{1-h(t)} = e^{\beta X} \frac{h_0(t)}{1-h_0(t)}$$

Stepanova and Thomas (2002) found that compared to the PH-Cox model with continuous time assumed, this discrete logistic model has a better fit to the data (in terms of log-likelihood) but "almost no difference in the parameter estimates and no difference in the number of correctly classified accounts between the methods". In most of their estimations, they assumed continuous time and used the Breslow approximation to handle the ties since this is computationally the fastest method. In the analysis we carried out we also assumed continuous time and used Breslow's approximation.

5.2.3.1 Extensions of Cox models

Departing from the assumption of homogeneity with respect to the baseline functions , we can assume that the same proportional hazard assumption holds for each individual strata with the individual baseline hazard function h_{0g} for strata g. The hazard function then becomes

$$h_g(t,X) = h_{0g}(t)e^{\beta X}$$

where $g = 1,2,....G$ stands for the strata. This model is called the Stratified Cox model. Although the baseline hazard functions are individually estimated for each strata, the parameter vector β, is still constrained to be the same across the groups. This maintains the compactness of the model.

Another way of extending the Cox Models is called the Time Dependent Cox model. In the previous proportional hazard models all the covariates have been assumed to be unchanged from time zero to the end. Under situations where some covariates change over time, the time independent assumption can be relaxed. The hazard function can be written as

$$h(t,X(t)) = h_0(t)e^{\beta X+\delta X(t)}$$

where $X(t)$ represents the vector X at time t. As this extension introduces more complexity into the model, the potential gain from this approach might be overshadowed by the risk of possible overfitting. This can be seen in the later estimation results.

5.3 Data description

We are going to model time to default and time to paying back early. The time period in which a borrower is said to have defaulted is the first month in which he/she became two payments overdue. A binary indicator is used to mark the presence of a default event observed at a given time. This variable equals 1 when default occurs and 0 when the payments are made on schedule or is closed early. For paying back early we model the month in which the balance was paid off.

In the dataset prepared for the survival analysis, for each customer with recorded payment performance there are two new variables to be constructed.

1. The length of the duration time of observing the account state of keeping payments up to date with consequent exposure to default or payback early possibilities.

2. Censoring status, whereby an account is censored if the outcome of interest is not observed, such as borrower making scheduled payments throughout the observation period, paying back early when default is the outcome of interest or having defaulted when paying back early is the outcome of interest. An account is not censored otherwise.

Other variables used in the data set include those variables used in the estimation of the probabilities of default and of acceptance.

In the data investigated, although the fixed terms of the loans range from 24 to 84 months, the longest observation period is 26 months. Therefore most of cases are censored as their terms are longer than 26 months. Figure 5.2 illustrates this censoring situation.

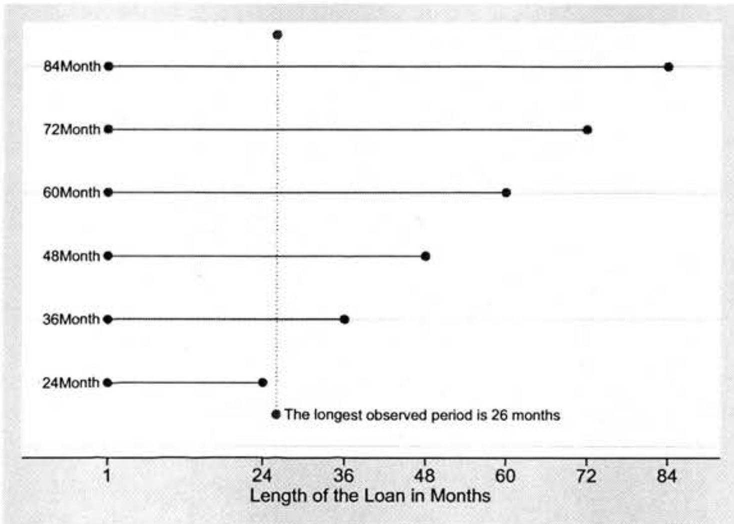

Figure 5.2: Observed loan terms

5.3.1 Description of the data using the Kaplan-Meier model

Since the Kaplan-Meier model makes no parametric assumptions, in the data exploration stage it can help us to investigate the overall hazard and survival functions without assuming distributional shapes in advance. The K-M survivor functions for default and paying back early are compared in Figure 5.3. Figure 5.4 compares the hazard functions of default and paying back early.

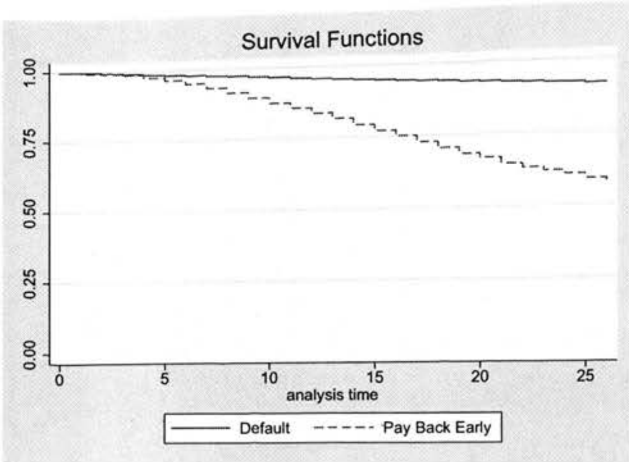

Figure 5.3: Kaplan-Meier survival functions, default and paying back early

Both figures indicate much higher hazard rates for paying back early than for default. While the shape of the default hazard function looks flat in Figure 5.4 because its relatively much smaller magnitude, in Figure 5.5 we can find the default hazard function at first quickly rises from 0.002 to 0.003 then at a slower pace increases and decrease until month 16, starting from when a sharp increase and decrease happens. The hazard function for the paying back early quickly increases from 0.003 to over 0.020 after 8 months. After 19 months, the paying back early hazard drops quickly from around 0.028 to 0.020 in three months before going up again. In the following sub sections we will compare the hazard functions of default and paying back early for different groups to explore the differences in more detail.

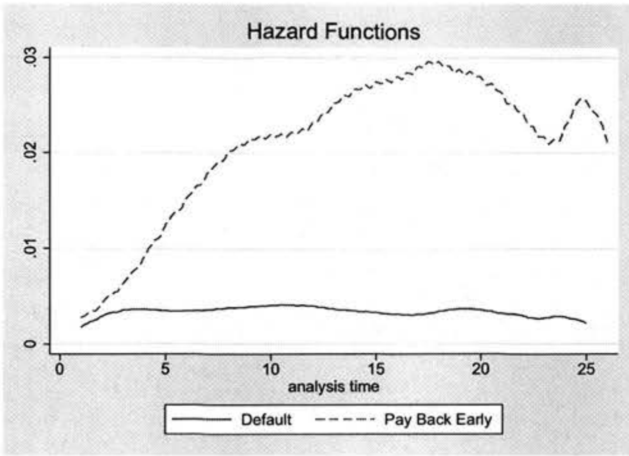

Figure 5.4: Kaplan-Meier hazard functions, default and paying back early

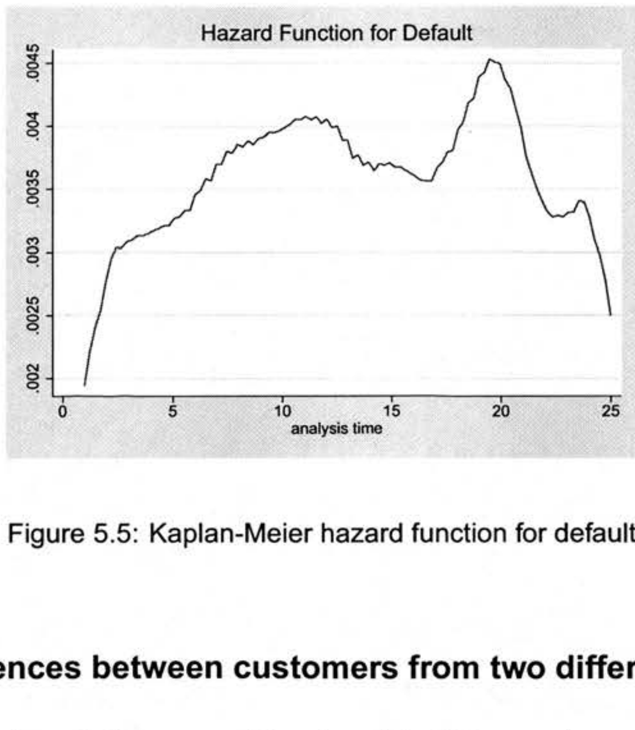

Figure 5.5: Kaplan-Meier hazard function for default

5.3.2 Differences between customers from two different brands

In the data investigated, there was still quite a big difference between the customers from two different brands. Brand1 has 8,823 customers while Brand2 has 22,549 customers. Their hazard functions shown in Figures 5.6, 5.7, 5.8 and 5.9 [1] reveal a large difference between the customers' behaviour of these two brands especially in terms of paying back early. Two possible explanations are available. The first is the length of the existing observation periods. For Brand1 customers, the maximum length of the observed period is 18 months while the maximum length of the observed period is 26 months for Brand2 customers. The second is the proportion of newly opened accounts. Of Brand1 accounts, 99.24% were new business while of Brand2 accounts, 74.22% were newly opened. One might expect that newly attracted customers would have a higher tendency to switch, but the paying back early hazard for Brand1 is actually lower than that for Brand2. Please note that the sudden rise of paying back early hazard for Brand1 customers after month 15 is due to the very few observations after that month.

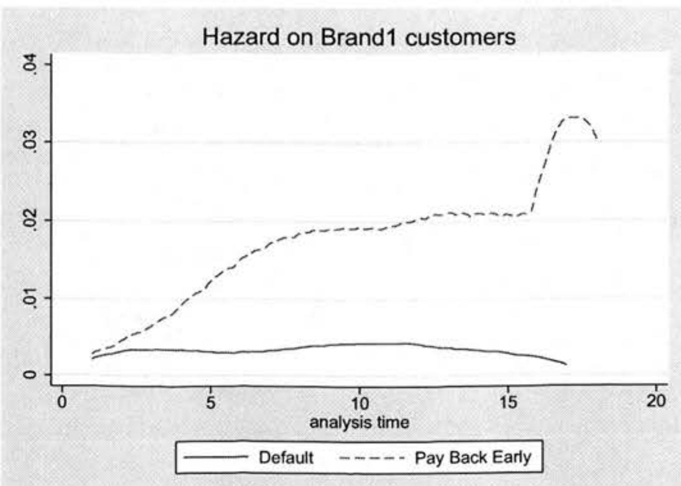

Figure 5.6: The KM hazard functions of the Brand1 customers for paying back early and default

[1] Figures 5.7 and 5.9 are supplemented because the much smaller magnitude of the default hazard functions render them look flat in comparison when plotted together with paying back early hazard functions.

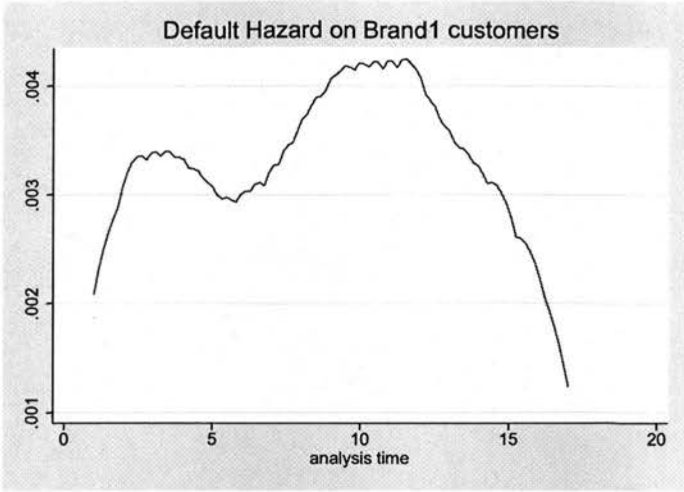

Figure 5.7: The KM hazard functions of the Brand1 customers for default

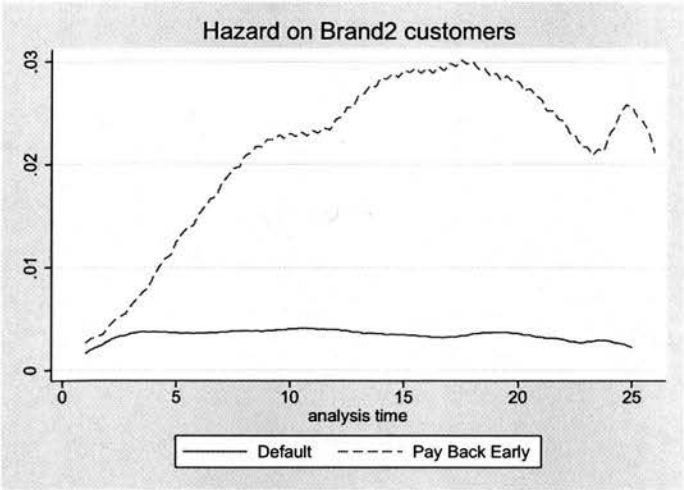

Figure 5.8: The KM hazard functions of the Brand2 customers for paying back early and default

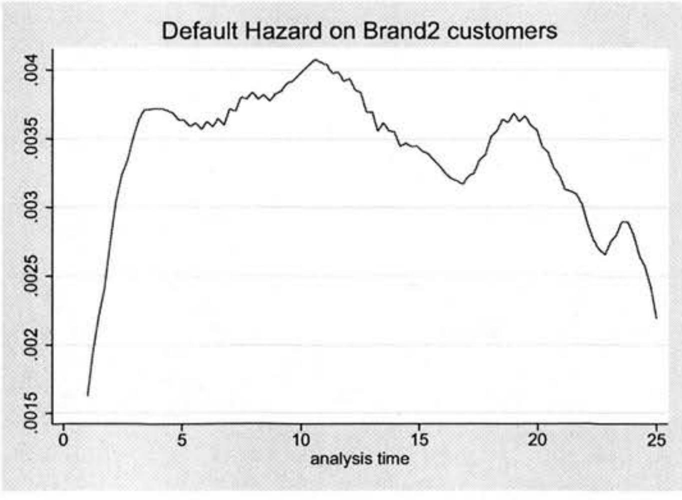

Figure 5.9: The KM hazard functions of the Brand2 customers for default

5.3.3 Difference between hazard functions with different loan terms

There are 6 categories of loan terms in the data set as shown in Table 5.1. Term 24 means the length of the loan is 24 months.

Loan Term	Frequency
24	2,149
36	8,038
48	5,864
60	11,109
72	867
84	3,345

Table 5.1: Size of different loan term groups

The hazard functions for different terms have different shapes. As shown in Figure

5.10, for customers taking the loan with the same term, the hazard functions for paying back early and default are not exactly same. Because of the much smaller magnitude of default hazard functions compared to that of paying back early hazard functions the default hazard functions look flat over the time but they are not so. They have been plotted separately in Figure 5.11.

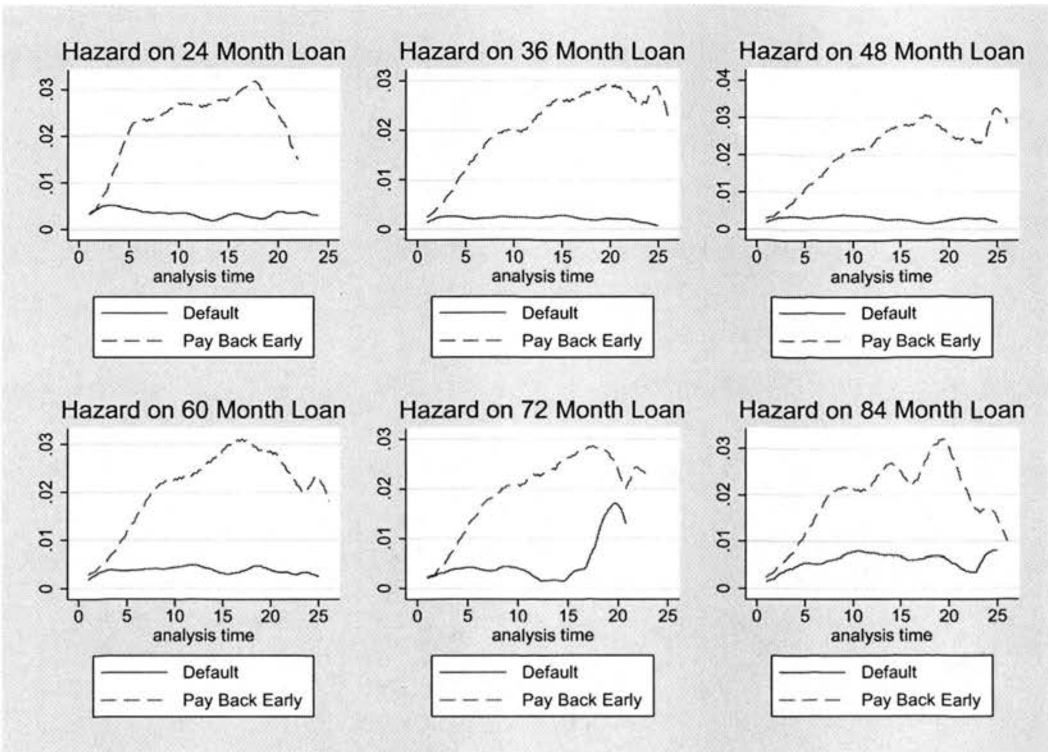

Figure 5.10: Compare the Hazard functions for loans with different terms.

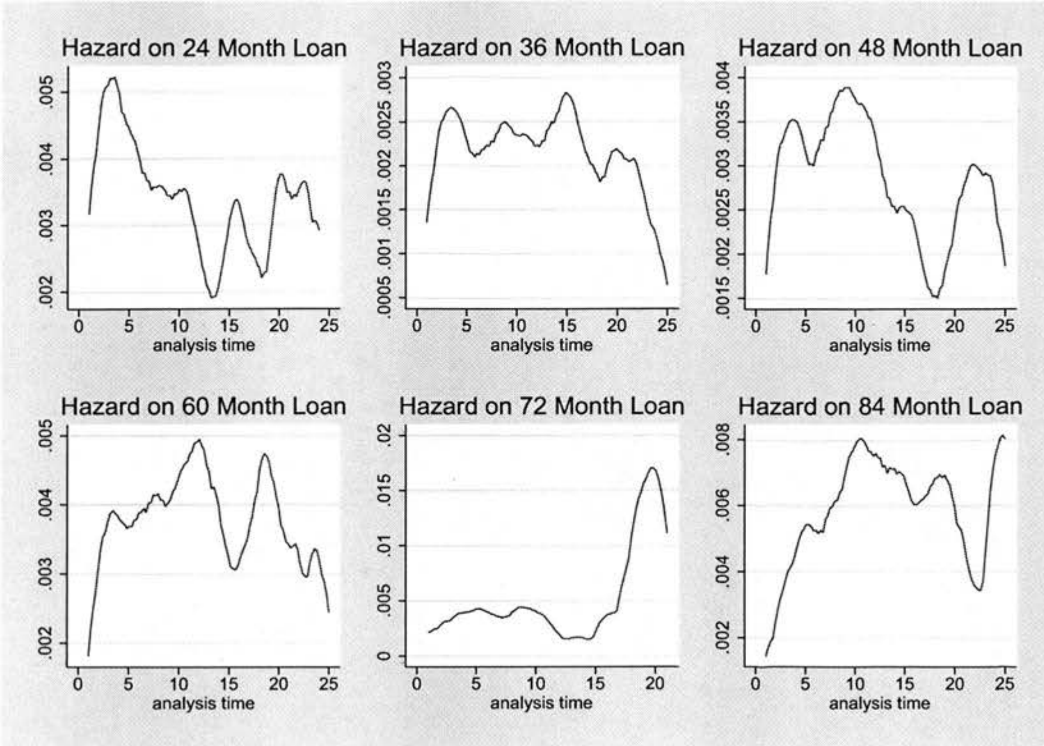

Figure 5.11: Compare the default hazard functions for loans with different terms.

5.4 Results of default modelling

This section explores the modelling of the time to default using parametric and semi-parametric models. To avoid dimensionality problems, the data is encoded using weights of evidence instead of dummy variables. Seventy percent of the data were randomly selected as a training set, on which the tests and estimations in the following sub sections were carried out. In total we have 21968 cases in the training set, 1,162 default, 6,063 paying back early and 14,743 Good cases. The cases for paying back early and Good were both treated together as Non-default.

5.4.1 Selection of explanatory variables

By testing the equality of the survival functions across the levels with the discrete explanatory variables, the log rank test or the Wilcoxon test can be carried out. If it is significant, then the null hypothesis that the survival functions are the same across the groups, can be rejected and this variable can be included into the model. However, when our variables have been recoded as continuous using weights of evidence, these tests were not carried out. Besides, the Stata manual for the survival analysis (release 9, page 300) suggests that "although it should be preferable to use log rank test, performing the log rank test or Cox (likelihood ratio) test makes little substantive difference with most datasets."

Forward stepwise selection was carried out by starting from fitting an empty model and one by one adding the most significant excluded term and then re-estimating the function. The test of significance is a Wald Test. The Wald Test is based on the estimated variance matrix of the estimators. The likelihood ratio test can also be used to test the significance of parameters and is preferred by many over the Wald test because fewer assumptions are made and the interpretation is easier. Our results showed that identical sets of variables have been selected by the Wald Test and the Likelihood ratio test. In the following sections we estimated hazard functions for default using alternative assumed distributions for the hazard functions: the Weibull and the Exponential. We then estimated PH Cox models.

5.4.2 Parametric regression using the Weibull distribution

Parametric regression estimation using the Weibull distribution was carried out and the results are reported in Table 5.2 and the variable dictionary can be found in Table 3.6.

The hazard and estimated survivor functions can be found in Figures 5.12 and 5.13.

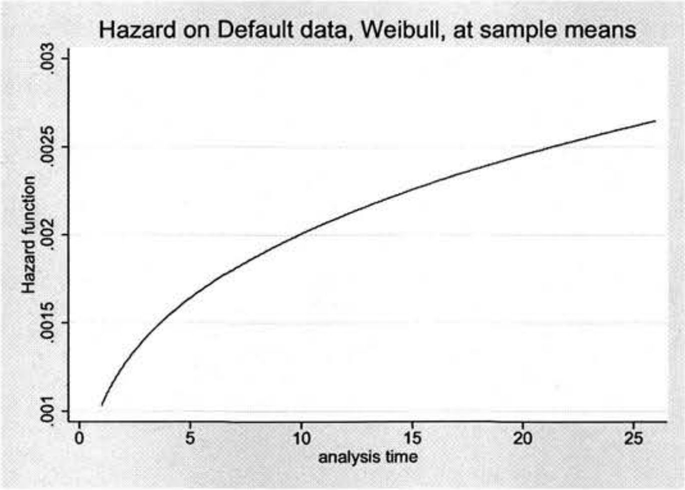

Figure 5.12: Hazard function, Weibull distribution, default data

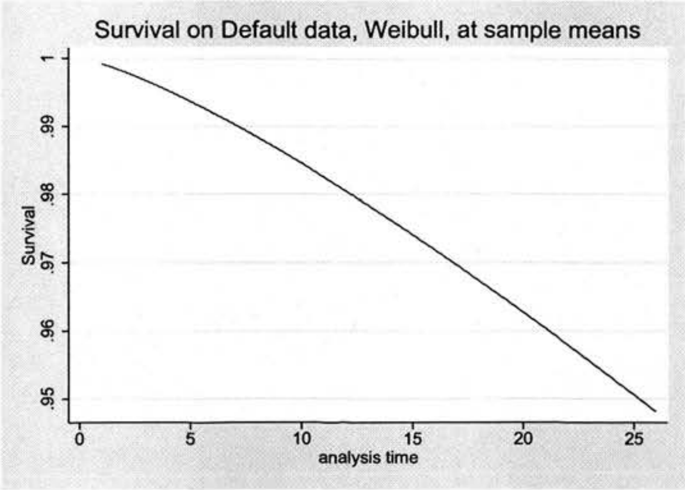

Figure 5.13: Survivor function, Weibull distribution, default data

Table 5.2: Weibull model estimates for default

_t	Coef.	Std. Err.	z	$P > \|z\|$	[95% Conf.	Interval]
loanapr1	-0.5348692	0.0353265	-15.14	0.000	-0.6041078	-0.4656306
cpi	-0.7395155	0.0388443	-19.04	0.000	-0.8156489	-0.6633822
term	-0.9456799	0.1094957	-8.64	0.000	-1.1602880	-0.7310723
timebank	-0.4994509	0.0839768	-5.95	0.000	-0.6640425	-0.3348594
spl6m12	-0.4935099	0.0654505	-7.54	0.000	-0.6217905	-0.3652294
ssrc4to6	-0.4304889	0.0701085	-6.14	0.000	-0.5678990	-0.2930789
loanbal4	-0.5236339	0.1091294	-4.80	0.000	-0.7375237	-0.3097442
spsetld	-0.5336005	0.0878954	-6.07	0.000	-0.7058724	-0.3613286
spl6m4	-0.4979006	0.1167386	-4.27	0.000	-0.7267040	-0.2690973
age	-0.4233402	0.1063859	-3.98	0.000	-0.6318528	-0.2148276
loanbal1	-0.6680406	0.1384581	-4.82	0.000	-0.9394134	-0.3966678
timadd1	-0.6679946	0.1801264	-3.71	0.000	-1.0210360	-0.3149533
inc_surp	-0.3277320	0.0970146	-3.38	0.001	-0.5178772	-0.1375867
searches	-0.4812325	0.2127586	-2.26	0.024	-0.8982318	-0.0642332
spvaldel	-0.4153225	0.1198893	-3.46	0.001	-0.6503012	-0.1803438
newbus	17.6266100	6.1234020	2.88	0.004	5.6249660	29.6282600
loanbal2	-0.4131220	0.1667517	-2.48	0.013	-0.7399493	-0.0862946
ccjgt500	-0.5696079	0.2531701	-2.25	0.024	-1.0658120	-0.0734037
brand	0.9648506	0.4020723	2.40	0.016	0.1768033	1.7528980
no_amex	-0.8484294	0.3949714	-2.15	0.032	-1.6225590	-0.0742997
mortbal	-0.5967126	0.2742134	-2.18	0.030	-1.1341610	-0.0592643
loanbal6	-0.2857244	0.1337473	-2.14	0.033	-0.5478644	-0.0235844
snball6m	-0.6260833	0.3564790	-1.76	0.079	-1.3247690	0.0726026
_cons	-6.4694950	0.1017219	-63.60	0.000	-6.6688660	-6.2701240
/ln_p	0.2541715	0.0264933	9.59	0.000	0.2022455	0.3060975
p	1.2893930	0.0341603			1.2241490	1.3581150
1/p	0.7755588	0.0205471			0.7363148	0.8168943

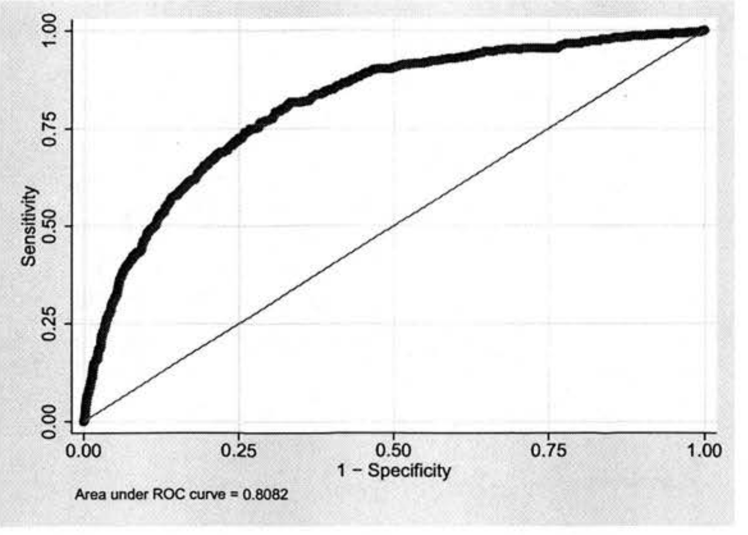

Figure 5.14: Area under ROC, Weibull distribution, default data

5.4.3 Parametric regression using the Exponential distribution

Parametric regression estimation using the Exponential distribution was carried out. The survival and hazard functions can be found in Figures 5.15 and 5.16.

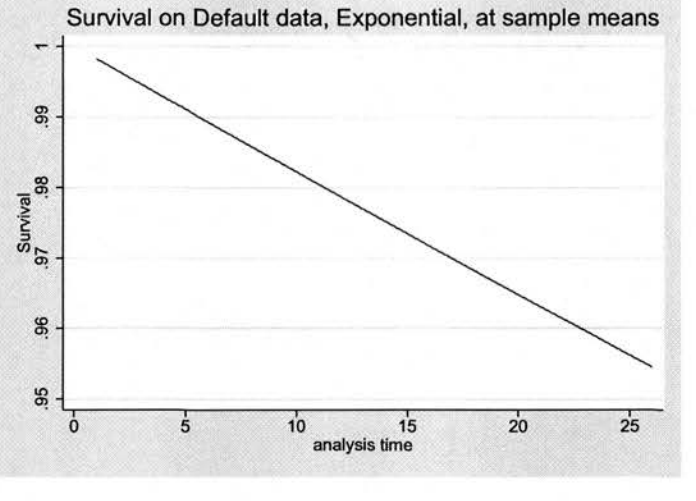

Figure 5.15: Survivor function, Exponential distribution, default data

The estimates are reported in Table 5.3. The area under the ROC curve is plotted in Figure 5.17. Although using a simpler model structure than a Weibull model, the AuROC on the holdout set is 0.8345, better than that of the Weibull model's 0.8082.

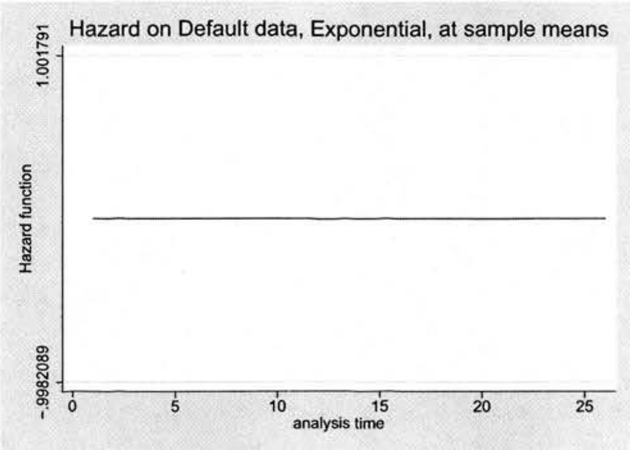

Figure 5.16: Hazard function, Exponential distribution, default data

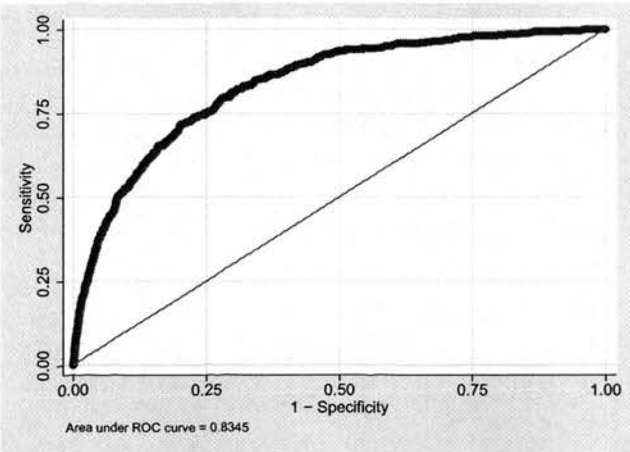

Figure 5.17: Area under ROC, Exponential distribution, default data

Table 5.3: Parametric regression results using Exponential Distribution on default

_t	Coef.	Std. Err.	z	$P > \|z\|$	[95% Conf.	Interval]
loanapr1	-0.5178199	0.0356924	-14.51	0.000	-0.5877758	-0.4478641
cpi	-0.7250035	0.0388431	-18.66	0.000	-0.8011347	-0.6488724
term	-0.9472073	0.1093950	-8.66	0.000	-1.1616170	-0.7327971
timebank	-0.4792742	0.0841515	-5.70	0.000	-0.6442081	-0.3143403
spl6m12	-0.4961234	0.0656931	-7.55	0.000	-0.6248795	-0.3673673
ssrc4to6	-0.4310955	0.0706320	-6.10	0.000	-0.5695317	-0.2926592
loanbal4	-0.5270958	0.1091070	-4.83	0.000	-0.7409415	-0.3132501
spsetld	-0.5602218	0.0891598	-6.28	0.000	-0.7349719	-0.3854718
spl6m4	-0.5125098	0.1219666	-4.20	0.000	-0.7515599	-0.2734597
age	-0.4046630	0.1064918	-3.80	0.000	-0.6133830	-0.1959430
loanbal1	-0.6682834	0.1382475	-4.83	0.000	-0.9392435	-0.3973233
timadd1	-0.6401188	0.1809254	-3.54	0.000	-0.9947260	-0.2855116
inc_surp	-0.3193763	0.0969892	-3.29	0.001	-0.5094716	-0.1292810
searches	-0.4943235	0.2128714	-2.32	0.020	-0.9115437	-0.0771032
spvaldel	-0.5027495	0.1294584	-3.88	0.000	-0.7564833	-0.2490157
newbus	19.1777700	5.9255270	3.24	0.001	7.5639540	30.7915900
loanbal2	-0.4180606	0.1668939	-2.50	0.012	-0.7451666	-0.0909547
ccjgt500	-0.5249048	0.2533820	-2.07	0.038	-1.0215240	-0.0282853
no_amex	-0.7933713	0.3964881	-2.00	0.045	-1.5704740	-0.0162689
loanbal6	-0.2929449	0.1337332	-2.19	0.028	-0.5550571	-0.0308326
mortbal	-0.5509227	0.2726835	-2.02	0.043	-1.0853730	-0.0164728
snball6m	-0.6134397	0.3571774	-1.72	0.086	-1.3134940	0.0866152
smo89	1.1302930	0.5553435	2.04	0.042	0.0418393	2.2187460
alcifdet	-0.5125166	0.2514563	-2.04	0.042	-1.0053620	-0.0196714
_cons	-5.6685750	0.0349650	-162.12	0.000	-5.7371050	-5.6000450

5.4.4 Cox proportional hazard model

The Cox proportional hazard model estimates are reported in Table 5.4. Figure 5.18 shows the baseline survivor function, where $S_0(t(j)) = \prod_{i=0}^{j-1}(1-h_j)$. h_j is the baseline hazard contribution. Figure 5.19 plots the ROC curve on the holdout set.

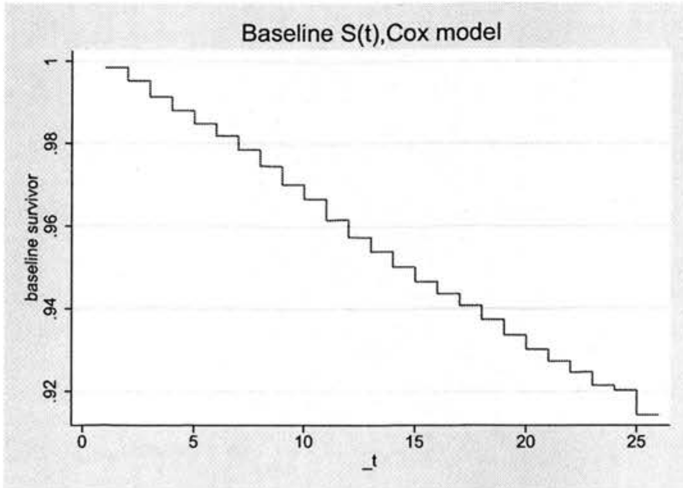

Figure 5.18: Baseline function S(t), Cox model, default data

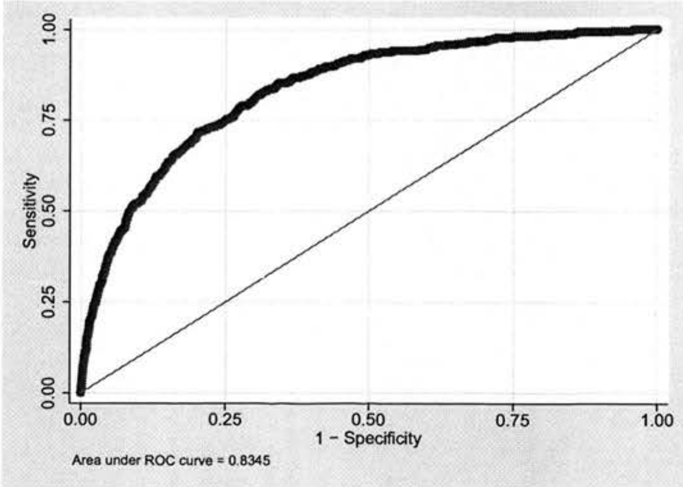

Figure 5.19: Area under ROC, Cox PH model, default data

Table 5.4: PH Cox model estimates on default

_t	Coef.	Std. Err.	z	$P > \|z\|$	[95% Conf.	Interval]
loanapr1	-0.535006	0.035208	-15.20	0.000	-0.604013	-0.466000
cpi	-0.728857	0.038837	-18.77	0.000	-0.804975	-0.652739
term	-0.942529	0.109279	-8.62	0.000	-1.156712	-0.728346
timebank	-0.483651	0.084039	-5.76	0.000	-0.648365	-0.318936
spl6m12	-0.487245	0.065402	-7.45	0.000	-0.615430	-0.359060
ssrc4to6	-0.419171	0.070046	-5.98	0.000	-0.556458	-0.281883
loanbal4	-0.519626	0.109106	-4.76	0.000	-0.733471	-0.305782
spsetld	-0.540380	0.087629	-6.17	0.000	-0.712129	-0.368631
spl6m4	-0.466104	0.116628	-4.00	0.000	-0.694689	-0.237518
age	-0.400648	0.106515	-3.76	0.000	-0.609413	-0.191882
loanbal1	-0.662516	0.138324	-4.79	0.000	-0.933626	-0.391406
timadd1	-0.653350	0.180490	-3.62	0.000	-1.007103	-0.299596
inc_surp	-0.317062	0.096913	-3.27	0.001	-0.507008	-0.127116
searches	-0.481334	0.212881	-2.26	0.024	-0.898573	-0.064095
spvaldel	-0.423302	0.119776	-3.53	0.000	-0.658058	-0.188546
newbus	19.333330	5.928404	3.26	0.001	7.713872	30.952790
loanbal2	-0.417280	0.166892	-2.50	0.012	-0.744382	-0.090178
ccjgt500	-0.555103	0.253246	-2.19	0.028	-1.051456	-0.058749
loanbal6	-0.289933	0.133673	-2.17	0.030	-0.551926	-0.027940
no_amex	-0.806257	0.396196	-2.03	0.042	-1.582786	-0.029727
mortbal	-0.551189	0.272541	-2.02	0.043	-1.085359	-0.017018
snball6m	-0.628288	0.357366	-1.76	0.079	-1.328713	0.072136

5.4.4.1 Predictive performance using area under ROC

The cumulative hazard probability in the first 12 months or 24 months can be used to represent the default behaviour. This cumulative hazard probability equals 1 minus the probability of survival until 12 months or 24 months. This survival function $S(t,x)$, the probability of survival until time t for a subject with explanatory variable vector x under the proportional hazards assumption, can be expressed as

$$S(t,x) = e^{-\int_0^t h_0(u)e^{\beta x}du} = e^{-e^{\beta x}*\int_0^t h_0(u)du} = S_0(t)^{e^{\beta x}}$$

where $h_0(t)$ is the baseline hazard function that is only related to the duration time variable t. Now,

$$Prob(default\ within\ time\ t) = 1 - S(t,x) = 1 - e^{-e^{\beta x}*\int_0^t h_0(u)du}$$

For a fixed given value of time t, the probability of default within time 0 to t is monotonically changing with the exponentiated linear prediction $e^{\beta x}$, so called the relative hazard. When measuring the predictive performance using Area under ROC curves on the binary outcome classifiers, it is only the relative size of the predicted numerical values that matters. This relative hazard (or the hazard ratio) value can then be used instead of the actual probability of default in calculating the Area under the ROC values on the training and holdout data.

The area under the ROC curve on the holdout sample is 0.8345. For comparison, using the Logistic Regression model, the area under the ROC curve on the holdout set is 0.8339. The previous Exponential model achieved an AuROC of 0.8345 on the holdout set. Considering the randomness, the difference in the predictive power between those models, in terms of the area under ROC curve, is mostly negligible. The Weibull model, however, is the poorest performing model with AuROC of only 0.8082.

5.4.5 Test of proportional hazard assumption

The proportional hazard assumption is extremely important for the Cox models and other parametric regression models that are consistent with the proportional hazard assumption, for example, the Exponential model estimated earlier.

5.4.5.1 Graphical assessment of PH assumption

Hosmer and Lemeshow (1998) (chapter 6.3) describes methods to test the proportional hazard assumptions. One of the methods that can be used to test the violations of the proportional hazard assumption on discrete variables in simpler models is the graphical assessment method. For each level of of the nominal variable, a curve can be plotted. This can be log-log plots ($-ln(-ln(survival\ probability))$ vs $ln(analysis\ time)$. Parallel curves indicate the non-violations of the proportional hazard assumption.

Or as pointed out by Garrett (1997), two curves can be plotted by displaying predicted survival probability from the Cox model along with the observed probability from Kaplan-Meier models. The closer the observed values are to the predicted, the less likely the assumption is to be violated. One problem with the graphical assessment method is that eyeballing is difficult and subjective. Another limitation is its applicability only to simpler models with nominal covariates. Besides, too many levels within those nominal covariates gives one graph with many curves that are difficult to tell apart.

5.4.5.2 Testing the PH Assumption using scaled Schoenfeld residuals

Grambsch and Therneau (1994) proposed that the test of a zero slope in a generalized linear regression of a scaled Schoenfeld residuals of time is equivalent to a test of the existence of the constant log hazard ratio over time. The results for the global scaled Schoenfeld residuals test, as displayed in Table 5.5, show that the null hypothesis that there is a zero slope, has to be rejected (Prob>chi2 = 0.0000).

Table 5.5: Test of proportional hazards assumption

	rho	chi2	df	*Prob > chi2*
loanapr1	0.04959	3.02	1	0.0821
cpi	-0.01456	0.26	1	0.6114
term	-0.08801	9.45	1	0.0021
timebank	-0.04636	2.57	1	0.1088
spl6m12	0.01102	0.14	1	0.7040
ssrc4to6	-0.04981	3.07	1	0.0797
loanbal4	-0.05680	3.58	1	0.0585
spsetld	0.07621	6.65	1	0.0099
spl6m4	-0.04137	2.21	1	0.1371
age	-0.00056	0.00	1	0.9844
loanbal1	0.00492	0.03	1	0.8653
timadd1	-0.04016	1.83	1	0.1758
inc_surp	-0.09487	10.64	1	0.0011
searches	0.03680	1.55	1	0.2127
spvaldel	0.01139	0.16	1	0.6905
newbus	-0.05848	3.85	1	0.0499
loanbal2	-0.04155	1.91	1	0.1674
ccjgt500	-0.02504	0.74	1	0.3883
loanbal6	0.00140	0.00	1	0.9623
no_amex	-0.03270	0.97	1	0.3235
mortbal	0.02903	0.99	1	0.3209
snball6m	-0.02022	0.39	1	0.5320
global test		74.15	22	0.0000

That means there are violations in the assumption of proportional hazards. When assessed individually, the covariates that clearly violate the proportional hazards assumption are term, spsetld, inc_surp, newbus (the p value threshold is set as 0.05). However, graphically it is still not easy to check the violation. Comparison of Figures 5.20 and 5.21 shows that it is not easy to graphically judge whether the slope is zero or not.

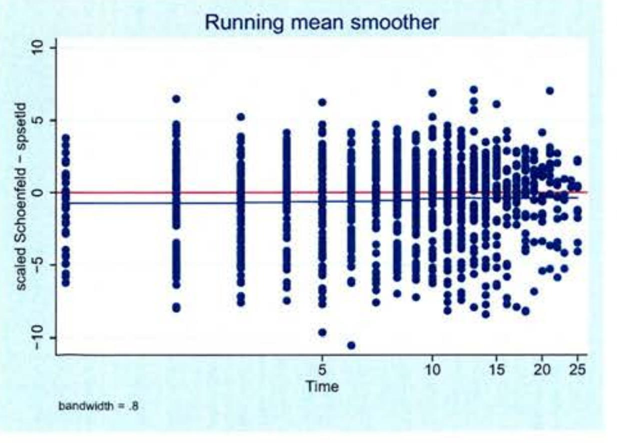

Figure 5.20: Scaled Schoenfeld Residuals for spsetld, violating PH Assumption, default data

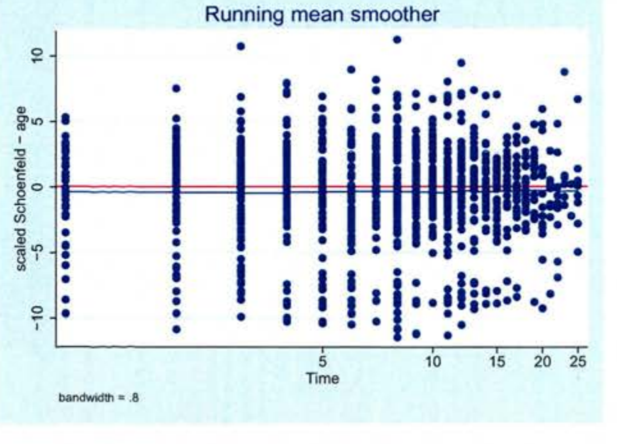

Figure 5.21: Scaled Schoenfeld Residuals for AGE, non violating PH Assumption, default data

5.4.5.3 Using time dependent covariates to test the PH assumption

Another method of testing the proportional hazard assumption was described by Hosmer and Lemeshow (1998) following Schoenfeld (1982) and Grambsch and Therneau (1994) by creating time dependent covariates which are the interactions between those covariates to be tested and the log of survival time. If those newly generated time dependent covariates enter the Cox model with significant parameters then the PH assumption is violated. Because of the limits of the software package used (Stata Stcox), which refuses to run when too many time dependent covariates are entered into the model, those newly created time dependent covariates were split into three groups, each group enters a Cox model with other time-independent covariates. The results in Appendix C: Tables C.1, C.2, and C.3 show the estimates for those three models.

Similar to findings using scaled Schoenfeld residuals, four variables (term, spsetld, inc_surp and newbus) were found to be significant. Another four covariates (loanbal4, spl6m4, loanbal2 and no_amex) were also found to be violating the PH assumption according to this test.

5.4.5.4 Performance of the Cox model with time dependent covariates

Previous tests on the proportional hazards assumptions indicated violations of the assumption and pointed out four covariates that may be time dependent. The four variables are term, spsetld, inc_surp and newbus. A Cox model with time dependent covariates is therefore constructed on the training data. The time dependent covariates were constructed as $X * t$, where t is the duration time and X is the set of four time dependent covariates that have been shown to be violating the PH assumption in both of previous tests.

The expectation is that the Cox model with time dependent covariates should improve the predictive performance on the holdout data. Surprisingly, this is not what was observed. The area under the ROC curve on the holdout set is 0.8289, lower than the area under the ROC curve from the original Cox model of 0.8345. Figure 5.22 plots the ROC curves on the holdout set and the estimates are listed in Appendix C: Table C.4. Therefore we retain these variables in the hazard function.

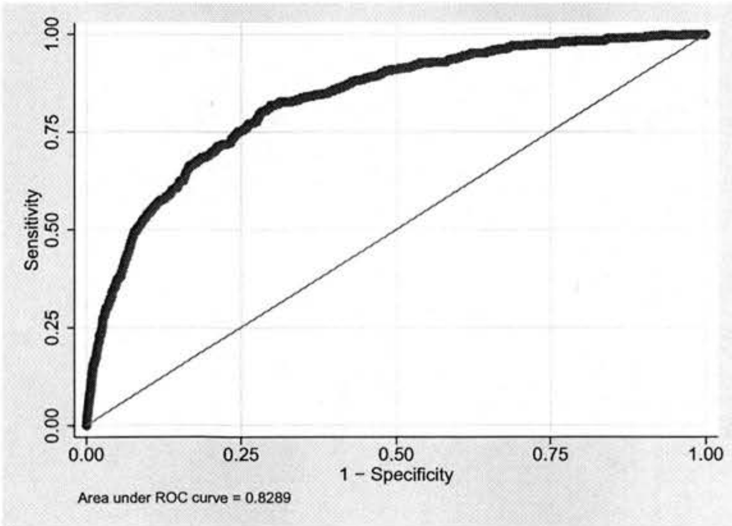

Figure 5.22: Area under ROC, Cox model with time dependent covariates, default data

5.4.6 Conclusion

In the previous analysis we have estimated the parameters for different survival analysis models of default. The predictive performance comparisons in terms of area under the ROC curve on the holdout set show that the Exponential model and the PH Cox model are as competitive as a Logistic Regression model in predicting default. More complex models such as the Cox model with time dependent covariates added, do not predict as well as expected. They might be suffering from the problem of over fitting.

5.5 Results of paying back early modelling

Previous results in section 5.3.1 have shown that the hazard functions for behaviours of default and paying back early are totally different. Stepanova and Thomas (2002) compared the modelling approaches of the paying back early behaviour using PH Cox models and Logistic Regression. They used two alternative definitions of paying back early in the modelling comparison. The first type is for the loan to be paid off early within the first 12 months. The second is for the loan to be paid off between month 12 and month 24 should the loan not have been paid off in the first 12 months. They found a stronger effect of the term arrangement (especially the remaining time-to-maturity of the loan) on the probability of paying back early than on the probability of default.

In our analysis, we will not predict the probability of paying back early from month 12 to month 24 since we are more interested in estimating profitability at the time of application. This profitability estimation requires the paying back early probabilities estimation at the time of application.

For the same reason, the behaviour of paying back early is defined as the observation that the loan has paid back within the whole duration of the loan. Because we have a limited observation period (only the first 26 months), our definition of paying back early is the observation of the outcome of the customer to pay back early within the first 26 months. Most of the variables used (except the two continuous variables, the loan amount L and loan APR) in the models are coded using the weights of evidence, which are calculated using the odds of paying back early.

Both parametric and semi parametric survival models will be fitted to the data. Two proportional hazards parametric survival models, Weibull and exponential models, will be used. Two accelerated failure time models, the Lognormal and Loglogistic models, will be tried as well. These two models are introduced because of the difference observed in the shape of the hazard functions compared to the hazard functions of the default models.

5.5.1 Parametric proportional hazards modelling results

The hazard functions for the Weibull and exponential models can be found in Figures 5.23 and 5.24. Because of the way the hazard functions are parameterized, their hazard functions structures are constrained. For the exponential model, the hazard function has to be held constant, which apparently differs from the real functional form of the hazard function as we have observed in the Kaplan-Meier model. It is therefore not surprising to find that the Lognormal and Loglogistic models (to be shown in the next sub section) have better fits for the hazard functions than Weibul and exponential models. The estimates for the two models can be found in Appendix C: Tables C.5 and C.6.

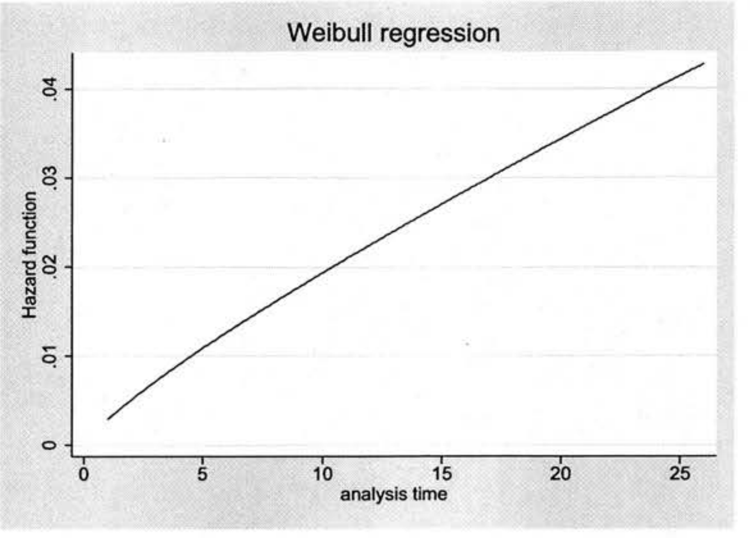

Figure 5.23: Hazard function, Weibull model, paying back early data

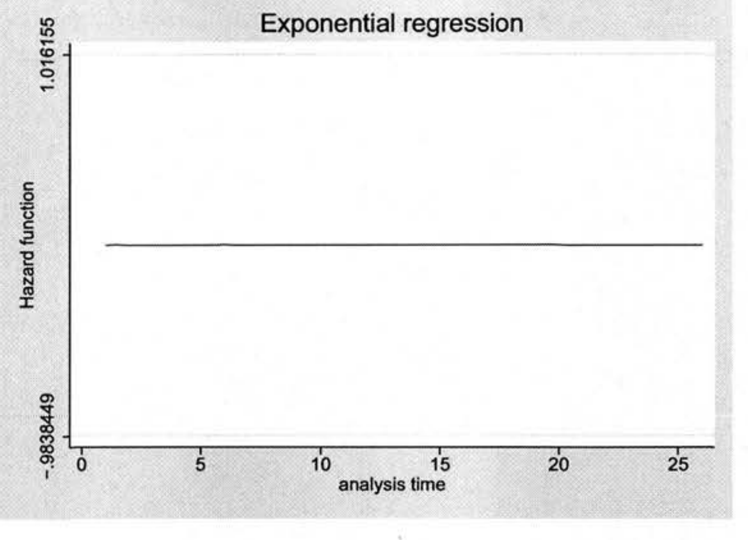

Figure 5.24: Hazard function, Exponential model, paying back early data

5.5.2 PH Cox model results

The hazard function for the PH Cox Models is shown in Figure 5.25. Because the model structure of PH Cox model is semi-parametric, the estimated hazard function is the closest to the observed hazard functions estimated by the Kaplan-Meier model. The hazard goes up initially and the speed of increase decreases until reaching around 19 months. After then the hazard of paying back early drops quickly. The estimates for the PH Cox model can be found in the Table C.7 within the appendix section.

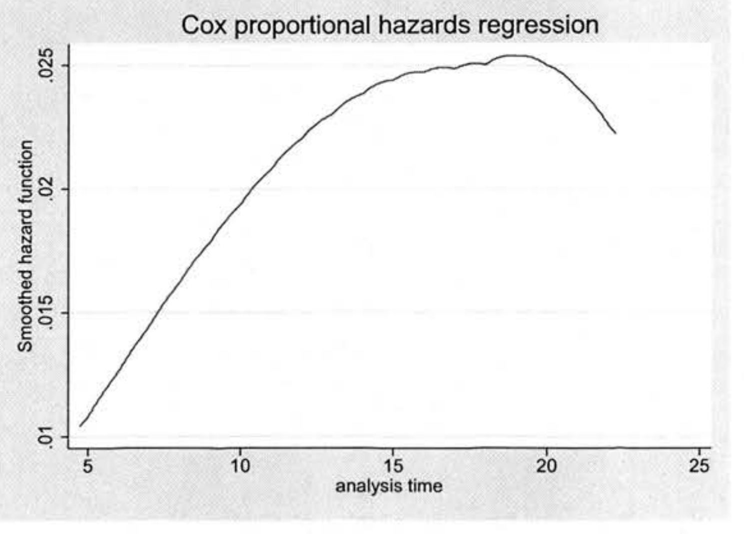

Figure 5.25: Hazard function, PH Cox model, paying back early data

5.5.3 Parametric accelerated failure time models

The hazard functions estimated from Lognormal and Loglogistic models (shown in Figures 5.26 and 5.27 respectively) more closely capture the shape of the observed real hazard functions than the Weibull and exponential models do. The hazard function rises at first then the speed of increase decreases slowly. However, both models have not captured the decrease in the hazard function after around 20 months as the semi-parametric PH Cox model does.

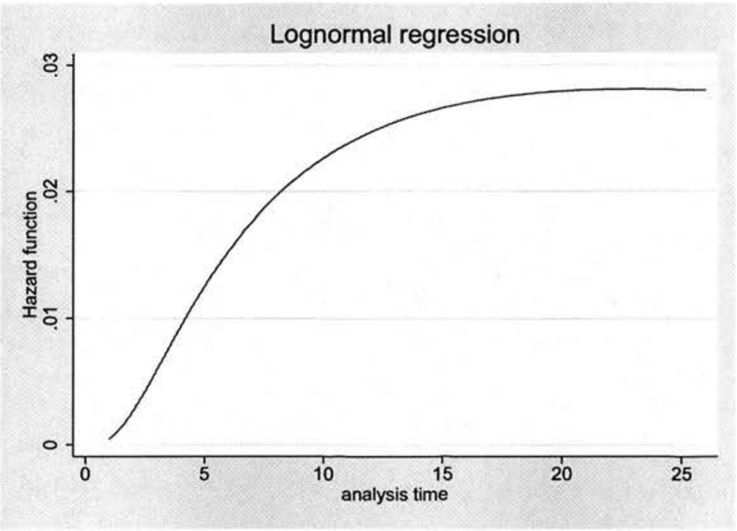

Figure 5.26: Hazard function, Lognormal model, paying back early data

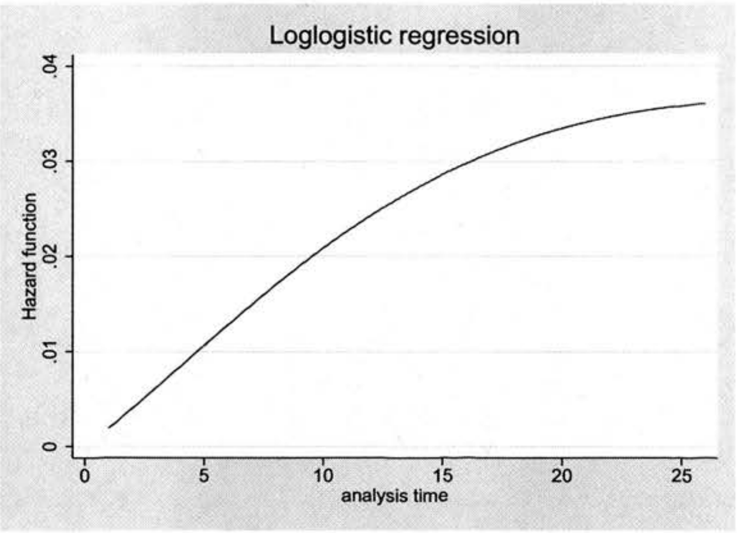

Figure 5.27: Hazard function, Loglogistic model, paying back early data

5.5.4 Model comparison

One way of comparing the goodness of fit of the survival models is by using Cox-Snell Residuals Cox and Snell (1968). For each observation j, the Cox-Snell residual r_j is

$$r_j = \widehat{H}_0(t_j)exp(\widehat{\beta}x_j)$$

where $\widehat{\beta}$ are the estimates of the survival model. $\widehat{H}_0(t_j)$ is the cumulative baseline hazard function up to t_j. This set of Cox-Snell residuals can be treated as observations from an exponential distribution with parameter λ equal to one if the β and $H_0(t)$ are the true estimates of the model parameters.

The fit of the model may be examined by comparing these Cox-Snell residuals to the empirical estimates of the cumulative hazard function. The Kaplan-Meier estimates of the survival function S, can be transformed into the empirical estimates of a cumulative hazard function where $H_0 = -ln(S)$. If the model has a good fit, the plot of Kaplan-Meier estimates against Cox-Snell residuals should be very close to a straight line with a slope of one.

The Figures 5.28, 5.29, 5.30 , 5.31 and 5.32 suggest that Weibull and exponential models have a poorer fit to the data compared to the Lognormal, Loglogistic and PH Cox model. The better model fit for the Lognormal and Loglogistic models can be explained by their suitable shapes of the hazard function forms, which see the hazard rates increase and then go down as observed in the paying back early behaviour. PH Cox model can also achieve a reasonable fit thanks to its semi-parametric estimation of the baseline hazard functions.

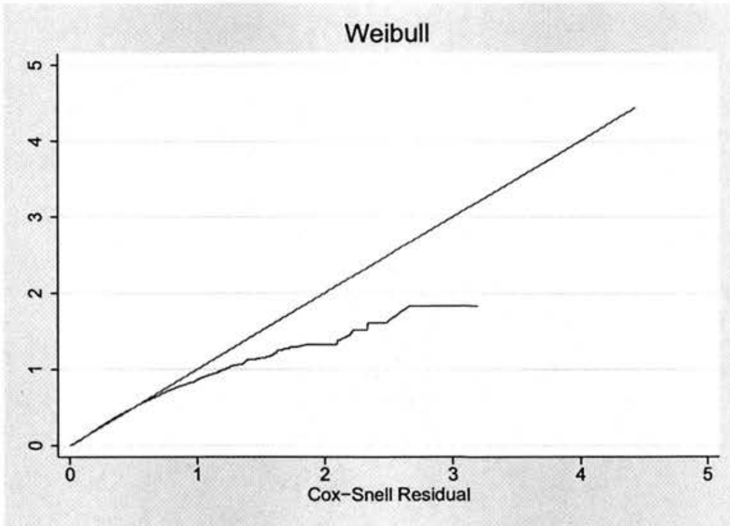

Figure 5.28: Using Cox-Snell residuals to check the fit of Weibull model

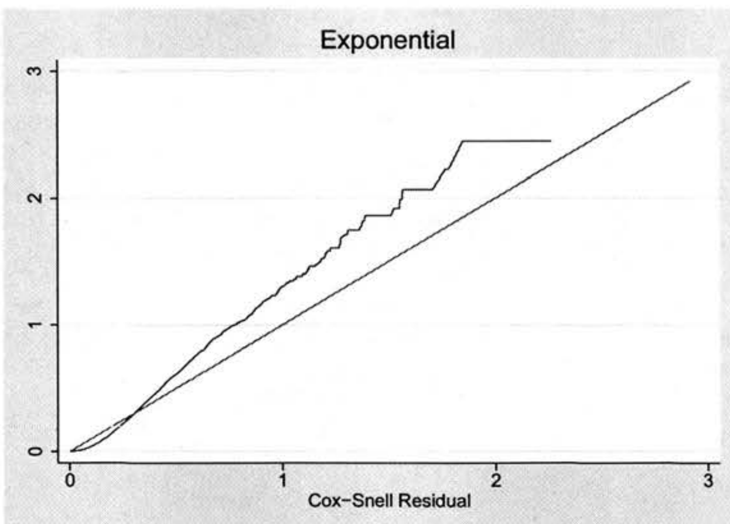

Figure 5.29: Using Cox-Snell residuals to check the fit of Exponential model

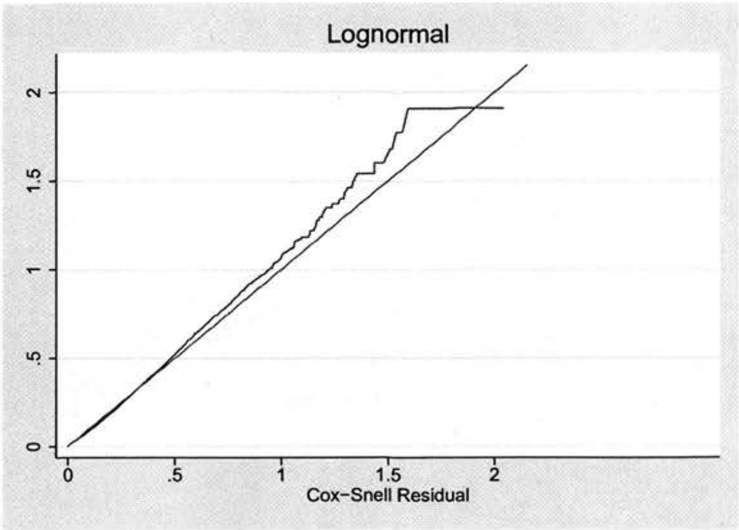

Figure 5.30: Using Cox-Snell residuals to check the fit of Lognormal model

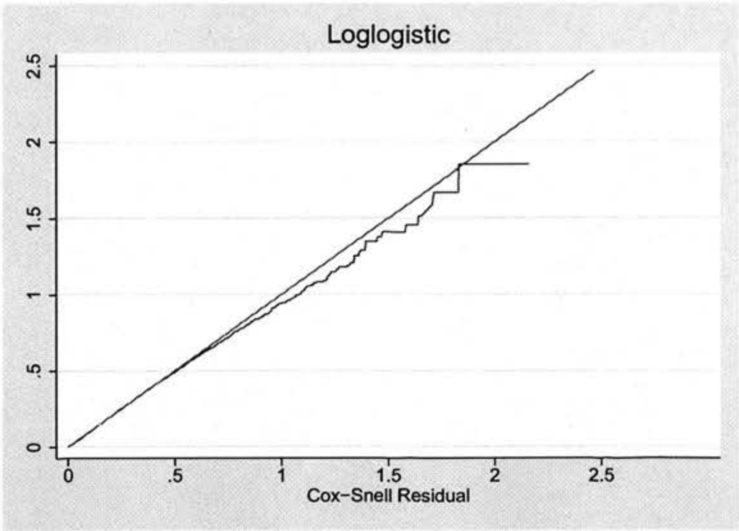

Figure 5.31: Using Cox-Snell residuals to check the fit of Loglogistic model

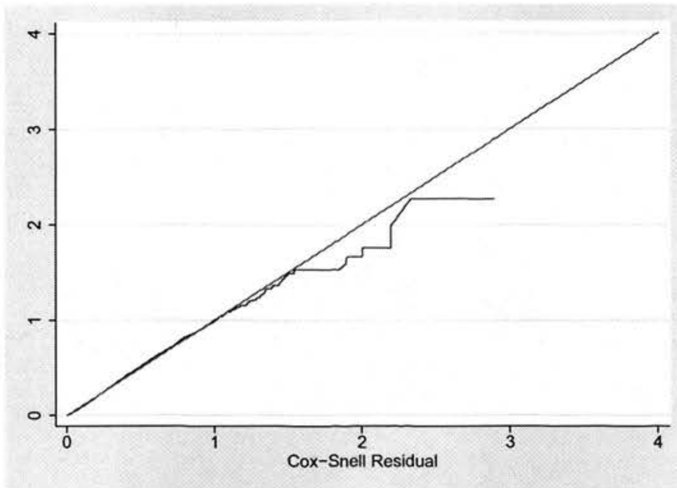

Figure 5.32: Using Cox-Snell residuals to check the fit of PH Cox model

Better fit of the model to the training data does not necessarily translate to a better predictor on the out-of-sample data. The predictive performances of those parametric and semi-parametric survival analysis models, measured by their abilities to differentiate the binary outcome of paying back early or not for the holdout set, are compared in Table 5.6 along with the performance of a Logistic Regression model as the benchmark.

Model	AuROC
Weibull	0.6600
Exponential	0.6700
Lognormal	0.6597
Loglogistic	0.6606
PH Cox	0.6641
Logistic Regression	0.6736

Table 5.6: Comparison of the model predictive performance on holdout set

As shown in Table 5.6, using AuROC as the performance measure, overall the predictive abilities of those models are very similar in terms of the ability of differentiating the non-paying back early from the paying back early customers within the 26 months observation period. The lowest AuROC value, 0.6597, was reported by a Lognormal model and the highest AuROC value, 0.6736, was reported by a Logistic Regression. The exponential model, despite its poor model fit indicated by the Cox-Snell residuals, achieves the second best AuROC.

5.5.5 Conclusion

Compared to the predictive performance of the default models, the paying back early models achieve much lower AuROC values. This is because of the lack of predictive variables explaining the paying back early behaviours. The dynamic competitive ranking data of the lenders' typical rates for each month and the rates charged by competing lenders , for example, might improve the predictions if they were available to be included in the models.

Due to the different shapes of the hazard functions observed, the parametric proportional hazard survival models like Weibull and exponential models are not found to fit the hazard functions well, despite still doing reasonable well in the binary outcome predictions. The parametric Accelerated Failure Time models like the Lognormal and Loglogisitic models, are found to fit the data better in terms of the Cox-Snell residuals. The proportional hazards Cox model, is also found to have good model fit (thanks to the semi-parametric model structure) and comparable predictive performance.

5.6 Results of competing risks modelling

In our data three different outcomes may happen, paying back early, continue to pay on time during the observation period (right censored) or default. The risks of paying back early and default are non-repeated failures by definition. Once the customer has paid back early or defaulted, for the analysis the account was considered closed and study time finished [2]. Each subject was either right censored or encountered one of the two events. As seen in Figure 5.3 and 5.4, the Kaplan-Meier survival and hazard functions for payback early and default are totally different.

Lunn and McNeil (1995) discussed different methods available for the estimation of parameters in modelling competing risks, either estimating the parameters for those events individually or jointly. The latter was preferred rather than the "separate estimation" approach, the drawback of which, they argued, is "it does not treat different types of failures jointly, complicating the comparison of parameter estimates corresponding to different failure types". They described two methods that model competing risks with parameters estimated jointly. One practical advantage of their models is that the two models do not require dedicated software packages. Instead, the models work by augmenting the data through duplication. For a model with two competing risks, the data will be doubled by duplicating rows. One row for one risk. One new binary covariate is introduced to indicate the risk type. The interactions with this new "risk type" covariate and other covariates x_i are also created to enter into the model.

The first of the two methods (called "Method A") proposed by Lunn and McNeil

[2] Many borrowers who defaulted on our definition (2 payments overdue), were actually allowed by the lender who supplied the data, to continue making payments.

assumed that the baseline hazard functions for each hazard function $h_{0j}(t)$ for each possible risk event differ by a constant ratio, e^{b_0}. When the assumption of a constant ratio between baseline hazard functions does not hold, Lunn and McNeil suggested an alternative method (called "Method B") that fits a stratified Cox PH model in which the data for each failure type forms a strata, assuming different baseline hazard functions and sharing the same set of parameters. Cleves (1999) suggested a simpler model (called "Method StataFAQ") for the analysis of multiple survival data. It is simpler because interaction covariates are dropped. The model is similar to Lunn and McNeil's "Method B" model in using the strata to model different baseline hazard functions for each type of risk. The tables of estimates are listed in Tables C.8, C.10 and C.9, which can be found in Appendix C.

The predictive performances for each event measured on the holdout set, in terms of area under ROC curves are reported in Table 5.7. All three models are trained using dummy variables since weights of evidence are outcome specific. The predictive variables are firstly stepwise selected from a PH Cox default model and a paying back early model, separately. The two sets of selected variables are then merged into the set of predictive variables used across the current three competing risk models. The weighted AuROC values are calculated as the sum of half of the AuROC values of defaults and half of those for payback early. Generally the paying back early was not predicted well compared to default in terms of AuROC values, as has been previously noted. The method StataFAQ is the worst performing model, with both default and paying back early having the lowest AuROC of the three models. Method A and Method B, on the other hand, are barely distinguishable.

AuROC	default	paying back early	weighted
Method A	0.8273	0.6691	0.7482
Method B	0.8287	0.6679	0.7483
Method StataFAQ	0.7543	0.6530	0.7037

Table 5.7: Comparison of the model performance

In conclusion, in the three competing risks models we have tested, the two models proposed by Lunn and McNeil provide reasonable predictive performance. Compared to the individually estimated approaches, PH Cox for example, the competing risks models are competitive in predicting paying back early. On the other hand, in the prediction of defaults, their predictive performance are lagging behind. Overall the benefits from using these competing risks models over individually estimated approaches are not significant.

5.7 Conclusion

In this chapter, different modelling approaches of default and paying back early have been tested. Overall, the semi-parametric proportional hazard Cox model is found to perform well in predicting both types of behaviours. In the parametric survival models, the exponential model is found to perform competitively in the prediction of both default and paying back early. However, the two types of proportional hazard parametric models, exponential and Weibull models, are found to fit the data less well than other models for the paying back early data, which has a bump shape in the hazard function. Finally, both types of the failure events have been modelled in a Competing Risk framework, which did not seem to bring much benefit in terms of the predictive performance compared with the approaches that model the events individually.

Chapter 6

Profitability Modelling

6.1 Introduction

This chapter calculates the unconditional expected profit for the lender at the time an application for credit is received and before making an offer to the customer. This chapter has the following structure.

- Section 6.2 gives the estimating equations for the profits of a fixed term loan product. Detailed results are discussed along with graphical presentations.
- Section 6.3 demonstrates the optimal decision policies the lender can employ to maximize profit or market share subject to the marketing strategies.
- Section 6.4 provides sensitivity tests on different segments. Specifically, we will analyse the difference between Internet and None Internet segments and compare the economic benefit of this segmentation.

The existing literature lacks an empirical methodology which a lender may use to choose the interest rate on a fixed term loan when its objective is to maximise unconditional expected profits at the time of application subject to a minimum market share. This chapter provides such a methodology and applies it to a dataset of actual choices made by applicants and a lender so that optimal decision policies can be applied subject to the constraints.

The results show the extent of the trade-off between market share and unconditional expected profits. Our results also demonstrate the possibility of segmenting the market and choosing the optimal interest rate for each loan amount requested can lead to markedly different policy decisions than by simply adapting a particular rate for all applicants. We also discuss how the specification of the models can affect the functions estimated and decision policies involved.

6.2 Estimating equations

The unconditional expected profit at the time of application ($t = c$) of a fixed term loan, but conditional on a vector of an applicant's characteristics, x, can be written as

$$E_{t=c}(\pi|x) = E_{t=c}(\pi|a|x)E_{t=c}(p(a)|x) + E_{t=c}(\pi|\overline{a}|x)(1 - E_{t=c}(p(a)|x)) \qquad (6.1)$$

where $a(\overline{a})$ = the potential borrower accepts (rejects) the offer and π = the present value of the profits at $t = c$. The second term is assumed to be zero. If the customer rejects the offer, the lender makes a profit of zero. The first term is the product of the expected profit conditional on acceptance (see section 6.2.1 below) and the acceptance probability (see section 6.2.2). We also assume that $E_{t=c}(\pi|a|x)$ and $E_{t=c}(p(a)|x)$ are independent. Although $E_{t=c}(\pi|a)$ is correlated with $p(a)$ since both are functions that share a same set of predictive variables, it does not necessarily imply that they are correlated with each other when conditional on this same set of applicant characteristics

x. It should also be noted that in later subsections, for the ease of notation, we use $p(a)$ in place for $E_{t=c}(p(a)|x)$.

6.2.1 Conditional expected profit

In the previous chapter on survival analysis we compared two different approaches to estimate survival probabilities of default and paying back early, independently, and simultaneously in competing risks models. The benefit gained from using competing risks models instead of independent estimation models was not significant when measuring the predictive performance of models by AuROC values on the holdout set. Therefore the basic assumption we have made in the beginning of this chapter is that the probability of default P_t^d and the probability of paying back early P_t^b are independent.

The conditional expected profit is the sum of four sources of expected revenue, each discounted at the opportunity cost of the funds, less the value of the loan. One rationale behind the discounting is that eventually we want to calculate the unconditional expected profit at the very time of the application before the offer is made by the lender to the applicant. The other reason is that, from the data supplied we cannot infer the exact figure for the cost of the lender supplying the loan. Using this discounting we assume the cost to the lender to supply the loan by borrowing the funds is a fixed interbank rate. [1]

The four sources of expected revenue are presented in detail as follows.

---

[1] This assumption of a fixed interbank rate can be wrong in a volatile market when the liquidity is under pressure.

1. The expected scheduled monthly payments when the borrower is still making these because he/she has not defaulted and not repaid early:

$$\sum_{t=1}^{T} S_t^b S_t^d \frac{M}{(1+i)^t}$$

where

S_t^b = probability that the borrower has not chosen to pay back early all of the loan on or before period t;

S_t^d = probability that the borrower has not defaulted on or before period t;

i = the interbank monthly interest rate, assumed to be 5% compound over 1 year; The opportunity cost of the fund cannot be omitted as we assume that, to service the loan, the lender need to borrow all or most of all (subject to regularities restrictions such as Basel I or II) the fund at an interbank rate.)

M = scheduled (fixed amount) monthly payment;

T = terminal period of the loan;

2. The expected balance to be repaid early provided the borrower has not defaulted before the early repayment date:

$$\sum_{t=1}^{T} S_t^d (S_{t-1}^b - S_t^b) \frac{B_t(1+2*j)}{(1+i)^t}$$

where

B_t = the expected balance to repay when the borrower wishes to settle early at the end of month t.

The fee for early repayment is assumed to be two months interest, at monthly interest rate j, on the rest of the balance.

3. The expected recovery amount if the borrower defaults in month t but has not

paid back early before t:

$$\sum_{t=1}^{T}(1-LGD)S_t^b(S_{t-1}^d - S_t^d)\frac{B_{t-1}}{(1+i)^t}$$

where LGD = Loss Given Default.

4. The expected receipts from insurance premia:

$$I * p(I) * p(U)$$

where

I = insurance income if the insurance is taken and no claim is made;

$p(I)$ = probability that the insurance is taken;

$p(U)$ = probability that no insurance claim is made.

The balance to pay when the customer wishes to settle early at the end of month t, B_t, is calculated according to the *Consumer Credit Early Settlement Regulations 2004* as:

$$B_t = L(1+j)^t - M\frac{(1+j)^t - 1}{j}$$

where

L = the loan amount requested

j = the monthly rate. Annual Percentage Rate (APR) is equal to $100((1+j)^{12}-1)$. So $j = e^{(log(1+APR/100))/12} - 1$.

The monthly payment, M, is

$$M = L * j * \frac{(1+j)^T}{(1+j)^T - 1}$$

where T is the term in the number of months.

The conditional insurance income, I, is calculated as an added margin as a percentage of the sum of expected monthly payments

$$I = \sum_{t=1}^{T} S_t^b S_t^d \frac{M}{(1+i)^t} IM$$

where IM = insurance margin.

We have assumed or estimated the following parameters listed below. When deciding what exact values to be chosen for parameters roughly assumed, we choose the values in a conservative way in their effects upon the expected profits.

- $LGD = 0.75$ This is an averaged Loss Given Default estimates assumed to be constant.
- $IM = 0.1$ The insurance margin indicates what percentage extra the lender will add to the monthly payments. The exact percentage the lender will charge is different, within range from 10% to 20%, depending on the decision policies constraints involved. A lower end value of 10% is chosen as it is conservative towards profit estimation.
- $p(U) = 0.8$ This is a conservative estimate by assuming only 20% of customers will claim. In fact, according to the 2006 report from The Office of Fair Trading about the UK Payment Protection Insurance (PPI) market titled 'The PPI Claim Ratio, percentage of premiums paid by consumers', the claim ratio was estimated to be as low as 15-20%. [2]
- $p(I) = 0.3$ The insurance take-up rate is the average value across all the sample instead of modelled.

[2] The web address for the market study report of the UK market of payment protection insurance (PPI) at 2006 can be found at http://www.oft.gov.uk/news/press/2006/148-06

- $i = 0.004074$ This monthly interest rate is equivalent to an annual rate of 5%. The averaged interbank rates on the market during the time period in which this fixed term loan product resides is below 5%. As the interbank rate constitutes the most important part of costs for the lender to fund the loan, an slightly conservative value of 5% is chosen to compensate possible increase of the rate and the administrative overheads in the fixed costs.
- $T = 24$ The time span of 24 months is chosen because the equations involved will be the simplest functional form to calculate in the Matlab Symbolic Toolbox.

Although 24 months is the choice of the loan term that we want to calculate the expected profits for, the samples on which the estimation routines were carried out include the samples with loan terms from 24 months up to 60 months. Using this larger pool of samples (totally 27160 cases including both training and holdout sets) avoids the possible bias introduced because of the very small sample size for the 24 months loan (totally 2149 cases including both training and holdout sets).

The survival probabilities for the Default and PayEarly were calculated using estimates from the Cox Proportional Hazard models [3].

$$\begin{aligned}\widehat{S_t^b}(t,L,APR) &= \widehat{S_0^b}(t)^{exp(\widehat{\beta_{APR}^b}*APR+\widehat{\beta_L^b}*L+\widehat{\beta_1^b}*log(L)*log(APR)+\widehat{\beta_0^b})} \\ \widehat{S_t^d}(t,L,APR) &= \widehat{S_0^d}(t)^{exp(\widehat{\beta_{APR}^d}*APR+\widehat{\beta_L^d}*L+\widehat{\beta_1^d}*log(L)*log(APR)+\widehat{\beta_0^d})}\end{aligned}$$

where $\widehat{S_0}(t)$ are the baseline survival functions which do not change with the predictive variables such as APR and Loan Amount L.

[3] This formulation includes an interaction between the loan amount L and APR. The tables of estimates for the Default and PayEarly hazard functions can be found in the appendix in Table D.13 and D.16

$\widehat{\beta_0^b}$ and $\widehat{\beta_0^d}$ are the estimated constants within the survival probability functions for paying back early and default, plus the added sum of the other variables' mean values multiplied by their corresponding coefficients estimated from Cox models. That is, if we calculate for the applicant with the mean values of the covariates:

$$\widehat{\beta_0^b} = \widehat{\beta_{b0}} + \sum \widehat{\beta_{bX}\overline{X}}$$
$$\widehat{\beta_0^d} = \widehat{\beta_{d0}} + \sum \widehat{\beta_{dX}\overline{X}}$$

For convenience we will refer to such an applicant as the 'mean' applicant or 'typical' applicant. The results for the remainder of this section and section 6.4 relate to the 'mean' applicant.

6.2.1.1 Results of conditional expected profit

The expected profits conditional on an offer having been accepted by customers, $E_{t=c}(\pi|a|x)$, have been calculated by summing up the four sources of revenues detailed in the previous subsection. Figure 6.1 and Figure 6.2 plot the results in a profit-loan amount-interest rate 3D space .

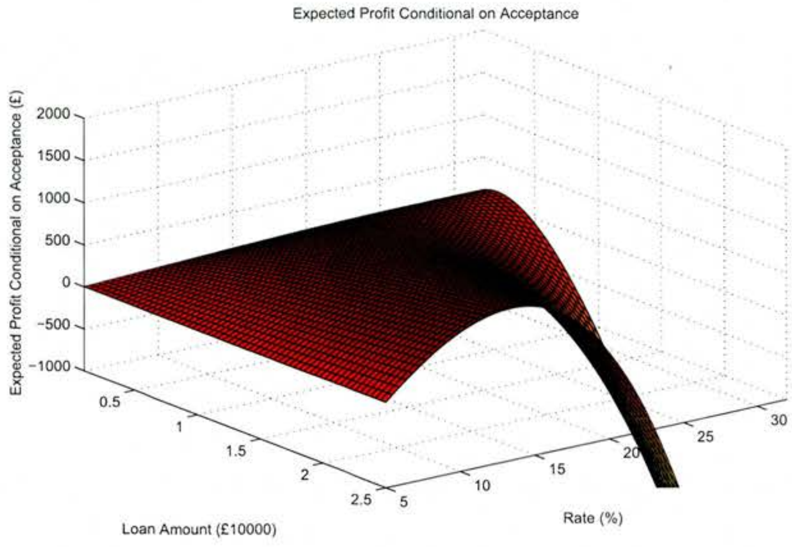

Figure 6.1: Expected profit conditional on the acceptance

In Figure 6.1, the X axis is Loan Amount, the Y axis is the interest rate APR charged and the Z axis is the expected profits conditional on the loan having been accepted by a customer. Figure 6.2 is the projection of Figure 6.1 into a 2 dimensional Loan Amount vs. Rate space. This is done by connecting points of different offers (Loan Amount and Rate) but with the same expected profits to give iso-expected-profit contours. The colours of the contours indicate the height of the conditional expected profits. Hot colours (red or yellow) stand for relatively higher profits while the cold colour (blue) stands for lower conditional expected profits.

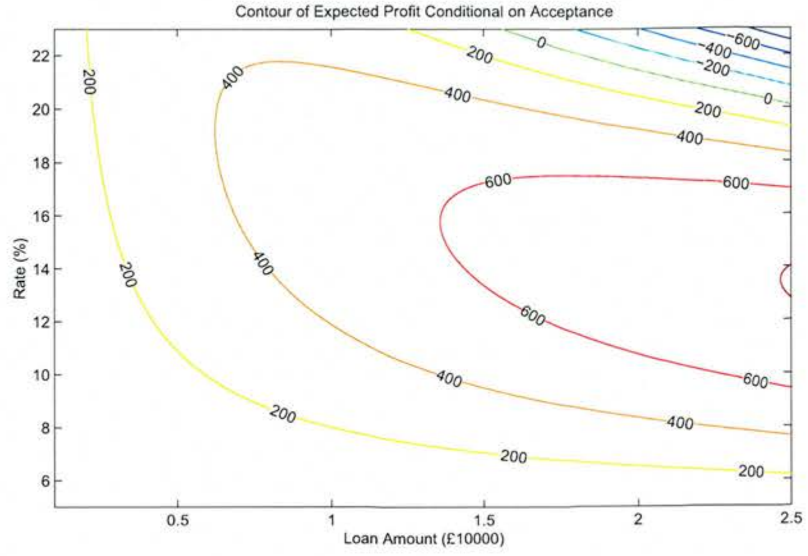

Figure 6.2: The contour of conditional expected profit shown in Figure 6.1 . [a]

[a]The colour of the contour indicates the height of the conditional expected profit. Hot colours (red or yellow) stand for relatively higher profit while cold colour (blue) stands for lower expected profit.

The figures show that up until an interest rate of approximately 14% the conditional expected profits increase along with the loan amount given an interest rate. The larger the loan amount, the larger the conditional expected profits. They also show that given a loan amount, a higher interest rate increases conditional expected profit up until a threshold point of the interest rate. Interest rates above that threshold actually reduce the conditional expected profit.

For different given loan amounts, this threshold interest rate above which the conditional expected profit is reduced rather than increased, is slightly different. As can be seen more clearly in Figure 6.2, the threshold interest rate is lower for larger loan amounts than for smaller loan amounts. For a £25000 Loan, the threshold interest rate will be around 14% while that rate will be slightly higher at around 19% for a loan of

£6000.

6.2.2 Acceptance probability

Logistic Regression was used to estimate the probability of acceptance on the samples with loan terms ranging from 24 months up to 60 months [4]. The equation that was estimated had the form of $log\frac{p}{1-p} = \beta X$. Therefore the estimated probability of acceptance $\widehat{p}(a)$ can be expressed as:

$$\widehat{p}(a) = \frac{e^{\widehat{\beta}X}}{1+e^{\widehat{\beta}X}} \tag{6.2}$$

Table D.1 in the appendix shows the estimated parameters. In the table, raw_loanapr1 and L are the Loan APR and Loan Amount in their original continuous values together with logLXAPR, the interaction term between the two variables APR and L that we found to significantly improve the acceptance model when it was included. This variable is coded as the product of their natural logarithms, log(L)*log(APR). All the other variables are coded using dummy variables.

Similar to the treatments we have used previously in the PH Cox models , $\widehat{\beta_0}$ is the constant plus the added sum of other variables' mean values multiplied by their correspondingly estimated coefficients: $\beta_0 + \beta X$.

$$\widehat{p}(a) = \frac{exp(\widehat{\beta_0} + \widehat{\beta_{APR}} * APR + \widehat{\beta_L} * L + \widehat{\beta_1} * log(L) * log(APR))}{(1 + exp(\widehat{\beta_0} + \widehat{\beta_{APR}} * APR + \widehat{\beta_L} * L + \widehat{\beta_1} * log(L) * log(APR))}$$

For an applicant with the mean values of the covariates, the 'mean' applicant: $\widehat{\beta_0} =$ 4.0175. The estimates for $\widehat{\beta_{APR}}$, $\widehat{\beta_L}$ and $\widehat{\beta_1}$ are -0.1098 , 0.1076 and -0.9015 respectively. Both negative signs on $\widehat{\beta_{APR}}$ and $\widehat{\beta_1}$ indicate that the applicant will be much

[4] The reason for this is that there are a limited number of 24 month cases in the data, too small to provide unbiased estimates. This choice of samples was discussed in section 6.2.1

less likely to accept an offer if the interest rate APR is high. The positive sign on $\widehat{\beta_L}$ combined with the negative sign on $\widehat{\beta_1}$ indicates that a more complex relationship exists between the loan amount and the acceptance probability. For a given interest rate APR, the acceptance probability is dependent on the relative sizes of two terms $\widehat{\beta_L} * L$ and $\widehat{\beta_1} * log(L) * log(APR)$. Plug in the estimates of $\widehat{\beta_L}$ and $\widehat{\beta_1}$

$$p(a) \propto 0.1098 * L - 0.9015 * log(L) * log(APR)$$

This means, when the Loan Amount is large, $p(a)$ will be dominated by the first term since the numerical values of $log(L) * log(APR)$ will be relatively much smaller. This leads to the conclusion that the acceptance probability will be higher when the loan amount is larger (though the maximum loan amount is £25K, meaning the largest $L = 25$). On the other hand, when the Loan Amount is small, the second term will be dominant and the combination of the two terms will bear negative sign. Under such circumstances, the applicant will be more likely to accept an offer when the loan amount is smaller.

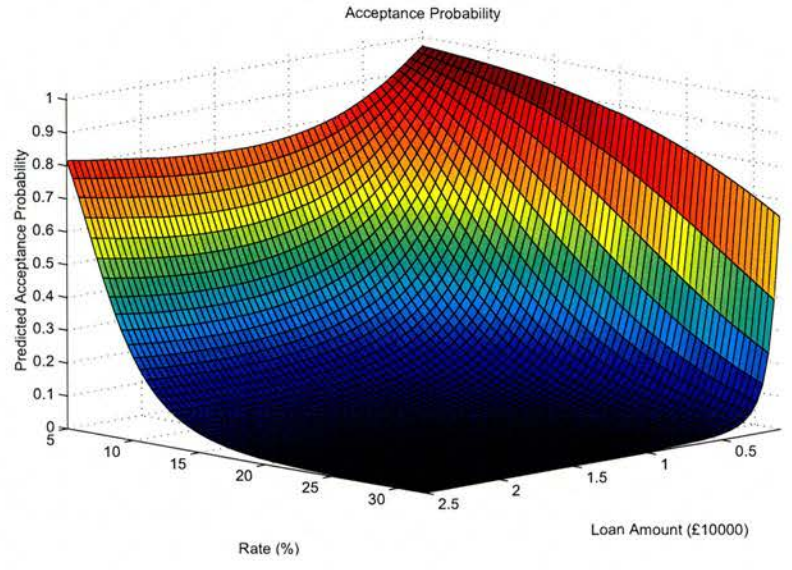

Figure 6.3: Estimated acceptance probability [a]

[a] The colour indicates the height of the numerical values of the estimated probability of acceptance. Hot colours (red or yellow) stand for relatively higher probability while cold colour (blue) stands for lower probability.

Figure 6.3 plots the results. Notice that the origin is in the *far corner*: lower interest rates and loan amounts occur further along the X and Y axis respectively. From the graph we can see that customers prefer a lower rate given a chosen loan amount and a lower loan amount given the interest rate, in accordance to the negative signs on the coefficients $\widehat{\beta_{APR}}$ and $\widehat{\beta_L}$. For example for a given loan amount of £25000, when the interest rate is increased from 5% to 10%, the probability of acceptance is decreased from 0.8 to around 0.3. While at an interest rate of 15%, when the loan amount is increased from £1000 to £5000, the probability of acceptance is decreased from around 0.9 to 0.3. The surface of the plot takes a S shape where the steepest part is located at smaller loan amounts and high interest rates. This is because the functional form of the probability of acceptance is a Logit function and the interaction term log(L)*log(APR) is included.

6.2.3 Unconditional expected profit

The unconditional expected profit at the time of application ($t = c$) of a fixed term loan was given in equation 6.1 and for convenience is reproduced here as $E_{t=c}(\pi|x) = E_{t=c}(\pi|a|x)E_{t=c}(p(a)|x)$. That is the product of the expected profit conditional on acceptance and the probability of acceptance. The expected profit conditional on acceptance $E_{t=c}(\pi|a|x)$ has been discussed in section 6.2.1. The probability of acceptance $p(a)$ has been discussed in section 6.2.2. Combining these two leads to the equation below

$$
\begin{aligned}
E_{t=c}(\pi|x) &= E_{t=c}(\pi|a|x)E_{t=c}(p(a)|x) \\
&= [\sum_{t=1}^{T} S_t^b S_t^d \frac{M}{(1+i)^t} + \sum_{t=1}^{T} S_t^d (S_{t-1}^b - S_t^b) \frac{B_t(1+2*j)}{(1+i)^t} \\
&+ \sum_{t=1}^{T} (1-LGD) S_t^b (S_{t-1}^d - S_t^d) \frac{B_{t-1}}{(1+i)^t} + \sum_{t=1}^{T} S_t^b S_t^d \frac{M}{(1+i)^t} IM * p(I) * p(U)] \\
&* \frac{exp(\widehat{\beta_0} + \widehat{\beta_{APR}} * APR + \widehat{\beta_L} * L + \widehat{\beta_1} * log(L) * log(APR))}{1 + exp(\widehat{\beta_0} + \widehat{\beta_{APR}} * APR + \widehat{\beta_L} * L + \widehat{\beta_1} * log(L) * log(APR))}
\end{aligned}
$$

The final Matlab Symbolic representation of the function of the unconditional expected profit, with all the estimated values, can be found in the appendix.

Figure 6.4 plots the results in 3D space, showing that a higher rate and loan amount do not necessarily bring higher expected profits from all applicants considering the very low acceptance rate at those points. In fact, a ridge of most profitable offers is shown, especially for those with higher loan amounts. At each given loan amount, the interest rate that maximizes the profit actually goes down when the amount is increasing.

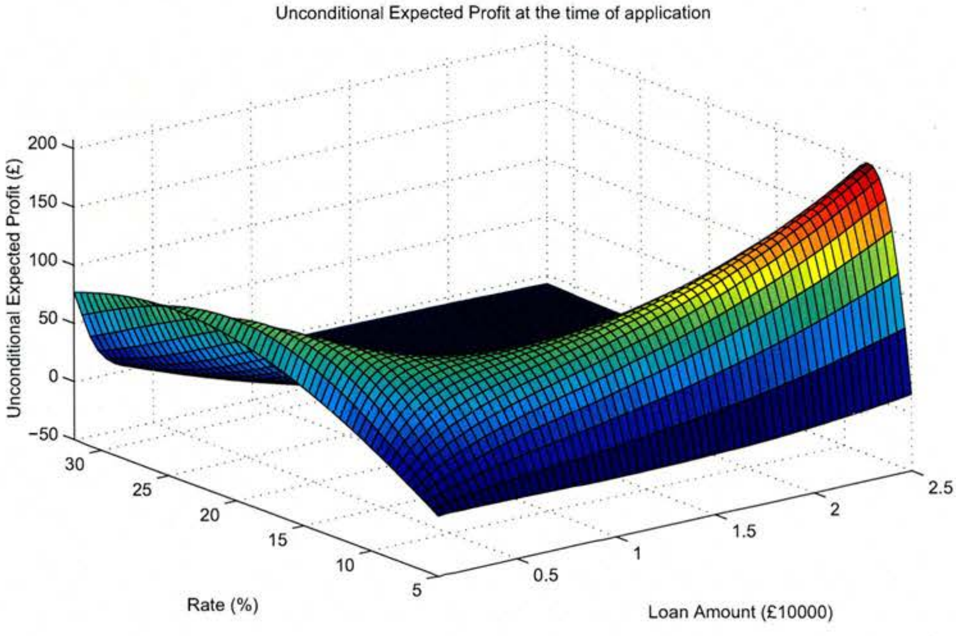

Figure 6.4: The unconditional expected profit of a loan at the time of application. [a]

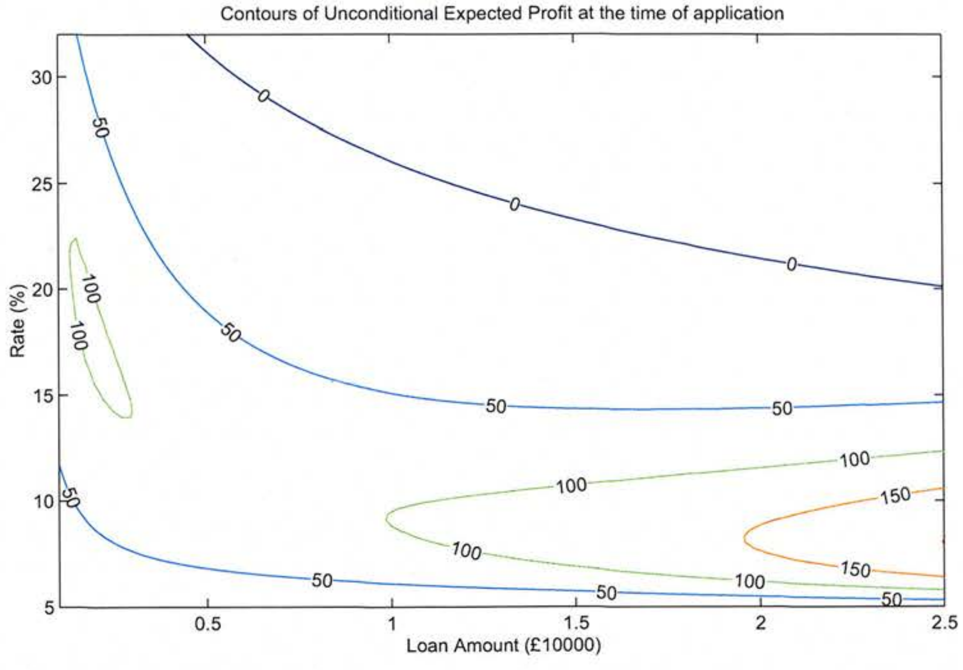

Figure 6.5: Contour of unconditional expected profit.

[a] The colour indicates the height of the expected profit. Hot colours (red or yellow) stand for relatively higher profit while cold colour (blue) stands for lower expected profit.

Figure 6.5 can be seen as the projection of Figure 6.4 into a 2 dimensional space of Loan Amount and Rate, with lines connecting loan amount - rate offers with the same expected profits to form contours. The same conclusion to that drawn from Figure 6.4 can also be drawn from this figure that a lower or mid-ranged interest rate is the most profitable for the lender. We can also see that for higher loan amounts the maximum expected profits are much less sensitive to the loan amounts than to the interest rates.

6.2.4 Iso-profit curves plotted together with iso-preference curves

We can further plot the iso-preference curves derived from the acceptance probabilities shown in Figure 6.3 and the iso-unconditional expected profit contours from Figure 6.5 on the same Figure, as shown in Figure 6.6. The iso-preference curves are stretching from top left corner to bottom right corner, marked with the acceptance probabilities (from 0.1 to 0.9). The iso-preference curves corresponding to a lower rate given a loan amount are those with higher acceptance probabilities.

If the lender wants to maintain market share by keeping the acceptance rates fixed while maximising profits, in other words, the lender wants to improve the profit without decreasing the probability of acceptance, from 0.6 as an example, then better offers can be made by moving along the iso-preference curve where $p(a) = 0.6$ from left to right to the region where the expected unconditional profit is higher. This assumes the lender can choose both rate the loan amount. The point of unconstrained maximum expected profit from an applicant can be found at the intersection at the point of the global maxima of profit (the top of the hill). In the next section we discuss the optimal decision policies under different constraints.

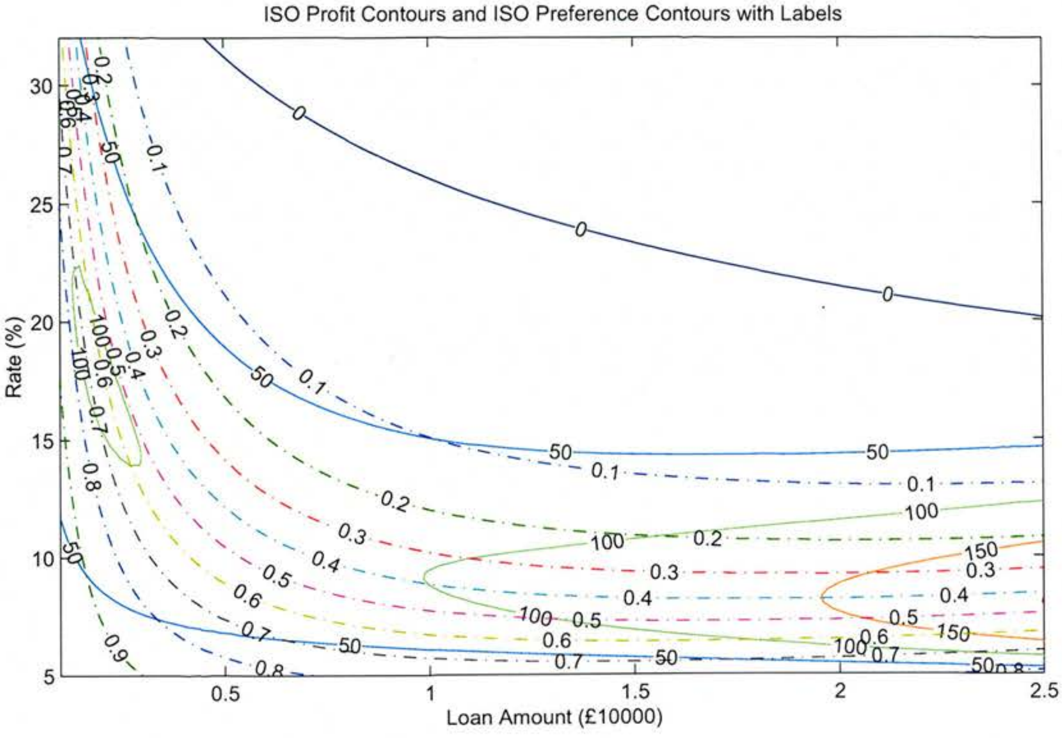

Figure 6.6: The iso-profit curves plotted together with the iso-preference curves.

6.3 Optimal decision policies

After a customer has applied for a loan and passed the credit check to be accepted by the lender, it is up to the lender to decide what as the characteristics of any offer it wishes to make. As shown in the flow chart below, Figure 6.7, the decision policies are dependent on its marketing strategies. A lender may want to maximize profit only, or increase market share only, or maximize profit subject to a certain minimum market share or possibly other combinations.

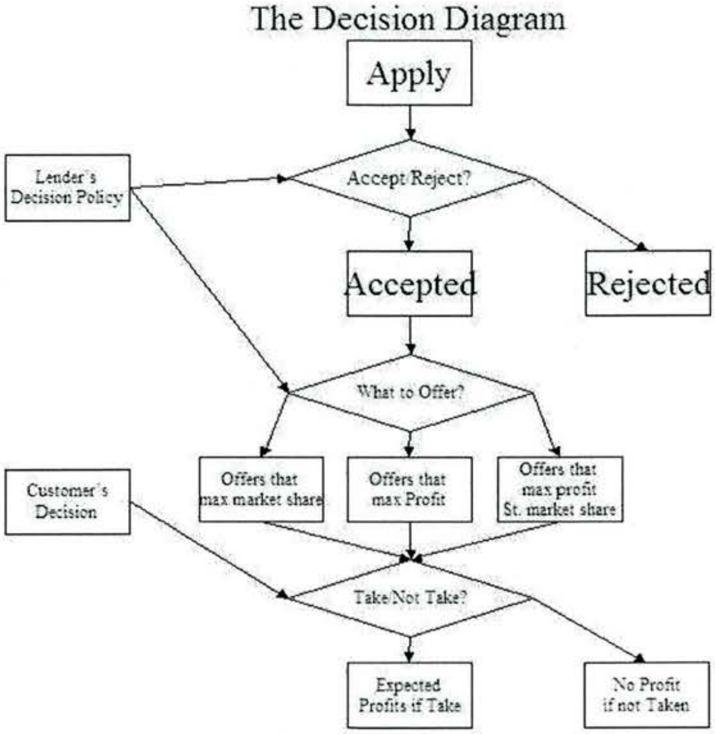

Figure 6.7: Decision diagram

Furthermore, the optimal decision policies to be employed by the lender are also dependent on certain constraints on aspects of the loan the lender can control. If the loan amount and APR are both variables adjustable by the lender, the optimal offers are no doubt to be found on the line of optimal offers suggested in Keeney and Oliver (2005).

When the expected profit is the only maximizing criteria (no market share concerns), the point at the top of the hill of the profits is the ideal choice. For example in Figure 6.8, the point marked with 'Peak' where Loan Amount is £25000 and Rate is 8.08% is such a maximum, assuming that £25000 is the largest loan amount that this loan product allows.

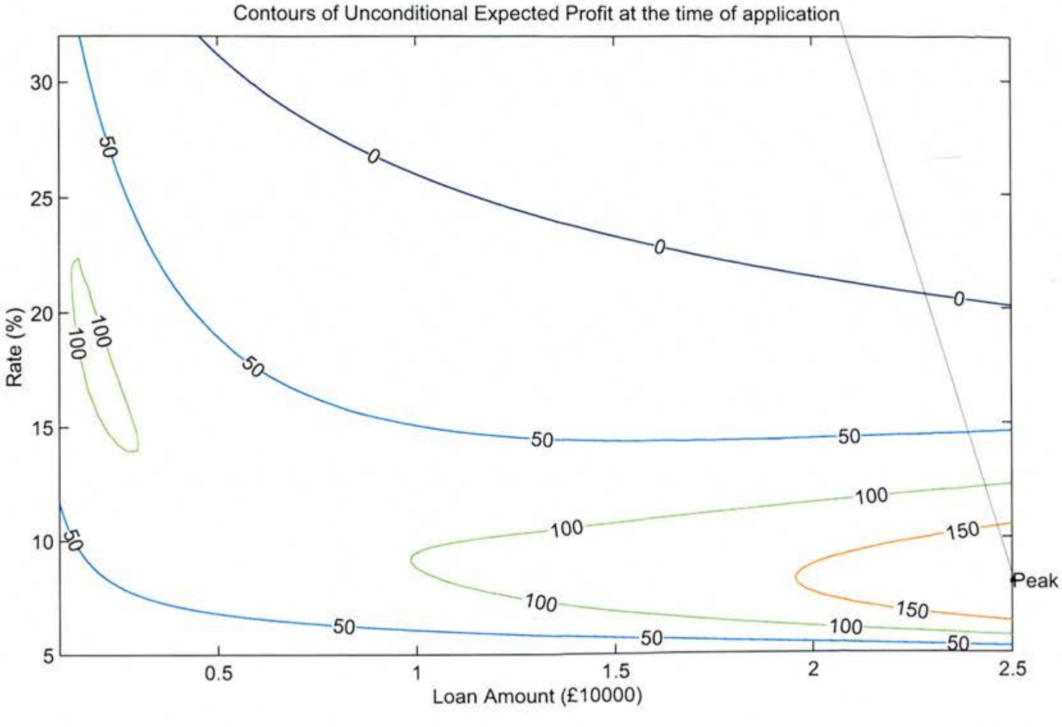

Figure 6.8: Contour of unconditional expected profit with peak point marked

When the lender wants to optimise the offers (by choices of interest rate) to maximise profit subject to a given market share ($p(a) \geq k$), the optimizing rate can be found as shown in section 6.3.1.

Table 6.1 lists the combinations of objectives to maximise given certain constraints. In reality, the amount of the loan is requested by the customer and rarely changed, except in order to allow the customer to pass affordability checks, or to offer a suggestion to the customer that they may want a bigger loan to cover an additional debt disclosed during a conversation. Since the loan amount is usually fixed by the applicant, the optimal policy is a choice of interest rate and is dependent on whether the market share is set as the constraint. The details are in subsection 6.3.2.

$p(a)$ unconstrained		
L fixed by borrower / r fixed by borrower	Yes	No
Yes	.	$\max_L E_{t=c}(\pi)$ st. $r = r^*$
No	$\max_r E_{t=c}(\pi)$ st. $L = L^*$	$\max_{L,r} E_{t=c}(\pi)$
$p(a) \geq k$		
L fixed by borrower / r fixed by borrower	Yes	No
Yes	.	$\max_L E_{t=c}(\pi)$ st. $p(a) \geq k$, $r = r^*$
No	$\max_r E_{t=c}(\pi)$ st. $p(a) \geq k, L = L^*$	$\max_{L,r} E_{t=c}(\pi)$ st. $p(a) \geq k$

Table 6.1: Matrix of optimal decision policies by lender

6.3.1 Optimal policies if choice is of APR rate s.t. p(a)[5]

Suppose the lender's objective is to maximize unconditional expected profit from an applicant by choice of interest rate, subject to a given minimum probability of acceptance $p(a) \geq k = 0.6$ and the loan amount chosen by the borrower will be adjusted accordingly. The optimal interest rate can be found by walking along the acceptance line $p(a) = k = 0.6$ until $L = 25000$ and $APR = 6.79\%$ with maximum of profit expected at 173.4, as seen in Figures 6.9 and 6.10.

[5] Please note that the discussion here is based on the results from the model that includes the interaction term between the loan amount L and APR in the predictive variables

Figure 6.9 shows the iso-profit contours with an iso-preference line with the probability of acceptance $p(a) = 0.6$. Figure 6.10 plots the unconditional expected profit against the loan amount when moving along the iso-preference line $p(a) = 0.6$ (The dotted curve in Figure 6.9 runs from from the top left corner to the right bottom corner with $p(a) = 0.6$ tag on). The unconditional expected profit increases then decreases and then increases again until reaching the point of maximum profit, which is found at the highest loan amount.

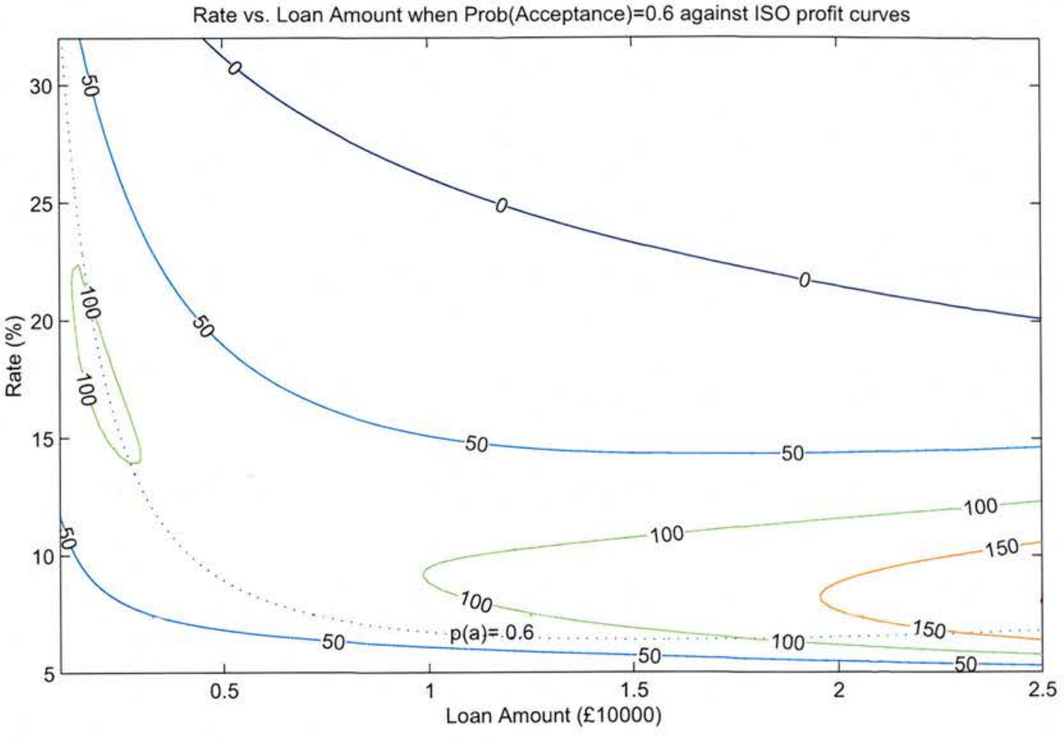

Figure 6.9: The acceptance line $p(a) = 0.6$ on the iso-profit contours.

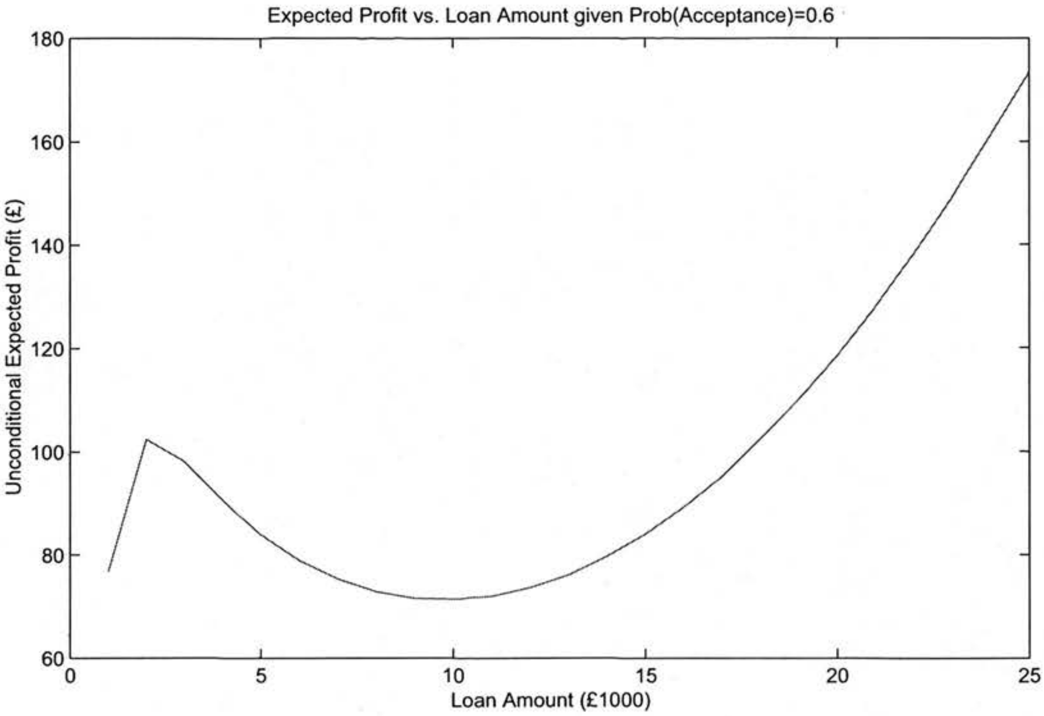

Figure 6.10: The loan amount that maximise the expected profit when $p(a) \geq 0.6$

The general solution to this optimisation problem can be set up as follows:

$$\begin{aligned} \max_{r} \quad E_{t=c}(\pi) &= f(L, APR) \\ s.t. \quad p(a) &\geq k \end{aligned}$$

where k is the minimum market share constraint. The Kuhn-Tucker condition shall be met and maxima can be found if f and $p(a) - k$ are concave. By observing the diagrams, it seems that f is not concave unless the decision region is split into two regions as there are two hills. If sufficient conditions are met, the optimisation problem can be written as follows:

$$\begin{aligned} \max_{r} \quad E_{t=c}(\pi) &= f(L, APR) \\ s.t. \quad k - p(a) &= g(L, APR) \leq 0 \end{aligned}$$

Letting the Lagrangian function $F = f(L,APR) - \lambda * g(L,APR)$ by introducing a multiplier λ and solving the Karush-Kuhn-Tucker conditions below shall yield the optimal solution.

$$
\begin{aligned}
\frac{\partial F}{\partial L} &= \tfrac{\partial f}{\partial L} - \lambda \tfrac{\partial g}{\partial L} = 0 \\
\frac{\partial F}{\partial APR} &= \tfrac{\partial f}{\partial APR} - \lambda \tfrac{\partial g}{\partial APR} = 0 \\
g &\leq 0 \\
\lambda * g &= 0 \\
\lambda &\geq 0
\end{aligned}
$$

Although all the functions here are differentiable, the functional form of the $E_{t=c}(\pi) = f(L,APR)$ is so complex (see Appendix) that differentiation procedures in the Matlab Symbolic Toolbox cannot be completed. This maybe due to memory overflow. Therefore, diagrams and small-stepped enumeration have been used to find the interest rate that maximizes the profit from an applicant.

6.3.2 Optimal policies if choice is of APR given loan amount

In the following cases the loan amount is chosen by the applicant and is fixed for the lender and is known by the lender before the lender chooses the interest rate.

6.3.2.1 If the market share p(a) is not considered[6]

If the market share is of no concern to the lender, we can just move across the iso-profit contours along a vertical line corresponding to the fixed loan amount and calculate the optimal rate. These rates are shown in Figure 6.11 for three different loan amounts.

[6] Please note that the discussion here is based on the results from the model includes the interaction term between loan amount L and APR in the predictive variables

Each of these lines represents a cross section through the iso-profit contours with different given loan amounts. Figure 6.11 shows that generally, for a given larger loan amount the interest rate that will maximize the profit is in fact lower.

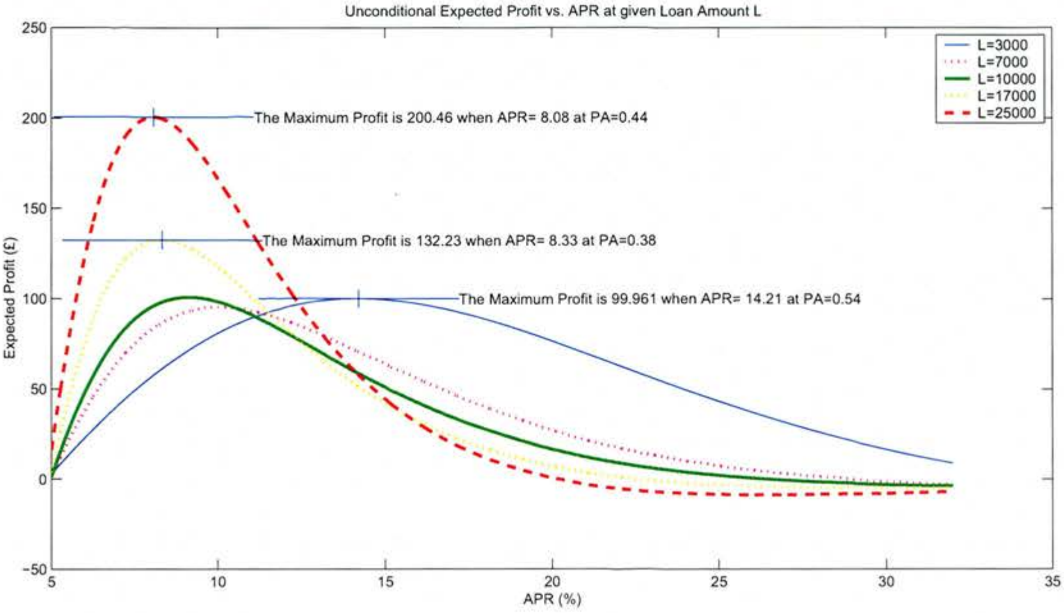

Figure 6.11: The optimal rate APR given Loan Amount L if p(a) is not considered.

6.3.2.2 If the market share p(a) is the constraint[7]

This situation is much more complex and needs discussion. Notice that if the loan amount is fixed by the borrower, the lender can choose whether to gain market share or profit subject to this constraint, but generally not both. Consider Figure 6.12 as an example, also assume the Loan Amount given is £10000:

1. If the minimum market share is $p(a) = 0.6$ then point A gives the optimal rate. Point A is at the intersection between the dotted iso-preference curve $p(a) = 0.6$

[7]please note that the discussion here is based on the results from the model includes the interaction term between loan amount L and APR in the predictive variables

and the vertical line of fixed loan amount L = 10000.

2. If maximising profit is regarded by the lender as more important than the market share then point B, where the highest profit given the loan amount is found, is the optimal choice. Point B is also at the intersection between the dotted iso-preference curve $p(a) = 0.38$ and the vertical line of fixed loan amount L = 10K. Notice that point A gives lower profit than B while gaining a higher market share.

3. Only if the minimum market share target implies a $p(a)$ that is lower than the $p(a)$ at the point B that will maximise the profit at the given loan amount, then the optimal choice is the point B, where both constraints on the market share and profit maximising target can be satisfied.

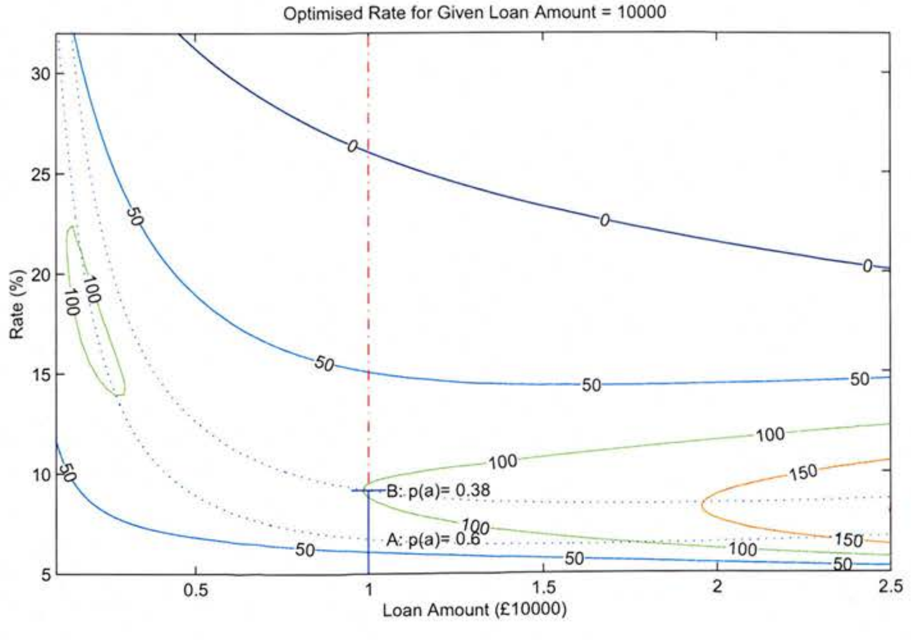

Figure 6.12: The optimal rate APR given Loan Amount L if p(a) is the constraint.

6.4 Segmentation

The results so far assume that one model for acceptance probability and one for each of the survival probabilities applies to all the applicants. To explore the sensitivity of this assumption, we separated the sample of applicants into different segments and observed the differences across the segments.

We wish to discover whether, if we choose an individual profit maximizing interest rate for the mean applicant from each segment separately, the unconditional expected profits are larger than the profits expected from the previous model without segmentation. Similarly, will the expected market share be larger after the segmentation?

6.4.1 Sensitivity test of the segmentation based on application channel

The equation used to calculate the unconditional expected profit from an applicant has to assume that the length of the loan term is 24 months. This is due to the limit imposed by the ability of the Matlab Symbolic Toolbox to handle large symbolic calculations. However, only a relatively few cases have been observed in our data with 24 month loan term. To achieve robust results with segmentation, it was decided to include more sample data with loan terms up to 60 months. In all the calculations and estimations hereafter, the variable of loan term was entered into the models. The coefficients accordingly have been used in the calculations to get the estimated probabilities given loan term equal to 24 months. The unconditional expected profits calculated are also based on a 24 month loan.

Based on this set of sample data with loan terms up to 60 months, applicants have been segmented into two sub samples depending on whether they applied through the Internet or they applied in some other way. Around 40 percent of customers applied through the Internet (18883 cases) and the rest did not (27199 cases).

Different methods are available to select the variables for the estimation of the acceptance model, the default model and the paying back early model when we consider segmented samples. Another factor to consider is the inclusion of interaction term between Loan Amount and Rate. Four methods have been considered and are discussed below.

Method 1 chose the sets of variables to enter the three probability functions after running stepwise procedures on the *combined* data set. Using the variables selected (with the Loan Amount variable forced in if necessary), the parameters for those three probabilities were then estimated separately for each segment as well as for the combined data. We call the resulting model, Model 1. In Model 1 the interaction term between Loan Amount and Rate was found to be statistically significant in predicting the probability of acceptance and survival probability of default but not so in predicting the survival function of early repayment.

Method 2 chose the sets of variables for each probability function *individually* using a separate stepwise procedure for each segment (with the Loan Amount variable forced in if necessary). The parameters were estimated using these sets of variables separately. The interaction term was excluded from the variable selection. We call the result Model 2.

Method 3 chose the sets of variables for each probability function *individually* using a separate stepwise procedure for each segment (with the Loan Amount variable forced in if necessary). The parameters were estimated using these sets of variables separately. The interaction term was included from the variable selection. We call the result Model 3.

Method 4 chose the sets of variables to enter the three probability functions after running stepwise procedures on the *combined* data set. Using the variables selected (with the Loan Amount variable forced in if necessary), the parameters for those three probabilities were then estimated separately for each segment as well as for the combined data. The interaction term was excluded from the variable selection. We call the resulting model, Model 4.

The estimation results for all three functions for both and each segments in Model 1, 2, 3 and 4 can be found in the appendix section. Panel a in Table 6.2 summarizes the difference between these 4 models concerning the inclusion of the interaction term and the choice of whether or not to conduct stepwise selection for each segment separately. Panel b in Table 6.2 compares the predictive performance between Model 1, 2, 3 and 4 across the Acceptance, Survival of Default and the Early Repayment models. The predictive performance is measured by the area under the ROC curve on the independently selected holdout data.

Panel a:			
Interaction term included		Yes	No
Stepwise for each segment separately	Yes	Model 3	Model 2
	No	Model 1	Model 4

Panel b:						
	Yes			No		
Yes	Model 3			Model 2		
	NonSegment	Acceptance	0.7874	NonSegment	Acceptance	0.7707
	Internet	Acceptance	0.8001	Internet	Acceptance	0.7848
	NonInternet	Acceptance	0.7799	NonInternet	Acceptance	0.7687
	NonSegment	Default	0.8331	NonSegment	Default	0.8295
	Internet	Default	0.8194	Internet	Default	0.8154
	NonInternet	Default	0.8294	NonInternet	Default	0.8334
	NonSegment	Payback early	0.6666	NonSegment	Payback early	0.6666
	Internet	Payback early	0.6379	Internet	Payback early	0.6373
	NonInternet	Payback early	0.6584	NonInternet	Payback early	0.6586
No	Model 1			Model 4		
	NonSegment	Acceptance	0.7874	NonSegment	Acceptance	0.7707
	Internet	Acceptance	0.8018	Internet	Acceptance	0.7886
	NonInternet	Acceptance	0.7802	NonInternet	Acceptance	0.7694
	NonSegment	Default	0.8331	NonSegment	Default	0.8295
	Internet	Default	0.8136	Internet	Default	0.8073
	NonInternet	Default	0.8332	NonInternet	Default	0.8294
	NonSegment	Payback early	0.6666	NonSegment	Payback early	0.6666
	Internet	Payback early	0.6546	Internet	Payback early	0.6543
	NonInternet	Payback early	0.6613	NonInternet	Payback early	0.6614

Table 6.2: Compare the AuROC values on the holdout set on different models

Model 1 and 3 include the interaction term between Loan Amount and Rate (which is significant in both the acceptance and default modelling) and Model 2 and 4 do not. This is the major reason why Model 1 and 3 are consistently more predictive on acceptance and default modelling than Model 2 and 4 across each individual segment and segments combined. For example, the AuROC of acceptance modelling in Model 3 is 0.7874 while that AuROC value in Model 2 is 0.7707.

On the other hand, the benefit of individual stepwise selection on each segment is not so evident by comparing the results from Model 1 against Model 3 and the results from Model 4 against Model 2. It was expected that Model 3 should achieve higher AuROC values on the segmented data than Model 1 since individual stepwise selection procedures have been used. However, this was not observed. The AuROC values from individually stepwise selected models are not larger or even lower than the AuROC values from models that use variables selected from the combined set. Take the acceptance modelling results for example, The AuROC is 0.8001 in Model 3 while the AuROC is 0.8018 in Model 1 for the Internet segment. The AuROC is 0.7799 in Model 3 while the AuROC is 0.7803 in Model 1 for the NonInternet segment.

Recognizing the high significance of the interaction term and overall higher AuROC values, the details of the Model 1 results will be presented and discussed in section 6.4.1.1. The implications for the economic benefits in terms of unconditional expected profits and market shares will be given later in section 6.4.2 where both models (Model 1 and Model 3) will be compared.

6.4.1.1 Comparison of the iso-preference and iso-profits contours for Model 1

Figures 6.13 and 6.14 show the acceptance probabilities accordingly for the Internet and Non-Internet segments for Model 1 respectively [8]. Figures 6.15 and 6.16 show the unconditional expected profit for the Internet and Non-Internet segments for Model 1 respectively. Figures 6.17 and 6.18 show the iso-profits and iso-preference contours for the Internet and Non-Internet segments for Model 1 respectively. The iso-perference curves in Figure 6.17 are mapped from Figure 6.13 and the iso-perference curves on Figure 6.18 are mapped from Figure 6.14. Those iso-preference curves are marked with the probabilities of acceptance in the map, ranging from 0.1 to 0.9. Comparison between these two sets of iso-preference curves shows that for any given loan amount, a higher interest rate can be charged to the Non-Internet segment than to the Internet segment to achieve the same probability of acceptance. In other words, the Internet applicants are harder to please and attract.

The slopes of each contour, indicating the trade-off of higher interest rate for lower loan amount to maintain the probability of acceptance, are not noticeably different between the segments for any given loan amount. The iso-profits contours, however, differ substantially between the two segments. These contours are mapped from Figure 6.15 and 6.16 respectively for the Internet and Non-Internet segments. Comparison between the iso-profit contours shows that generally for a given requested loan amount, the highest unconditional expected profit that can be earned per customer is much higher for the Internet segment than that can be earned from the Non-Internet segment. For example for a requested loan amount of £20K, with the profit maximizing interest rate charged,

[8] Please note that all the results here are relating to a 'mean' applicant with covariates (except Loan Amount, Rate) assigned to their mean values.

the unconditional expected profits per applicant are around £110 for a Non-Internet applicant and around £220 if the applicant applied from Internet.

If the lender has a strategy of minimum level of profitability given a loan amount, we can compare the two segments from a different perspective . We can say that for a requested loan amount and given unconditional expected profit, the probability of acceptance (market share for the lender) will be higher for the Internet segment that for the Non-Internet segments. For example for a loan amount of £5K and unconditional expected profit per applicant of £90 the probability of acceptance is 0.67 for the Internet applicants (see point A in Figure 6.17) and around 0.51 for the Non-Internet group (see point B in Figure 6.18).

To achieve the highest market share (probability of acceptance) for a given unconditional expected profit level one has to find the tangency points between the corresponding iso-profit curve and the geometrically lowest iso-preference curve. Both Figure 6.17 and 6.18 show the situation for loans up to £25K, above which the frequency of observation in the data becomes very low. Unfortunately the tangency points for many levels of profits occur around this maximum loan amount due to the convexity of both curves [9]. Nevertheless differences appear from the figures. For example, to maximize the probability of acceptance subject to gaining an unconditional expected profit of £100 would require a requested loan amount of around £1K and interest rate at 20.69% for the Internet applicants (see point C in Figure 6.17), but a loan amount around £25K and rate at 7.06% for the Non-Internet applicants (see point D in Fig-

[9]The convexity of the iso-preference curve is due to the inclusion of the interaction term which was highly significant in the regression.

ure 6.18). Of course since the loan amount is requested by the applicant and rarely changed by the lender, it is not possible for a lender to freely choose which of the tangency points it wishes to locate at.

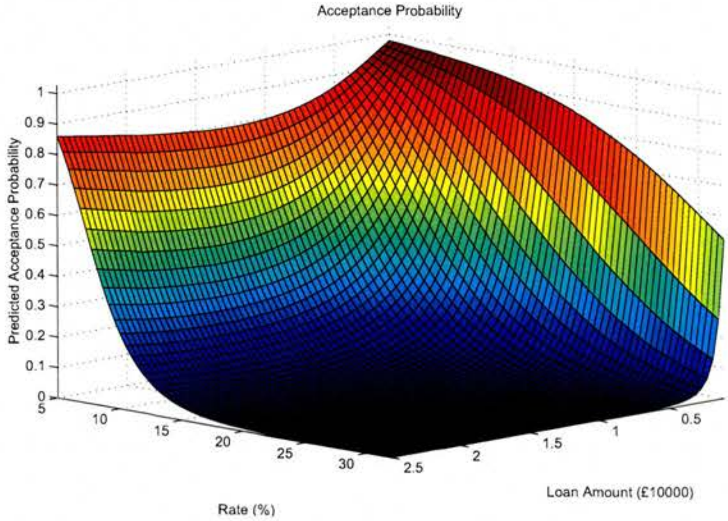

Figure 6.13: Acceptance probabilities for customers applying through Internet

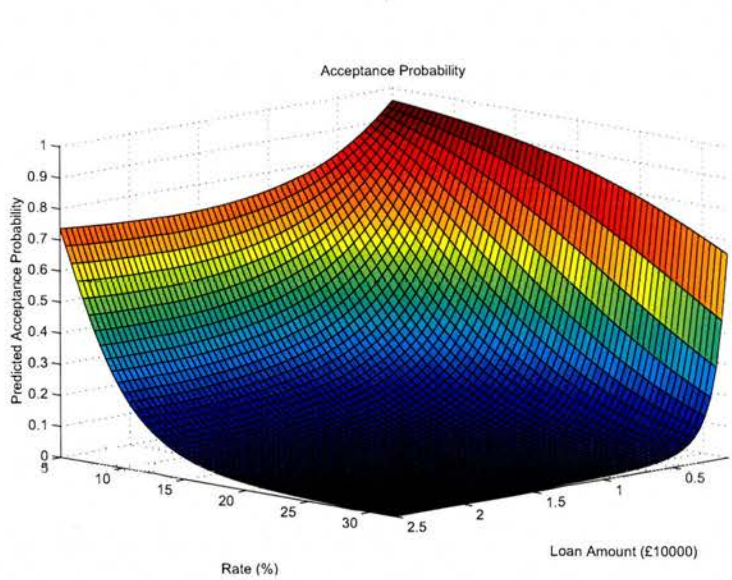

Figure 6.14: Acceptance probabilities for customers applying through Non-Internet

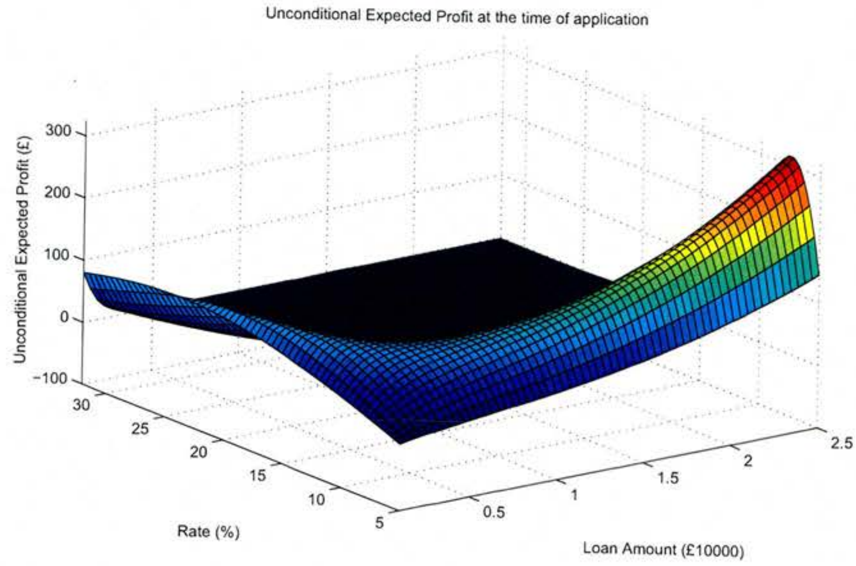

Figure 6.15: Unconditional expected profits for customers applying through Internet

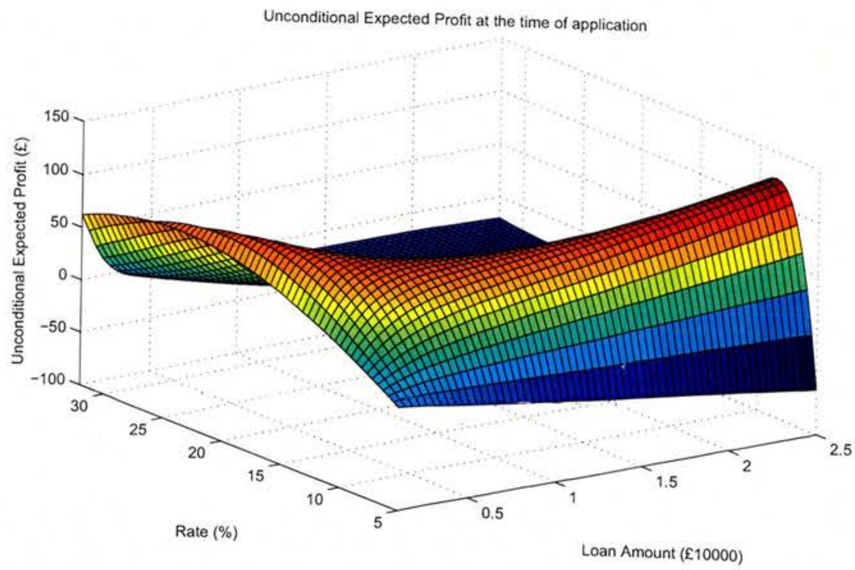

Figure 6.16: Unconditional expected profits for customers applying through Non-Internet

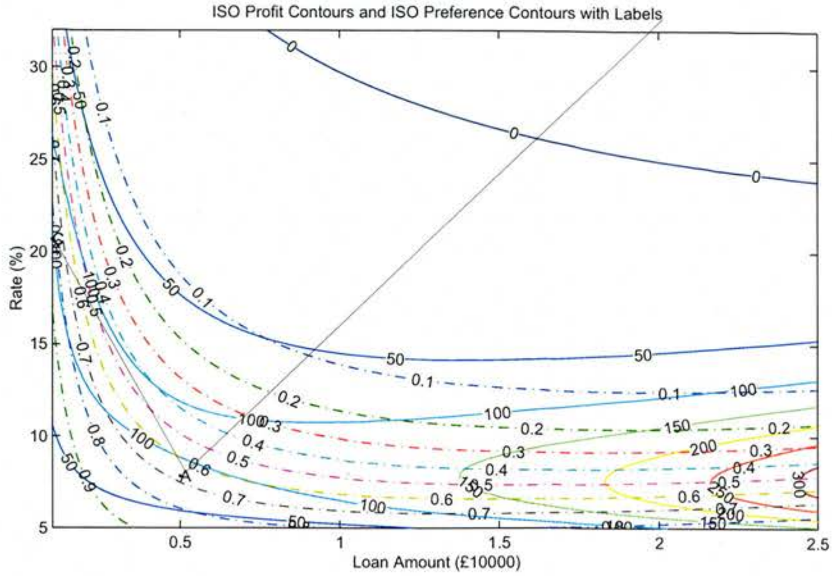

Figure 6.17: Iso-profit and iso-acceptance curves (marked with p(a) from 0.1 to 0.9)for customers applying through Internet

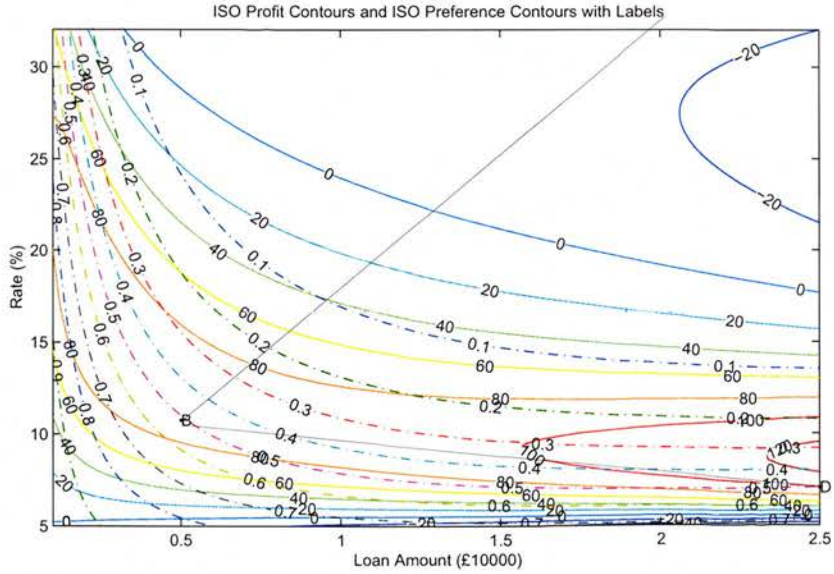

Figure 6.18: Iso-profit and iso-acceptance curves (marked with p(a) from 0.1 to 0.9) for customers applying through Non-Internet

6.4.2 Economic benefit of the segmentation

Tables 6.3 and 6.4 compare the economic benefit of the segmentation for each of the five different loan amounts, using results from Model 1 and Model 3 accordingly. The economic benefit is mainly measured by the total unconditional expected profits, although the expected total number of accepted customers is also included in the comparison since a lender may have objectives that relate to both.

The expected total unconditional profit for each given loan amount for each segment (or all segments combined) is calculated as the product of the expected unconditional profit per applicant within that specific segment (or all segments combined) and the number of customers, which is the number of applicants observed to apply for that specific given loan amount. Notice that we have assumed all applicant will receive an offer regardless of the probability of default. Of course we could modify this to estimate the unconditional expected profits from a subset of applicants who meet certain criteria such as positive profits or a probability of default.

The expected total number of customers who accept the offer is the total number of customers that apply for the offer that are expected to accept it. This is calculated as the probability of acceptance at each given loan amount times the number of applicants. Please note that the number of applicants at each given loan amount is observed rather than predicted. The assumption is that the distribution of applicants between segments at each given loan amount remains as observed.

With one exception, all of the relevant predictive variables (except Loan Amount and Interest Rate) in the calculation of the conditional expected profit per customer and

Table 6.3: Benefits of Segmentation Model 1

L=3000	Unconditional Profit Expected per applicant	Optimal Interest Rate	P(A)	Number of applicants	Total Profit Expected	Expected number of applicants who accept offer
Internet	106.26	12.93	0.56	302	32090.31	169.54
NonInternet	96.10	15.16	0.53	1148	110318.78	609.93
2 Segments Combined			0.54		142409.09	779.48
NonSegmentation	99.61	14.27	0.54	1450	144435.08	786.92
Gains from Segmentation					-2025.99	-7.44
L=7000	**Unconditional Profit Expected per applicant**	**Optimal Interest Rate**	**P(A)**	**Number of applicants**	**Total Profit Expected**	**Expected number of applicants who accept offer**
Internet	109.23	9.09	0.46	1096	119718.27	501.31
NonInternet	91.27	11.09	0.38	1455	132793.78	556.39
2 Segments Combined			0.41		252512.05	1057.70
NonSegmentation	95.17	10.06	0.40	2551	242786.83	1032.39
Gains from Segmentation					9725.21	25.31
L=10000	**Unconditional Profit Expected per applicant**	**Optimal Interest Rate**	**P(A)**	**Number of applicants**	**Total Profit Expected**	**Expected number of applicants who accept offer**
Internet	122.51	8.27	0.44	2260	276871.02	996.21
NonInternet	91.96	10.07	0.34	2879	264757.45	990.66
2 Segments Combined			0.39		541628.46	1986.87
NonSegmentation	100.64	9.13	0.38	5139	517165.83	1934.32
Gains from Segmentation					24462.63	52.55
L=17000	**Unconditional Profit Expected per applicant**	**Optimal Interest Rate**	**P(A)**	**Number of applicants**	**Total Profit Expected**	**Expected number of applicants who accept offer**
Internet	183.83	7.60	0.47	125	22979.35	58.51
NonInternet	102.93	9.08	0.32	96	9881.03	30.86
2 Segments Combined			0.40		32860.38	89.38
NonSegmentation	136.15	8.26	0.38	221	30089.06	84.42
Gains from Segmentation					2771.32	4.95
L=25000	**Unconditional Profit Expected per applicant**	**Optimal Interest Rate**	**P(A)**	**Number of applicants**	**Total Profit Expected**	**Expected number of applicants who accept offer**
Internet	315.10	7.52	0.54	472	148729.09	254.64
NonInternet	125.54	8.64	0.34	338	42431.40	116.27
2 Segments Combined			0.46		191160.49	370.92
NonSegmentation	208.76	8.01	0.44	810	169098.68	356.64
Gains from Segmentation					22061.81	14.27

Table 6.4: Benefits of Segmentation Model 3

L=3000	Unconditional Profit Expected per applicant	Optimal Interest Rate	P(A)	Number of applicants	Total Profit Expected	Expected number of applicants who accept offer
Internet	102.42	13.02	0.56	302	30930.39	170.24
NonInternet	95.43	15.05	0.54	1148	109554.44	616.13
2 Segments Combined			0.54		140484.83	786.37
NonSegmentation	99.61	14.27	0.54	1450	144435.08	786.92
Gains from Segmentation					-3950.25	-0.55
L=7000	**Unconditional Profit Expected per applicant**	**Optimal Interest Rate**	**P(A)**	**Number of applicants**	**Total Profit Expected**	**Expected number of applicants who accept offer**
Internet	94.40	9.11	0.46	1096	103467.66	502.19
NonInternet	90.82	11.04	0.38	1455	132140.04	558.43
2 Segments Combined			0.42		235607.71	1060.62
NonSegmentation	95.17	10.06	0.40	2551	242786.83	1032.39
Gains from Segmentation					-7179.13	28.23
L=10000	**Unconditional Profit Expected per applicant**	**Optimal Interest Rate**	**P(A)**	**Number of applicants**	**Total Profit Expected**	**Expected number of applicants who accept offer**
Internet	98.46	8.30	0.44	2260	222517.57	993.72
NonInternet	91.91	10.03	0.35	2879	264597.66	993.26
2 Segments Combined			0.39		487115.23	1986.98
NonSegmentation	100.64	9.13	0.38	5139	517165.83	1934.32
Gains from Segmentation					- 30050.61	52.66
L=17000	**Unconditional Profit Expected per applicant**	**Optimal Interest Rate**	**P(A)**	**Number of applicants**	**Total Profit Expected**	**Expected number of applicants who accept offer**
Internet	130.05	7.60	0.47	125	16256.64	58.76
NonInternet	104.76	9.05	0.32	96	10056.67	30.98
2 Segments Combined			0.41		26313.31	89.74
NonSegmentation	136.15	8.26	0.38	221	30089.06	84.42
Gains from Segmentation					-3775.75	5.32
L=25000	**Unconditional Profit Expected per applicant**	**Optimal Interest Rate**	**P(A)**	**Number of applicants**	**Total Profit Expected**	**Expected number of applicants who accept offer**
Internet	207.97	7.42	0.56	472	98161.56	262.48
NonInternet	131.89	8.60	0.35	338	44579.33	117.56
2 Segments Combined			0.47		142740.88	380.04
NonSegmentation	208.76	8.01	0.44	810	169098.68	356.64
Gains from Segmentation					- 26357.79	23.39

probability of acceptance have been assumed to take their mean values. The only exception is the values used for the Internet variable. To account for the interaction between the Internet variable and the Loan Amount variable, the mean values of the Internet variable used for calculating the $\widehat{\beta_0}$, $\widehat{\beta_0^b}$ and $\widehat{\beta_0^d}$ in Model 1 and Model 3 on the non-segment data at each given loan amount have been adjusted to use the mean values of Internet variable observed at each given loan amount.

Overall the expected unconditional profit per applicant is increasing with loan amount in Model 1 and Model 3 when the loan amount is larger than £3000, as seen in Table 6.3 and 6.4. Both models show that the unconditional expected profit per applicant for the Internet segment is higher than that for the Non-Internet segment. The difference in the unconditional expected profit between the two segments is also generally increasing with the loan amount in the two models. When the loan amount becomes larger, the Internet applicants are expected to be more profitable that their Non-Internet peers.

This shift of profitability between Internet and Non-Internet segments, becomes greater and greater when the loan amount increases, and can be observed in both models with one exception in Model 3. In Table 6.3 for Model 1, at loan amounts of £3K, the difference is $106.26 - 96.10 = 10.16$. For loan amounts of £7K, £10K, £17K, £25K, the difference are 17.96, 30.55, 80.91, 189.57, indicating the increasing profitability of the Internet applicants for the lender when the loan amount is larger. In Table 6.4 for Model 3, for loan amount of £3K, £7K, £10K, £17K, £25K, the difference are 6.99, 3.59, 6.55, 25.30, 76.08. Here one exception is observed when the loan amount increases from £3K to £7K the difference shrinks from 6.99 to 3.59.

What is not changing when loan amount gets larger is that the optimal interest rates that maximize the unconditional expected profits from Internet applicants is always lower than that for the Non-Internet segment, as can be shown in Tables 6.3. The difference between those optimal interest rates, though, is getting smaller and smaller when the loan size gets bigger. For loan amount of £3K, £7K, £10K, £17K, £25K, the differences between the Internet and Non-Internet optimal rate are -2.23, -2.00, -1.80, -1.48 and -1.12. This could be possibly explained away by the observation that in non-segment data, the optimal interest rates are getting smaller for larger loan amounts, therefore the difference between Internet and Non-Internet rates shall get smaller to keep the percentage of the difference at roughly a constant level.

The benefit of segmentation can be illustrated using gains in the total expected unconditional profits. For Model 1, as shown in Table 6.3, the gains measured as a percentage of the total profit before segmentation, are -1.40%, 4.01%, 4.73%, 9.21% and 13.05% at loan amounts of £3K, £7K, £10K, £17K and £25K respectively. For Model 3, as shown in Table 6.4, the gains measured as a percentage of the total profit before segmentation, are -2.74%, -2.96%, -5.81%, -12.55% and -15.59% at loan amounts of £3K, £7K, £10K, £17K and £25K respectively.

The benefit of segmentation can also be illustrated using the generally increased expected number of customers accepting the offer. For Model 1, as shown in Table 6.3, the total number of accepts, for loans of £3K, £7K, £10K, £17K and £25K, are changed by a percentage [10] of -0.95%, 2.45%, 2.72%, 5.87% and 4.00% accordingly. For Model 3, as shown in Table 6.4, the total number of accepts, for loans of £3K,

[10] Percentage of the number of accepts before the segmentation

£7K, £10K, £17K and £25K, are changed by a percentage of -0.07%, 2.73%, 2.72%, 6.30% and 6.56% accordingly.

The different results in Tables 6.3 and 6.4 demonstrate the importance of the specification of the models that are used to model the hazard functions and the probability of acceptance functions. Our initial expectation was that if all of these functions are individually estimated using stepwise selection routines and if interest rates are chosen to maximize unconditional expected profits at each loan amount separately for each segment, these segmentation would yield higher profits. In Model 1 we find it does. But it is possible that it may not do so for several reasons. One reason is that the estimated functions may fit the data less well at some loan amounts than at others. Another is that when the functions are estimated for each segment separately the smaller number of observations within the segment compared with larger sample size of the segments combined may cause the functions to fit less well in the former case than in the latter case.

6.5 Conclusion

This chapter provides an empirical methodology which a lender can employ to estimate the unconditional expected profits and expected acceptance probabilities at the time of application for an individual applicant by combining the estimates from acceptance modelling, Survival probabilities of the defaults and early repayment behaviours. This chapter also discussed potential optimal decision policies for a lender subject to different constraints. The results have also shown that it is possible, using proper model specification and careful interpretation, to segment the market by choosing the optimal interest rate for each segment to meet the lender's marketing objectives, being either to maximize the expected total profits overall all applicants or the market share (acceptance probability).

Chapter 7

Conclusion

Based on the theoretical model proposed by Keeney and Oliver, this thesis explores and models the relationships between offers of credit products, credit scores, consumers' acceptance decisions and expected profit using data that records the actual choices made by customers and their monthly account status after being accepted. This concluding chapter is divided into three sections. The first section will summarize the research findings by answering the questions raised earlier in the Introduction chapter. The second section will reiterate the contribution to the knowledge. The third section will discuss the limitations within our modelling approaches and outline some possible research directions in the future.

7.1 Summary of findings

In Chapter 1, the following research questions were asked:

- Why do we need to model the profitability of making a loan, unconditional on the acceptance by the applicants, and how can iso-profit and iso-acceptance contours be empirically estimated and presented?

- Is acceptance inference needed?
- How do novel approaches like support vector machines (SVMs) perform (compared to logistic regression) in predicting default and acceptance ?
- How to model the chance of default and paying back early and how to incorporate them into a profit estimation?

The first question involving the acceptance and profitability modelling is the key one in this thesis. The other questions naturally emerge during the course of the investigation into the modelling of acceptance and profitability. The following subsections will summarise our findings as answers to these questions.

7.1.1 How can we model the profitability of making a loan, unconditional on the acceptance by the applicants, and how can iso-profit and iso-acceptance contours be empirically estimated and presented?

Keeney and Oliver's theoretical work provided a foundation for our model. Keeney and Oliver's work empathized two objectives for lenders: profitability and market share (the probability of the acceptance of offers). Their concept of profitability unconditional on the acceptance of offers differs from previous research where the profitability analysis focused on customers who have already accepted offers. Earlier research omitted the probability that an applicant will accept an offer in the analysis of potential profit to a lender and so such could not estimate expected profit at the time of application.

To model this profitability, unconditional on the acceptance of offers, both behaviours, acceptance and profitability, need to be modelled. In Chapter 2, we reviewed the previous literature on acceptance modelling and profitability modelling.

We estimate the expected profits unconditional on acceptance in Chapter 6 by combining previous results in the acceptance and survival analyses of default and paying back early. The results for a single applicant are presented in a three dimensional space of Profit vs Rate vs Loan Amount. The iso-profit contours are drawn by connecting all the points representing the same amount of expected profit in a two dimensional space of Loan Amount vs Rate. Similarly, the iso-preference contours are drawn by connecting all the offers with the same level of acceptance probabilities in the same two dimensional space of Loan Amount vs Rate. By examining those iso-profit and iso-acceptance contours, different profit optimising decision policies can be derived under various constraints.

A further segmentation analysis has also been conducted by separating the samples into two different groups, Internet and Non-internet applicants, with parameters estimated individually. On each segment, different profit-maximizing interest rates were found and we found some economic benefit when aggregated over all applicants can be achieved through this segmentation exercise.

7.1.2 Is acceptance inference needed?

The need of acceptance inference has been explored in Chapter 4 via fitting models of bivariate Probit sample selection. The results indicate that our model does not suffer from the sample selection bias [1] unless using a lean model in which a significant sample selection bias has been observed. The conclusion is that acceptance inference may not be needed for our data.

7.1.3 How do novel approaches like support vector machines (SVMs) perform (compared to logistic regression) in predicting default and acceptance ?

In the newly introduced credit scoring methodologies, SVMs have received a lot of interest and been reported to be quite competitive in the literature (such as Baesens (2003) and Baesens et al. (2003)). SVMs with various kernels have been used to predict the default in Chapter 3 and the acceptance of offers in Chapter 4. In the default prediction, we found SVMs are not competitive compared to the logistic regression (the performance measure used is the area under the ROC curve on the holdout set). In acceptance prediction, the performance of SVMs were observed to be as predictive as logistic regression (using the area under the ROC curve on the holdout set as performance measure). One of the reasons behind this may be the difference in the class distributions of default and acceptance. Another explanation for the better performance of logistic regression may be its appropriate size of model complexity compared to that of SVMs to avoid the danger of over fitting. As a conclusion, we have not found that SVMs out-perform logistic regression based on our data.

[1] Assuming that the residuals of the two equations are normally distributed.

7.1.4 How to model the chance of default and paying back early and how to incorporate them into a profit estimation?

Accurate profit estimation requires estimating whether the applicant will default, but also when the default is expected to happen. An earlier defaulter causes more loss than a later one as the latter has made more payments and therefore left a smaller balance as loss to the lender. Binary outcome predictors like static logistic regression models, lack the capability to predict the timing of an event whereas this is the aim of survival analysis models. The latter have been shown to be equally predictive in tasks of binary outcome predictions over specific outcome windows as logistic regression.

Paying back early was observed to be much more frequent in our data than the events of default and paying back early events are also very important to lenders in terms of profitability, in spite of being rarely modelled in the literature. A customer with a higher probability to pay back early brings less profit for the lender and therefore impacts adversely on the profitability of the whole portfolio.

Chapter 5 presents survival analysis results of modelling the default and paying back early events. Different semi-parametric and parametric models have been compared and Cox PH models were the best performing models. Chapter 5 also models the survival probabilities of default and paying back early under the competing risk framework. Three different competing risk models have been compared. However, no improvement has been observed in the predictive performance.

Together with the coefficients estimated from Acceptance models, the coefficients from the Cox PH models for default and paying back early events estimated separately, are then plugged into the equation to calculate unconditional expected profits for a fixed term loan. The details have been given in Chapter 6.

7.2 Contributions to knowledge reiterated

This section will reiterate the eight major contributions this thesis has made to knowledge. First, this thesis is the first empirical academic study to estimate expected profits at the time of application. Unlike previous studies, which predict the profits of a customer who has already accepted an offer, this thesis estimated the expected profits at the time of application by combining the results from acceptance modelling, survival analysis of default and paying back early.

Second, the customers' acceptance behaviours are estimated based on a data set that contains *the actual acceptance choices made by customers of a real financial product.* Instead of using hypothetical data as in previous research, findings based on our data shall be closer to what will be observed in the practical retail lending industry.

Third, this thesis found that the iso-preference curves drawn from empirically estimated results indicate a preference towards lower loan amounts rather than higher amounts. This appears contrary to those assumed in Keeney and Oliver's theoretical model, where larger credit lines are presumed to be preferred over lower credit lines.

Fourth, this thesis discussed different profit maximizing strategies the lender may choose under different marketing objectives. What has made this thesis different from previous researches is that this thesis has used estimates from industry sourced data for the profit calculation rather than assumed numerical cost and profit figures.

Fifth, this thesis provided a segmentation analysis based on separately estimating profits on Internet and Non-Internet groups. This segmentation practice was demonstrated to lead to markedly different policy decisions compared with the decisions drawn before the segmentation.

Sixth, the possible existence of sample selection bias introduced in the process of acceptance has been explored using bivariate Probit sample selection models. Previous literature focused on the study of reject inference and paid less attention to the corresponding scenario of acceptance inference. Our results suggested that acceptance inference might not be necessary.

Seventh, this thesis revisited the topic of comparing the classification methods, SVMs vs. logistic regression. Varied predictive performances of SVMs on different prediction tasks (default and acceptance) have been observed. SVMs were found to perform poorly against logistic regression in predicting default, in contrast to the good performance reported in the literature (Baesens (2003) and Baesens et al. (2003)). Another novelty in this thesis is the application of SVMs to predict acceptance. We found that SVMs produced similar predictive performance as logistic regression did. One possible explanation for this varied performance on different data by SVMs is the difference in the class distributions. The default data is much more unbalanced than the acceptance data, hence more challenging to SVMs, which are more sensitive to the class distribution.

Eighth, this thesis compared competing risk survival models against those separately estimated survival models in the prediction of default and paying back early and observed little improvement in the predictive performance from competing risk models.

7.3 Limitations and future research

There are still some limitations that exist in our work. First, the cost for the lender to service the loan in our profitability model has been assumed to consist of only the running cost for the lender to borrow at the inter-bank rate. This simplification is adopted because detailed data relating to fixed costs are unfortunately not available.

Second, this profitability model has not considered impacts on the profits of the economic cycle. Adding macro economic variables into the set of predictors, such as what the models in Bellotti and Crook (2007a) do, might offer a more robust model at different stages in the economic cycles. However, in our data the longest duration of observed performance is 26 months, which is too short to cover a whole economic cycle.

A number of possible extensions can be suggested for future research. For example, only one type of segmentation analysis is done, the Internet Non-Internet segmentation. It would be interesting to calculate the expected profits for other segments and to see how much the economic benefits can be increased through segmenting onto other groups and how different optimal decision strategies can be derived.

Profitability analysis can also be applied to other products. The profitability analysis in this thesis is based only on a fixed term loan product. Extending profitability analysis to other types of credit products such as mortgage or credit cards could be very interesting.

Our model has not considered the capital adequacy requirement the lender is bound to abide by. Under Basel II, the latest capital requirement, the lender has to cover the unexpected loss by setting aside a minimum amount of capital which is a function of PD and LGD and other parameters and the type of product. More capital required means less return on economic capital for the lender. Different optimal decision policies might be needed under such capital requirements.

Finally, the confidence intervals of AuROC estimates have not been calculated. Providing such estimates could provide facility to check if the difference in predictive performance is significant. One way to generate such estimates is to sample the data using sampling method like bootstrapping and report the AuROC results distribution, from which confidence intervals can be drawn.

References

Albert, A. and Anderson, J. (1984). On the existence of maximum likelihood estimates in logistic regression models. *Biometrika.*

Andreeva, G. (October 2006). European generic scoring models using survival analysis. *Journal of the Operational Research Society*, 57:1180–1187(8).

Andreeva, G., Ansell, J., and Crook, J. (2007). Modelling profitability using survival combination scores. *European Journal of Operational Research*, 127(3):1537–1549.

Andreeva, G., Ansell, J., and Crook, J. N. (September 2005). Modelling the purchase propensity: analysis of a revolving store card. *Journal of the Operational Research Society*, 56:1041–1050(10).

Attanasio, O. P. (1999). *Consumption, Handbook of Macroeconomics, Volume 1, Part 2, Pages 741-812.* Elsevier.

Baesens (2003). Developing intelligent systems for credit scoring using machine learning techniques. *PhD Thesis, Katholieke Universiteit, Leuven.*

Baesens, B., Gestel, T. V., Stepanova, M., den Poel, D. V., and Vanthienen, J. (2005). Neural network survival analysis for personal loan data. *Journal of Operational Research Society (2005) 56, 1089-1098.*

Baesens, B., Gestel, T. V., 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, Vol. 54, No. 6. , pp. 627-635.*

Banasik, J. and Crook, J. (2005). Credit scoring, augmentation and lean models. *Journal of the Operational Research Society (2005) 56, 1072-1081.*

Banasik, J., Crook, J., and Thomas, L. (1999). Not if but when will borrowers default. *Journal of the Operational Research Society 50: 1185-1190.*

Banasik, J., Crook, J., and Thomas, L. (2003). Sample selection bias in credit scoring models. *Journal of the Operational Research Society (2003) 54, 822C83.*

Beling, P., Covaliu, Z., and Oliver, R. (2005a). Optimal scoring cutoff policies and efficient frontiers. *Journalof the Operational Research Society (2005) 56, 1016-1029.*

Beling, P., Jiang, W., Oliver, R., and Overstreet, G. (2005b). Retail credit portfolio optimisation under risk. *Presentation at Credit Scoring and Credit Control IX con ference, Edinburgh,September.*

Bellotti, T. and Crook, J. (2007a). Credit scoring with macroeconomic variables. *Credit Scoring Conference 2007.*

Bellotti, T. and Crook, J. (2007b). Credit scoring with macroeconomic variables using survival analysis. *Credit Scoring Research Center working paper.*

Bellotti, T. and Crook, J. (2007c). Support vector machines for credit scoring and discovery of significant features. *Credit Research Centre Working Paper.*

Blochlinger, A. and Leippold, M. (2006). Economic benefit of powerful credit scoring. *Journal of Banking and Finance Volume: 30 Issue: 3 Pages: 851-873.*

Boyes, W. J., Hoffman, D. L., and Low, S. A. (1989). An econometric analysis of the bank credit scoring problem. *Journal of Econometrics 40(1989) 3-14. North-Holland.*

Breslow, N. (1974). Covariance analysis of censored survival data. *Biometrics, Vol. 30 No.1 pp. 89-90.*

Burges, C. J. C. (1998). A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2(2):121–167.

Chang, C.-C. and Lin, C.-J. (2001). *LIBSVM: a library for support vector machines.* Software available at http://www.csie.ntu.edu.tw/ cjlin/libsvm.

Chapelle, O., Haffner, P., and Vapnik, V. N. (1999). Support vector machines for histogram-based image classification. *IEEE TRANSACTIONS ON NEURAL NETWORKS, VOL. 10, NO. 5, SEPTEMBER.*

Chow, C. and Liu, C. (1968). Approximating discrete probability distributions with dependence trees. *IEEE Transactions on Information Theory.*

Cleves, M. (1999). Analysis of multiple failure-time survival data. *StataCorp http://www.stata.com/support/faqs/stat/stmfail.html.*

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society, Series B , Vol.34, No.2.*

Cox, D. R. and Snell, E. J. (1968). A general definition of residuals. *Journal of the Royal Statistical Society. Series B (Methodological).*

Crook, J. and Banasik, J. (2004). Does reject inference really improve the performance of application scoring models? *Journal of Banking & Finance Volume 28, Issue 4, Retail Credit Risk Management and Measurement, April, Pages 857-874.*

Crook, J., Hamilton, R., and L.C.Thomas (1992). The degradation of the scorecard over the business cycle. *IMA Journal of Mathematics applied in business Industry.*

Crook, J. N., Edelman, D. B., and Thomas, L. C. (2007). Recent developments in consumer credit risk assessment. *European Journal of Operational Research 183 14471465.*

Cyert, R., Davidson, H., and Thompson, G. (1962). Estimation of the allowance for doubtful accounts by markov chains. *Management Science.*

DeLong, E. R., DeLong, D. M., and Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: A non-parametric approach. *Biometrics, Vol. 44, No. 3 (Sep., 1988), pp. 837-845.*

Efron, B. (1977). The efficiency of cox's likelihood function for censored data. *Journal of the American Statistical Association, Vol. 72, No. 359, pp. 557-565.*

Friedman, N., Geiger, D., and Goldszmidt, M. (1997). Bayesian network classifiers. *Machine Learning*, 29(2-3):131–163.

Frydman, H. (1984). Maximum likelihood estimation in the mover-stayer model. *Journal of the American Statistical Association Vol.79, NO.387, pp.632-638.*

Frydman, H., Kallberg, J. G., and Kao, D. L. (1985). Testing the adequacy of markov chain and mover-stayer models as representations of credit behavior. *Operations Research, Vol.33, NO.6 pp.1203-1214.*

Garrett, J. M. (1997). Graphical assessment of the cox model proportional hazards assumption. *Stata Technical Bulletin 35: 9-14.*

Grambsch, P. and Therneau, T. (1994). Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika.*

Greene, W. (1998). Sample selection in credit-scoring models. *Japan and the World Economy 10(1998) 299-316.*

Greene, W. H. (1992). A statistical model for credit scoring. *New York University, Leonard N. Stern School of Business, Department of Economics in its series Working Papers with number 92-29.*

Hand, D. and Kelly, M. (2001). Lookahead scorecards for new fixed term credit products. *Journal of the Operational Research Society, Volume 52, Number 9, 1 September, pp. 989-996(8).*

Hand, D. J. (1997). *Construction and assessment of classification rules.* Wiley.

Hand, D. J. (2006). Classifier technology and the illusion of progress. *Statistical Science.*

Hand, D. J. and Henley, W. E. (1993). Can reject inference ever work? *IMA Journal of Mathematics Applied in Business and Industry (1993/4) 4, 45-55.*

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica, Vol.47, No. 1.*

Ho, J., Thomas, L. C., Pomroy, T., and Scherer, W. T. (2004). *Segmentation in Markov chain consumer credit behavioural models in Thomas, Edelman, Crook J (eds) Readings in Credit Scoring.* Oxford University Press.

Hopper, M. and Lewis, E. (1992). Development and use of credit profit measures for account management. *IMA Journal of Mathematics Applied in Business and Industry, 4,3-17.*

Hosmer, D. W. and Lemeshow, S. (1998). *Applied Survival Analysis, Regression Modeling of Time to Event Data.* Wiley.

Hosmer, D. W. and Lemeshow, S. (2000). *Applied Logistic Regression.* Wiley.

Jung, K. M., Thomas, L., and Thomas, S. D. A. (2003). Modelling acceptance probabilities. *Proceedings of Credit Scoring and Credit Control 7.*

Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association, Vol. 53, NO.282. pp 457-481.*

Keeney, R. L. and Oliver, R. M. (2004). *Improving lender offers using consumer preferences in Thomas, Edelman, Crook J (eds) Readings in Credit Scoring.* Oxford University Press.

Keeney, R. L. and Oliver, R. M. (2005). Designing win-win financial loan products for consumers and businesses. *Journal of the Operational Research Society 56: 1030-1040.*

Kelly, M. G. and Hand, D. J. (1999). Credit scoring with uncertain class definitions. *IMA Journal of Management Mathematics 1999 10(4):331-345;.*

Kelly, M. G., Hand, D. J., and Adams, N. M. (1999). The impact of changing populations on classifier performance. *Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining.*

Keshk, O. M. G. (2003). Cdsimeq: A program to implement two-stage probit least squares. *Stata Journal,Issue (Month): 2 (June),Pages: 157-167.*

Lambrecht, B. M., Perraudin, W. R., and Satchell, S. (2003). Mortgage default and possession under recourse: A competing hazards approach. *Journal of Money, Credit and Banking, Issue: 3.*

Li, H. G. and Hand, D. (2002). Direct versus indirect credit scoring classifications. *Journal of Operational Research Society ,53, 1-8.*

Lia, S.-T., Shiueb, W., and Huang, M.-H. (2006). The evaluation of consumer loans using support vector machines. *Expert Systems with Applications 30 (2006) 772782.*

Liebman, L. H. (1972). A markov decision model for selecting optimal credit control policies. *Management Science, Vol.18, NO.10, pp. B519-B525.*

Lunn and McNeil (1995). Applying cox regression to competing risks. *Biometrics 51 524-532, June.*

Maddala (1983). *Limited-dependent and qualitative variables in econometrics.* Cambridge University Press.

Marshall, K. T. and Oliver, R. M. (1995). *Decision Making and Forecasting.* MaGraw-Hill.

Meng, C.-L. and Schmidt, P. (1985). On the cost of partial observability in the bivariate probit model. *International Economic Review 26, P71-85.*

Montgomery, A. L. (2001). Applying quantitative marketing techniques to the internet. *Interfaces.*

Narain, B. (1992). Survival analysis and the credit granting decision. *Credit Scoring and Credit Control 109-122.*

Oliver, R. (1993). Effects of calibration and discrimination on profitabilityscoring. *Proceedings of Credit Scoring and Credit Control.*

Oliver, R. and Wells, E. (2001). Efficient frontier cutoff policies in credit portfolios. *Journal of Operational Research Society.*

Oliver, R. M. and Thomas, L. C. (2005). How basel will affect optimal cut-offs. *CREDIT SCORING AND CREDIT CONTROL X.*

Ong, C.-S., Huang, J.-J., and Tzeng, G.-H. (2005). Building credit scoring models using genetic programming. *Expert Systems with Applications Volume 29, Issue 1 , July 2005, Pages 41-47.*

Osuna, E., Freund, R., and Girosi, F. (1997). Support vector machines: Training and applications. Technical Report AIM-1602.

Rossi, P. E., McCulloch, R. E., and Allenby, G. M. (1996). The value of purchase history data in target marketing. *Marketing Science.*

Schebesch, K. B. and Stecking, R. (2005a). Support vector machines for classifying and describing credit applicants: detecting typical and critical regions. *Journal of Operational Research Society,56, 1082-1088.*

Schebesch, K. B. and Stecking, R. (2005b). Support vector machines for credit scoring:extension to non standard cases. *Innovations in Classification, Data Science and Information Systems.*

Schoenfeld, D. (1982). Partial residuals for the proportional hazards regression model. *Biometrika.*

Seow, H.-V. and Thomas, L. C. (2005). To ask or not to ask, that is the question. *Credit Scoring Conference 2005.*

Somers, M. and Whittaker, J. (2007). Quantile regression for modelling distributions of profit and loss. *European Journal of Operational Research Volume 183, Issue 3, 16 December 2007, Pages 1477-1487.*

Srinivasan, C. and Kim, Y. H. (1987). Credit granting: A comparative analysis of classification procedures. *The Journal of Finance, Vol. 42, No. 3,.*

Stepanova, M. and Thomas, L. (1999). Survival analysis methods for personal loan data. *Credit Scoring and Credit Control VI.*

Stepanova, M. and Thomas, L. (2001). Phab scores: Proportional hazards analysis behavioural scores. *Journal of Operational Research Society 52: 1007-1016.*

Stepanova, M. and Thomas, L. (2002). Survival analysis methods for personal loan data. *OPERATIONS RESEARCH Vol. 50, No. 2, March-April, pp. 277-289.*

Thomas, L., Jung, K. M., Thomas, S. D., and Wu, Y. (2006). Modelling consumer acceptance probabilities. *Expert Systems with Applications 30(2006) 499-506.*

Thomas, L. C., Edelman, D. B., and N.Crook, J. (2002). *Credit Scoring and Its Applications.* SIAM.

Till, R. J. and Hand, D. J. (2001). Markov models in consumer credit. *Credit Scoring and Credit Control 7 Conference, Edinburgh.*

Trench, M. S., Pederson, S. P., Lau, E. T., Ma, L., Wang, H., and Nair, S. K. (2003). Managing credit lines and prices for bank one credit card. *Interfaces, Volume: 33 , Issue: 5.*

Vapnik, V. N. (1995). *The Nature of Statistical Learning Theory.* Springer.

Winterfeldt, D. and Edwards, W. (1982). Costs and payoffs in perceptual research. *Psychological Bulletin, 91.*

Yobas, M. B., Crook, J. N., and Ross, P. (2000). Credit scoring using neural and evolutionary techniques. *IMA Journal of Management Mathematics 2000 11(2):111-125.*

Appendix A

Appendix for Chapter 3

Variable	Dummy Variable	MLE estimates	Pr > ChiSq
Intercept		-3.6218	< .0001
Rate		12.2737	< .0001
Insurance	N	-0.5967	< .0001
LOAN_AMT	(1.2e+04,1.5e+04]	-0.0860	0.3907
LOAN_AMT	(1.5e+04,2.3e+04]	-0.0840	0.4805
LOAN_AMT	(1e+04,1.2e+04]	-0.1434	0.2623
LOAN_AMT	(2.3e+04,2.5e+04]	0.3544	0.0111
LOAN_AMT	(5e+03,6.5e+03]	0.0710	0.5207
LOAN_AMT	(6.5e+03,8e+03]	0.1578	0.0953
LOAN_AMT	(8e+03,1e+04]	-0.0369	0.6621
TERM	24	-0.4330	0.0012
TERM	36	-0.2416	0.0059
TERM	48	-0.0373	0.6625
TERM	60	0.1981	0.0035
TERM	72	0.0384	0.8303
newbus	0	-0.1485	0.0024
ALCIFDET	EMP	0.2847	0.0157
CCJGT500	(22,27]	-0.0422	0.6516
CCJGT500	(27,32]	0.2038	0.0140
CCJGT500	(32,37]	-0.1849	0.0579
CCJGT500	(37,43]	0.2147	0.0142
CCJGT500	(43,58]	-0.0569	0.5286
CCJGT500	EMPTY	-0.1334	0.0583
LOANBAL1	(1.29e+04,6.49e+04]	-0.1379	0.1690
LOANBAL1	(1.37e+06,8.72e+06]	0.4300	0.0005
LOANBAL1	(1.53e+05,3e+05]	-0.0875	0.3610
LOANBAL1	(3e+05,4.89e+05]	-0.00468	0.9606
LOANBAL1	(4.89e+05,7.4e+05]	-0.00170	0.9852
LOANBAL1	(6.49e+04,1.53e+05]	-0.2884	0.0062
LOANBAL1	(7.4e+05,1.37e+06]	0.1479	0.0925
NETINCM	(1.05e+03,1.2e+03]	-0.1730	0.0556
NETINCM	(1.2e+03,1.38e+03]	-0.2400	0.0288
NETINCM	(1.38e+03,1.55e+03]	-0.0469	0.6287
NETINCM	(1.55e+03,1.8e+03]	-0.0197	0.8418
NETINCM	(1.8e+03,2.2e+03]	0.0591	0.5511
NETINCM	(2.2e+03,3.5e+03]	0.2635	0.0082
NETINCM	(3.5e+03,9.37e+05]	0.6437	< .0001
NETINCM	(900,1.05e+03]	-0.1742	0.0823
TOSETTL6	EMPTY	-0.1518	0.1438
TOSETTL6	N	0.2665	0.0230

Table A.1: Maximum likelihood estimates of parameters by Logistic Regression across all bands. To be continued in table A.2

Variable	Dummy Variable	MLE estimates	Pr > ChiSq
AGE	(27,31]	0.1423	0.1526
AGE	(31,34]	0.0211	0.8409
AGE	(34,37]	-0.1060	0.3194
AGE	(37,41]	0.0292	0.7453
AGE	(41,45]	-0.1155	0.2420
AGE	(45,50]	-0.0892	0.3777
AGE	(50,57]	-0.3926	0.0007
AGE	(57,64]	0.0658	0.7216
TIMEBANK	(1.1e+03,1.41e+03]	0.1652	0.1055
TIMEBANK	(1.41e+03,1.61e+03]	-0.0107	0.9187
TIMEBANK	(1.61e+03,2e+03]	0.0749	0.3966
TIMEBANK	(2.4e+03,3e+03]	0.0407	0.7191
TIMEBANK	(2e+03,2.4e+03]	-0.2843	0.0803
TIMEBANK	(3e+03,8.2e+03]	-0.8381	0.0014
TIMEBANK	(500,900]	0.2211	0.0217
TIMEBANK	(900,1.1e+03]	0.1914	0.0496
SPSETLD	0	0.5221	< .0001
SPSETLD	1	0.2555	0.0052
SPSETLD	2	0.1540	0.0961
SPSETLD	3	0.0917	0.3592
SPSETLD	4	-0.1486	0.1703
SPSETLD	5	0.0308	0.7831
SPSETLD	6	-0.0905	0.4891
SPSETLD	7	-0.1121	0.4337
SPSETLD	8	-0.3901	0.0354
SPVALDEL	-1	1.0130	0.0066
SPVALDEL	0	0.3766	0.2589
SPVALDEL	1	0.8392	0.0191
SSRC4TO6	0	-0.4315	< .0001
SSRC4TO6	1	-0.2540	0.0024
SSRC4TO6	2	0.0499	0.6391
SSRC4TO6	3	0.0393	0.8070
SWRSTCUR	0	-0.0126	0.8837
SWRSTCUR	EMP	0.4754	0.0002
SWRSTCUR	N	-0.2818	0.1883
WRST46AL	0	0.1486	0.4468
WRST46AL	1	-0.3887	< .0001
WRST46AL	2	-0.3354	0.0237
WRST46AL	3	0.0669	0.5270

Table A.2: Following table A.1. Maximum likelihood estimates of parameters by Logistic Regression across all bands.

Variable	Dummy Variable	MLE estimates	Pr > ChiSq
Intercept		0.7011	0.2529
Insurance	N	-1.0505	< .0001
SOCNOACT	EMPTY	1.9042	0.0019
INC_SURP	(1.12e+03,1.38e+03]	0.5048	0.3566
INC_SURP	(1.38e+03,1.79e+03]	-1.1836	0.0600
INC_SURP	(1.79e+03,2.96e+03]	1.4598	0.0115
INC_SURP	(2.96e+03,9.37e+04]	-0.1744	0.8771
INC_SURP	(395,565]	-0.0345	0.9338
INC_SURP	(565,726]	-0.3132	0.4683
INC_SURP	(726,910]	-0.0784	0.8536
INC_SURP	(910,1.12e+03]	-1.2085	0.0848

Table A.3: Maximum likelihood estimates of parameters by Logistic Regression after stepwised selection based on 223 cases in band 10. [1]

Variable	Dummy Variable	MLE estimates	Pr > ChiSq
Intercept		1.9976	0.9924
Insurance	N	-0.5738	0.0001
TOSETTL4	EMPTY	0.7048	0.0068
TOSETTL4	N	-0.0884	0.7590
AGE	(27,31]	-0.0309	0.9343
AGE	(31,34]	-0.6688	0.1647
AGE	(34,37]	-0.4486	0.4459
AGE	(37,41]	1.1564	0.0010
AGE	(41,45]	1.4336	0.0002
AGE	(45,50]	-0.5223	0.3746
AGE	(50,57]	-0.6051	0.3138
AGE	(57,64]	-0.3528	0.7346
SPVALDEL	-1	-3.1530	0.9881
SPVALDEL	0	-4.1920	0.9841
SPVALDEL	1	-2.8148	0.9893

Table A.4: Maximum likelihood estimates of parameters by Logistic Regression after stepwised selection based on 389 cases in band 20.

Variable	Dummy Variable	MLE estimates	Pr > ChiSq
Intercept		-2.7428	$<.0001$
Rate		8.5002	0.0075
Insurance	N	-0.5847	$<.0001$
TERM	24	-0.4375	0.2826
TERM	36	-0.0925	0.7016
TERM	48	0.2013	0.3970
TERM	60	0.2751	0.1645
TERM	72	-0.6943	0.2739
SEARCHES	(15,32]	-0.1104	0.6524
SEARCHES	(32,49]	-0.0686	0.7727
SEARCHES	(49,65]	-0.2743	0.2277
SEARCHES	(65,80]	0.1117	0.6139
SEARCHES	(80,95]	0.2383	0.2786
SEARCHES	(95,100]	0.6558	0.0351
TOSETTL1	EMPTY	0.4997	0.0029
TOSETTL1	N	0.00893	0.9421

Table A.5: Maximum likelihood estimates of parameters by Logistic Regression after stepwised selection based on 997 cases in band 30.

Variable	Dummy Variable	MLE estimates	Pr > ChiSq
Intercept		-2.9717	< .0001
Rate		11.4369	0.0140
Insurance	N	-0.5551	< .0001
LOAN_AMT	(1.2e+04,1.5e+04]	0.3109	0.2365
LOAN_AMT	(1.5e+04,2.3e+04]	0.1445	0.6185
LOAN_AMT	(1e+04,1.2e+04]	0.1555	0.6335
LOAN_AMT	(2.3e+04,2.5e+04]	0.0584	0.8996
LOAN_AMT	(5e+03,6.5e+03]	-1.0087	0.0315
LOAN_AMT	(6.5e+03,8e+03]	0.8449	0.0002
LOAN_AMT	(8e+03,1e+04]	0.0517	0.8336
INC_SURP	(1.12e+03,1.38e+03]	-1.0187	0.0165
INC_SURP	(1.38e+03,1.79e+03]	-0.9583	0.0259
INC_SURP	(1.79e+03,2.96e+03]	0.3283	0.2462
INC_SURP	(2.96e+03,9.37e+04]	0.5199	0.2135
INC_SURP	(395,565]	0.4628	0.0382
INC_SURP	(565,726]	0.2670	0.2775
INC_SURP	(726,910]	-0.5160	0.1377
INC_SURP	(910,1.12e+03]	0.4539	0.0896
SPL6M4	0	-0.5685	0.0641
SPL6M4	1	-0.3942	0.3923
SPL6M4	EMPTY	1.4446	0.1416
SPL6M4	N	0.0932	0.7582
SSRC4TO6	0	-0.4648	0.0265
SSRC4TO6	1	-0.3155	0.1442
SSRC4TO6	2	0.2700	0.3124
SSRC4TO6	3	-0.5614	0.2102

Table A.6: Maximum likelihood estimates of parameters by Logistic Regression after stepwised selection based on 1327 cases in band 40.

Variable	Dummy Variable	MLE estimates	Pr > ChiSq
Intercept		-3.9773	< .0001
Rate		14.1654	0.0008
Insurance	N	-0.4936	< .0001
TERM	24	-0.3190	0.3719
TERM	36	-0.6210	0.0109
TERM	48	0.1662	0.4223
TERM	60	0.3625	0.0226
TERM	72	0.2218	0.5527
LOANBAL4	(2.2e+04,3.55e+05]	-0.2979	0.1034
LOANBAL4	(3.55e+05,3.83e+06]	0.6149	0.0020
MOR_RENT	(156,227]	-0.4059	0.1360
MOR_RENT	(227,300]	0.0418	0.8442
MOR_RENT	(300,360]	0.3197	0.2146
MOR_RENT	(360,450]	0.1904	0.4020
MOR_RENT	(450,577]	-0.4461	0.1547
MOR_RENT	(577,900]	-0.6067	0.0535
MOR_RENT	(900,1.3e+0]	0.7644	0.0137
AGE	(27,31]	-0.3390	0.2257
AGE	(31,34]	0.1946	0.4228
AGE	(34,37]	0.0374	0.8876
AGE	(37,41]	0.1343	0.5592
AGE	(41,45]	-0.1658	0.5477
AGE	(45,50]	-0.0228	0.9327
AGE	(50,57]	-0.9720	0.0138
AGE	(57,64]	0.5020	0.3725
SMO89	0	-0.9325	0.1277
SNW12TV	0	-1.0634	0.0365
SWRSTCUR	0	-0.0175	0.9131
SWRSTCUR	EMP	0.6050	0.0202
SWRSTCUR	N	0.1785	0.5623

Table A.7: Maximum likelihood estimates of parameters by Logistic Regression after stepwised selection based on 3209 cases in band 50.

Variable	Dummy Variable	MLE estimates	Pr > ChiSq
Intercept		-3.5447	< .0001
Rate		10.7774	0.0004
Insurance	N	-0.6322	< .0001
LOAN_AMT	(1.2e+04,1.5e+04]	-0.2795	0.0852
LOAN_AMT	(1.5e+04,2.3e+04]	0.2446	0.1512
LOAN_AMT	(1e+04,1.2e+04]	-0.5738	0.0122
LOAN_AMT	(2.3e+04,2.5e+04]	0.6144	0.0027
LOAN_AMT	(5e+03,6.5e+03]	0.0469	0.7776
LOAN_AMT	(6.5e+03,8e+03]	0.0765	0.6018
LOAN_AMT	(8e+03,1e+04]	0.0304	0.8090
TERM	24	-0.4929	0.0230
TERM	36	-0.0160	0.8973
TERM	48	-0.1645	0.2216
TERM	60	0.0559	0.6099
TERM	72	0.0376	0.8964
newbus	0	-0.1682	0.0224
ALCIFDET	EMP	0.3896	0.0275
CCJGT500	(22,27]	0.0719	0.5944
CCJGT500	(27,32]	0.3022	0.0125
CCJGT500	(32,37]	-0.2112	0.1516
CCJGT500	(37,43]	0.1845	0.1587
CCJGT500	(43,58]	-0.1693	0.2334
CCJGT500	EMPTY	-0.2633	0.0187
NETINCM	(1.05e+03,1.2e+03]	-0.2688	0.0597
NETINCM	(1.2e+03,1.38e+03]	-0.1341	0.3941
NETINCM	(1.38e+03,1.55e+03]	-0.0128	0.9278
NETINCM	(1.55e+03,1.8e+03]	-0.1377	0.3445
NETINCM	(1.8e+03,2.2e+03]	-0.1196	0.4385
NETINCM	(2.2e+03,3.5e+03]	0.3746	0.0066
NETINCM	(3.5e+03,9.37e+05]	0.8909	< .0001
NETINCM	(900,1.05e+03]	-0.2922	0.0725
SOCSETT	EMPTY	-0.4652	0.0166
AGE	(27,31]	0.3381	0.0227
AGE	(31,34]	0.0145	0.9281
AGE	(34,37]	0.0199	0.8943
AGE	(37,41]	0.00469	0.9703
AGE	(41,45]	-0.3292	0.0261
AGE	(45,50]	-0.1417	0.2925
AGE	(50,57]	-0.4148	0.0036
AGE	(57,64]	0.2017	0.3273

Table A.8: Maximum likelihood estimates of parameters by Logistic Regression after stepwised selection based on 15766 cases in band 60. To be continued.

Variable	Dummy Variable	MLE estimates	Pr > ChiSq
SPL6MACT	0	-0.1774	0.2479
SPL6MACT	1	0.3217	0.0517
SPL6MACT	EMPTY	0.9194	0.0117
SPL6MACT	N	-0.8553	0.0148
SSRC4TO6	0	-0.5930	< .0001
SSRC4TO6	1	-0.3222	0.0511
SSRC4TO6	2	-0.0914	0.6666
SSRC4TO6	3	0.1686	0.6243
SWRSTCUR	0	-0.1876	0.1089
SWRSTCUR	EMP	0.5766	0.0030
SWRSTCUR	N	0	.

Table A.9: Following table A.8. Maximum likelihood estimates of parameters by Logistic Regression after stepwised selection based on 15766 cases in band 60.

Variable	Dummy Variable	MLE estimates	Pr > ChiSq
Intercept		-0.7224	0.0583
Insurance	N	-0.4457	0.0008
brand	dlfs	0.2855	0.0370
internet	0	0.3357	0.0193
AGE	(27,31]	-0.1139	0.7402
AGE	(31,34]	-0.3634	0.3786
AGE	(34,37]	-0.3644	0.4792
AGE	(37,41]	0.9108	0.0036
AGE	(41,45]	0.8991	0.0101
AGE	(45,50]	0.0297	0.9446
AGE	(50,57]	-0.5848	0.2725
AGE	(57,64]	-0.6786	0.5134
SVALCAIS	0	-0.7585	0.0267
WORST12	0	-0.5096	0.0286
WORST12	1	-0.1522	0.3961

Table A.10: Maximum likelihood estimates of parameters by Logistic Regression after stepwised selection based on data of band 0 and 20

APR	Coef.	Std. Err.	t	$P>\|t\|$	[95% Conf.	Interval]
I_Default	0.101254	0.067156	1.51	0.132	-0.03037	0.232883
raw_apradj	1.036216	0.007658	135.31	0.000	1.021206	1.051226
loan_amt6 03	-0.36626	0.033634	-10.89	0.000	-0.43218	-0.30033
internet1	-0.51071	0.019706	-25.92	0.000	-0.54934	-0.47209
cpiy	0.349419	0.044769	7.8	0.000	0.261669	0.437168
brandlomb	0.266039	0.020584	12.92	0.000	0.225694	0.306384
term	-0.01412	0.000703	-20.09	0.000	-0.0155	-0.01274
loan_amt8 04	-0.29979	0.031185	-9.61	0.000	-0.36092	-0.23867
inc_surp0395	0.342868	0.02919	11.75	0.000	0.285654	0.400081
loan_amt2 04	0.306278	0.053768	5.7	0.000	0.200891	0.411665
netincm10 03	-0.02061	0.025631	-0.8	0.421	-0.07084	0.02963
newbus1	-0.24536	0.028024	-8.76	0.000	-0.30029	-0.19043
loan_amt8 03	1.191471	0.032696	36.44	0.000	1.127385	1.255556
socsett02	-0.34227	0.05138	-6.66	0.000	-0.44298	-0.24157
spl6m12n	0.142824	0.035236	4.05	0.000	0.073759	0.211888
tosettl1em y	-0.0362	0.02726	-1.33	0.184	-0.08963	0.01723
wrst46al20	-0.19312	0.041188	-4.69	0.000	-0.27385	-0.11239
timadd11100	0.11173	0.02811	3.97	0.000	0.056634	0.166827
spl6mact00	-0.01198	0.020982	-0.57	0.568	-0.0531	0.029149
tosettl2y	-0.07268	0.02774	-2.62	0.009	-0.12705	-0.01831
inc_surp17 3	-0.12677	0.027802	-4.56	0.000	-0.18126	-0.07228
inc_s 726910	-0.02879	0.027662	-1.04	0.298	-0.08301	0.025428
spl6m12u	-0.10446	0.033068	-3.16	0.002	-0.16928	-0.03965
timeb 500900	0.065156	0.029057	2.24	0.025	0.008204	0.122108
timebank2e 3	0.017199	0.038602	0.45	0.656	-0.05846	0.092861
ssrc4to610	-0.11954	0.060767	-1.97	0.049	-0.23864	-0.00043
loan_amt5 03	-0.40821	0.036762	-11.1	0.000	-0.48026	-0.33615
loanbal14 05	-0.02214	0.026868	-0.82	0.410	-0.0748	0.030523
loan_amt12 4	-0.11312	0.034235	-3.3	0.001	-0.18022	-0.04601
ssrc4to600	-0.14718	0.060213	-2.44	0.015	-0.2652	-0.02916
ssrc4to620	-0.1125	0.065111	-1.73	0.084	-0.24012	0.015121
mor_rent0156	-0.02571	0.022109	-1.16	0.245	-0.06905	0.017621
swrstcurem y	0.053772	0.055867	0.96	0.336	-0.05573	0.163272
sncais3mem y	-0.03702	0.058797	-0.63	0.529	-0.15226	0.078224
noopen601	-0.00557	0.02096	-0.27	0.790	-0.04666	0.035508
_cons	8.405523	0.179857	46.73	0.000	8.052997	8.758048

Table A.11: Equation for APR in Simultaneous Equations with apr adjustment variable

Default	Coef.	Std. Err.	z	$P > \|z\|$	[95% Conf.	Interval]
I_APR	0.027213	0.018953	1.440	0.151	-0.00993	0.06436
raw_apradj	0.063977	0.020553	3.110	0.002	0.023694	0.104261
cpiy	0.560055	0.031586	17.730	0.000	0.498147	0.621962
wrst46al30	0.194317	0.04471	4.350	0.000	0.106687	0.281947
socworst00	-0.21736	0.039904	-5.450	0.000	-0.29557	-0.13915
loan_amt2 04	0.265667	0.06295	4.220	0.000	0.142288	0.389045
spvaldel00	-0.33224	0.06417	-5.180	0.000	-0.45801	-0.20647
age3741	0.00492	0.038709	0.130	0.899	-0.07095	0.080788
wrst46al20	-0.10458	0.070333	-1.490	0.137	-0.24243	0.033272
loanbal17 06	0.043432	0.040286	1.080	0.281	-0.03553	0.122391
spsetld	-0.0323	0.005034	-6.420	0.000	-0.04217	-0.02244
tosettl6n	0.23468	0.05033	4.660	0.000	0.136034	0.333326
inc_surp13 3	-0.17994	0.048165	-3.740	0.000	-0.27434	-0.08554
timeb 500900	0.10139	0.039015	2.600	0.009	0.024922	0.177858
ssrc4to6em y	0.375853	0.0948	3.960	0.000	0.190048	0.561658
swrstcurem y	0.236494	0.059098	4.000	0.000	0.120665	0.352324
term	0.004485	0.001008	4.450	0.000	0.00251	0.00646
mor_r 156227	-0.1126	0.042278	-2.660	0.008	-0.19547	-0.02974
_cons	-2.06113	0.178601	-11.540	0.000	-2.41118	-1.71108

Table A.12: Equation for Default in Simultaneous Equations with apr adjustment variable

APR	Coef.	Std. Err.	t	$P > \|t\|$	[95% Conf.	Interval]
I_Default	2.066874	0.169633	12.18	0.000	1.734386	2.399362
loan_amt6 03	-0.39192	0.117188	-3.34	0.001	-0.62162	-0.16223
internet1	-0.11097	0.068446	-1.62	0.105	-0.24512	0.023189
cpiy	-0.33139	0.129165	-2.57	0.010	-0.58456	-0.07822
brandlomb	0.272648	0.070521	3.87	0.000	0.134424	0.410872
term	-0.01247	0.002393	-5.21	0.000	-0.01716	-0.00778
loan_amt8 04	-0.19384	0.108511	-1.79	0.074	-0.40652	0.01885
inc_surp0395	0.289899	0.09066	3.2	0.001	0.112201	0.467596
loan_amt2 04	-0.13056	0.16374	-0.8	0.425	-0.4515	0.190379
netincm10 03	0.10031	0.086499	1.16	0.246	-0.06923	0.269852
newbus1	-0.79597	0.090501	-8.8	0.000	-0.97335	-0.61858
loan_amt8 03	1.534303	0.112716	13.61	0.000	1.313375	1.75523
socsett02	-0.53127	0.186071	-2.86	0.004	-0.89598	-0.16657
spl6m12n	0.410265	0.109917	3.73	0.000	0.194823	0.625706
tosettl1em y	-0.23293	0.091214	-2.55	0.011	-0.41171	-0.05414
wrst46al20	-0.0249	0.153433	-0.16	0.871	-0.32563	0.275838
timadd11100	0.456138	0.092864	4.91	0.000	0.274121	0.638155
spl6mact00	-0.20657	0.071755	-2.88	0.004	-0.34721	-0.06593
tosettl2y	0.21492	0.090498	2.37	0.018	0.03754	0.3923
inc_surp17 3	-0.31014	0.098731	-3.14	0.002	-0.50366	-0.11662
inc_s 726910	0.014688	0.09666	0.15	0.879	-0.17477	0.204145
spl6m12u	-0.02656	0.125598	-0.21	0.833	-0.27274	0.219619
timeb 500900	0.428248	0.092688	4.62	0.000	0.246576	0.609921
timebank2e 3	0.231935	0.15089	1.54	0.124	-0.06382	0.527686
ssrc4to610	-0.9958	0.175959	-5.66	0.000	-1.34068	-0.65091
loan_amt5 03	-0.24802	0.130183	-1.91	0.057	-0.50318	0.007142
loanbal14 05	0.072639	0.091784	0.79	0.429	-0.10726	0.252539
loan_amt12 4	0.009586	0.119498	0.08	0.936	-0.22463	0.243806
ssrc4to600	-1.18404	0.177833	-6.66	0.000	-1.5326	-0.83548
ssrc4to620	-0.8841	0.182831	-4.84	0.000	-1.24246	-0.52575
mor_rent0156	-0.08755	0.071996	-1.22	0.224	-0.22867	0.053562
swrstcurem y	0.802323	0.160225	5.01	0.000	0.488276	1.116369
sncais3mem y	0.56678	0.179544	3.16	0.002	0.214865	0.918694
noopen601	0.025002	0.070821	0.35	0.724	-0.11381	0.163813
_cons	14.19966	0.464848	30.55	0.000	13.28854	15.11078

Table A.13: Equation for APR in Simultaneous Equations without apr adjustment variable

Default	Coef.	Std. Err.	z	$P > \|z\|$	[95% Conf.	Interval]
I_APR	0.109739	0.012056	9.1	0.000	0.08611	0.133368
cpiy	0.493173	0.031372	15.72	0.000	0.431685	0.554661
wrst46al30	0.194647	0.044112	4.41	0.000	0.108189	0.281105
socworst00	-0.2434	0.038644	-6.3	0.000	-0.31914	-0.16766
loan_amt2 04	0.23279	0.061501	3.79	0.000	0.112249	0.35333
spvaldel00	-0.30918	0.062943	-4.91	0.000	-0.43255	-0.18581
age3741	0.003102	0.037798	0.08	0.935	-0.07098	0.077186
wrst46al20	-0.07969	0.069019	-1.15	0.248	-0.21497	0.055586
loanbal17 06	0.050535	0.039337	1.28	0.199	-0.02656	0.127634
spsetld	-0.03747	0.005069	-7.39	0.000	-0.04741	-0.02754
tosettl6n	0.245969	0.049513	4.97	0.000	0.148925	0.343012
inc_surp13 3	-0.15898	0.047131	-3.37	0.001	-0.25136	-0.06661
timeb 500900	0.101738	0.038764	2.62	0.009	0.025763	0.177714
ssrc4to6em y	0.324367	0.097269	3.33	0.001	0.133725	0.51501
swrstcurem y	0.205048	0.060343	3.4	0.001	0.086778	0.323317
term	0.006744	0.000864	7.8	0.000	0.00505	0.008438
mor_r 156227	-0.12054	0.041102	-2.93	0.003	-0.2011	-0.03998
_cons	-2.73474	0.139228	-19.64	0.000	-3.00762	-2.46186

Table A.14: Equation for Default in Simultaneous Equations without apr adjustment variable

Appendix B

Appendix for Chapter 4

Table B.1: Maximum Likelihood Estimates

Variable	Dummy Variable	DF	Estimate	Error	Chi-Square	Pr > ChiSq
Intercept		1	2.9725	0.2549	135.9394	< .0001
R1		1	-21.1886	0.4186	2562.605	< .0001
CPI	N	1	0.5154	0.014	1351.445	< .0001
LOAN_AMT	(1.2e+04,1.5e+04]	1	-0.2633	0.0334	62.1473	< .0001
LOAN_AMT	(1.5e+04,2.3e+04]	1	-0.4308	0.0408	111.6346	< .0001
LOAN_AMT	(1e+04,1.2e+04]	1	-0.1396	0.0396	12.4294	0.0004
LOAN_AMT	(2.3e+04,2.5e+04]	1	-0.4509	0.0604	55.6452	< .0001
LOAN_AMT	(5e+03,6.5e+03]	1	0.2872	0.0376	58.2157	< .0001
LOAN_AMT	(6.5e+03,8e+03]	1	0.0913	0.0325	7.886	0.005
LOAN_AMT	(8e+03,1e+04]	1	0.0436	0.0291	2.235	0.1349
TERM	24	1	-0.1228	0.0444	7.6367	0.0057
TERM	36	1	-0.1898	0.0289	43.1785	< .0001
TERM	48	1	-0.1234	0.0292	17.8485	< .0001
TERM	60	1	-0.0816	0.0242	11.394	0.0007
TERM	72	1	0.3115	0.0635	24.048	< .0001
brand	dlfs	1	-0.0788	0.014	31.797	< .0001
internet	0	1	0.1071	0.0155	47.8113	< .0001
newbus	0	1	0.2676	0.0258	107.7105	< .0001
CCJGT500	(22,27]	1	-0.0103	0.0327	0.1001	0.7517
CCJGT500	(27,32]	1	0.0406	0.0323	1.584	0.2082
CCJGT500	(32,37]	1	0.00434	0.0324	0.018	0.8933
CCJGT500	(37,43]	1	-0.064	0.0329	3.788	0.0516
CCJGT500	(43,58]	1	0.0271	0.0323	0.7047	0.4012
CCJGT500	EMPTY	1	-0.0646	0.0235	7.5343	0.0061
GDSCDE2	(111,222]	1	0.0342	0.0301	1.2946	0.2552

Table B.1: Maximum Likelihood Estimates(continued)

Variable	Dummy Variable	DF	Estimate	Error	Chi-Square	Pr > ChiSq
GDSCDE2	(222,999]	1	0.1634	0.03	29.5759	< .0001
GDSCDE3	0	1	-0.1238	0.0632	3.8295	0.0504
GDSCDE3	111	1	-0.0881	0.1123	0.6146	0.4331
GDSCDE3	200	1	0.1767	0.1454	1.477	0.2242
GDSCDE3	222	1	-0.3798	0.1444	6.9211	0.0085
LOANBAL1	(1.29e+04,6.49e+04]	1	-0.0575	0.0373	2.3734	0.1234
LOANBAL1	(1.37e+06,8.72e+06]	1	0.1974	0.058	11.5776	0.0007
LOANBAL1	(1.53e+05,3e+05]	1	0.037	0.0365	1.0265	0.311
LOANBAL1	(3e+05,4.89e+05]	1	-0.0412	0.0362	1.2949	0.2552
LOANBAL1	(4.89e+05,7.4e+05]	1	0.0187	0.0354	0.2788	0.5975
LOANBAL1	(6.49e+04,1.53e+05]	1	-0.0339	0.0362	0.8753	0.3495
LOANBAL1	(7.4e+05,1.37e+06]	1	0.0613	0.0366	2.8057	0.0939
LOANBAL2	(1.05e+04,6.27e+04]	1	-0.0622	0.0367	2.8648	0.0905
LOANBAL2	(1.05e+06,6.34e+06]	1	0.341	0.0572	35.5606	< .0001
LOANBAL2	(1.69e+05,3.68e+05]	1	-0.0435	0.0363	1.4383	0.2304
LOANBAL2	(3.68e+05,1.05e+06]	1	0.0566	0.038	2.2158	0.1366
LOANBAL2	(6.27e+04,1.69e+05]	1	-0.1402	0.0361	15.1271	0.0001
LOANBAL3	(1.59e+05,6.17e+05]	1	-0.0228	0.0376	0.3679	0.5442
LOANBAL3	(2.57e+04,1.59e+05]	1	0.0131	0.0365	0.1286	0.7199
LOANBAL3	(6.17e+05,7.29e+06]	1	0.2809	0.0574	23.9684	< .0001
LOANBAL4	(2.2e+04,3.55e+05]	1	-0.043	0.0387	1.2335	0.2667
LOANBAL4	(3.55e+05,3.83e+06]	1	0.3096	0.0569	29.5751	< .0001
LOANBAL5	(1.98e+05,8.5e+06]	1	0.1078	0.0436	6.1054	0.0135
LOANBAL6	(8.91e+04,2.5e+06]	1	0.1002	0.0415	5.8325	0.0157
MORTBAL	(1.05e+05,1e+07]	1	0.0774	0.0465	2.7647	0.0964
MORTBAL	(4.8e+04,1.05e+05]	1	-0.152	0.0332	20.9423	< .0001
MOR_RENT	(156,227]	1	-0.0253	0.036	0.4956	0.4814
MOR_RENT	(227,300]	1	0.0166	0.0316	0.2743	0.6005
MOR_RENT	(300,360]	1	-0.0363	0.0379	0.9149	0.3388
MOR_RENT	(360,450]	1	-0.0396	0.0332	1.4219	0.2331
MOR_RENT	(450,577]	1	0.0551	0.0361	2.3296	0.1269
MOR_RENT	(577,900]	1	0.0553	0.0351	2.4755	0.1156
MOR_RENT	(900,1.3e+0	1	0.1081	0.062	3.0414	0.0812
NETINCM	(1.05e+03,1.2e+03]	1	0.0356	0.033	1.1637	0.2807
NETINCM	(1.2e+03,1.38e+03]	1	-0.00227	0.0371	0.0037	0.9513
NETINCM	(1.38e+03,1.55e+03]	1	-0.0122	0.0339	0.1302	0.7182
NETINCM	(1.55e+03,1.8e+03]	1	-0.0465	0.0328	2.0016	0.1571
NETINCM	(1.8e+03,2.2e+03]	1	-0.00261	0.0343	0.0058	0.9393
NETINCM	(2.2e+03,3.5e+03]	1	-0.0668	0.0366	3.3215	0.0684

Table B.1: Maximum Likelihood Estimates(continued)

Variable	Dummy Variable	DF	Estimate	Error	Chi-Square	Pr > ChiSq
NETINCM	(3.5e+03,9.37e+05]	1	-0.0401	0.0664	0.3647	0.5459
NETINCM	(900,1.05e+03]	1	-0.00159	0.0376	0.0018	0.9663
NOOPEN6	(0,1]	1	0.0597	0.0261	5.2263	0.0222
NOOPEN6	(1,2]	1	-0.0816	0.0348	5.4948	0.0191
NOOPEN6	(2,9]	1	-0.1013	0.0537	3.5597	0.0592
SNBALALL	(1,3]	1	-0.1565	0.0554	7.9719	0.0048
SNBALALL	(104,245]	1	0.0885	0.0403	4.8287	0.028
SNBALALL	(13,30]	1	-0.00009	0.04	0	0.9983
SNBALALL	(3,13]	1	-0.0454	0.0403	1.2689	0.26
SNBALALL	(30,57]	1	0.0321	0.0381	0.7121	0.3987
SNBALALL	(57,104]	1	0.0922	0.0376	6.0171	0.0142
SNBALALL	EMPTY	1	0.019	0.0757	0.0627	0.8023
SNBALL6M	(1,3]	1	0.3842	0.0751	26.1688	< .0001
SNBALL6M	(10,24]	1	0.0639	0.0705	0.821	0.3649
SNBALL6M	(133,738]	1	-0.4445	0.1165	14.5697	0.0001
SNBALL6M	(24,44]	1	-0.2622	0.0704	13.8736	0.0002
SNBALL6M	(3,10]	1	0.3925	0.0723	29.4695	< .0001
SNBALL6M	(44,72]	1	-0.0867	0.0695	1.5569	0.2121
SNBALL6M	(72,133]	1	-0.3528	0.0712	24.5532	< .0001
SOCBAL	EMPTY	1	-0.1128	0.0422	7.1575	0.0075
SOCSETT	EMPTY	1	0.5024	0.0528	90.408	< .0001
TIMADD1	(1.2e+03,1.71e+03]	1	-0.00408	0.0354	0.0133	0.9083
TIMADD1	(1.71e+03,2.61e+03]	1	0.1317	0.0369	12.7426	0.0004
TIMADD1	(100,200]	1	-0.1731	0.0361	22.9409	< .0001
TIMADD1	(2.61e+03,5.9e+03]	1	0.275	0.0596	21.2978	< .0001
TIMADD1	(200,306]	1	-0.0741	0.0346	4.5809	0.0323
TIMADD1	(306,506]	1	-0.1068	0.0333	10.3107	0.0013
TIMADD1	(506,800]	1	0.0212	0.0337	0.3933	0.5306
TIMADD1	(800,1.2e+03]	1	0.0316	0.0343	0.8508	0.3563
TOSETTL1	EMPTY	1	-0.1002	0.0436	5.2795	0.0216
TOSETTL1	N	1	-0.0758	0.0243	9.7162	0.0018
TOSETTL2	EMPTY	1	-0.0158	0.0404	0.154	0.6948
TOSETTL2	N	1	-0.0889	0.0252	12.4507	0.0004
TOSETTL3	EMPTY	1	0.0209	0.0384	0.2962	0.5863
TOSETTL3	N	1	-0.098	0.0281	12.1829	0.0005
TOSETTL4	EMPTY	1	0.1475	0.0456	10.4613	0.0012
TOSETTL4	N	1	-0.1257	0.0339	13.7455	0.0002
AGE	(27,31]	1	-0.0703	0.036	3.8093	0.051
AGE	(31,34]	1	-0.0597	0.0368	2.6343	0.1046

Table B.1: Maximum Likelihood Estimates(continued)

Variable	Dummy Variable	DF	Estimate	Error	Chi-Square	Pr > ChiSq
AGE	(34,37]	1	-0.00425	0.0363	0.0137	0.9067
AGE	(37,41]	1	0.082	0.0331	6.1289	0.0133
AGE	(41,45]	1	0.0689	0.035	3.8729	0.0491
AGE	(45,50]	1	0.0861	0.0363	5.6392	0.0176
AGE	(50,57]	1	0.0437	0.0376	1.3512	0.2451
AGE	(57,64]	1	0.0692	0.069	1.005	0.3161
TIMEBANK	(1.1e+03,1.41e+03]	1	0.0583	0.0363	2.5787	0.1083
TIMEBANK	(1.41e+03,1.61e+03]	1	0.0182	0.0343	0.2803	0.5965
TIMEBANK	(1.61e+03,2e+03]	1	-0.00793	0.0292	0.0739	0.7857
TIMEBANK	(2.4e+03,3e+03]	1	0.013	0.0367	0.1257	0.7229
TIMEBANK	(2e+03,2.4e+03]	1	0.0439	0.0492	0.7988	0.3714
TIMEBANK	(3e+03,8.2e+03]	1	0.1358	0.0612	4.9196	0.0266
TIMEBANK	(500,900]	1	-0.0471	0.0345	1.8585	0.1728
TIMEBANK	(900,1.1e+03]	1	-0.0821	0.0342	5.7612	0.0164
NO_STORE	0	1	-0.3339	0.1944	2.952	0.0858
NO_STORE	1	1	-0.1436	0.1986	0.5233	0.4694
NO_STORE	2	1	-0.1791	0.2242	0.6381	0.4244
NO_STORE	3	1	-0.0123	0.2997	0.0017	0.9673
SNRECACT	1	1	1.3558	0.4147	10.6873	0.0011
SPL6M12	0	1	-0.1768	0.0677	6.8184	0.009
SPL6M12	1	1	0.0199	0.0891	0.0497	0.8236
SPL6M12	EMPTY	1	0.3056	0.2471	1.5294	0.2162
SPL6M12	N	1	0.1596	0.0797	4.012	0.0452
SPL6MACT	0	1	-0.2741	0.1783	2.3647	0.1241
SPL6MACT	1	1	-0.6286	0.1854	11.4968	0.0007
SPL6MACT	EMPTY	1	-0.5415	0.2816	3.6988	0.0545
SPL6MACT	N	1	1.8661	0.6691	7.7769	0.0053
SPSETLD	0	1	0.1615	0.0379	18.1829	< .0001
SPSETLD	1	1	0.122	0.0341	12.7568	0.0004
SPSETLD	2	1	0.0875	0.0333	6.8885	0.0087
SPSETLD	3	1	-0.0178	0.0341	0.272	0.602
SPSETLD	4	1	0.0227	0.0354	0.4105	0.5217
SPSETLD	5	1	0.0144	0.0388	0.1384	0.7098
SPSETLD	6	1	-0.0573	0.0424	1.8267	0.1765
SPSETLD	7	1	-0.0171	0.0484	0.1249	0.7238
SPSETLD	8	1	-0.126	0.053	5.6409	0.0175
SPVALDEL	-1	1	0.0367	0.2379	0.0238	0.8773
SPVALDEL	0	1	-0.2175	0.2381	0.8343	0.361
SPVALDEL	1	1	0.0872	0.2462	0.1255	0.7231

Default	Coef.	Std. Err.	z	$P > \|z\|$	[95% Conf.	Interval]
loanapr1	0.3117450	0.0213706	14.59	0.000	0.2698593	0.3536306
cpi	0.5949605	0.0365988	16.26	0.000	0.5232281	0.6666930
wrst46al	0.5469332	0.0802858	6.81	0.000	0.3895758	0.7042906
timebank	0.3395312	0.0491891	6.90	0.000	0.2431223	0.4359402
ssrc4to6	0.3364142	0.0692591	4.86	0.000	0.2006688	0.4721596
socworst	0.3045035	0.0567139	5.37	0.000	0.1933464	0.4156607
loanbal2	-0.6521232	0.1653033	-3.95	0.000	-0.9761117	-0.3281347
loanbal6	-1.5275050	0.4034007	-3.79	0.000	-2.3181550	-0.7368537
spsetld	0.3421178	0.0889288	3.85	0.000	0.1678205	0.5164151
term	0.9258050	0.2473159	3.74	0.000	0.4410748	1.4105350
netincm	-1.2971430	0.3497322	-3.71	0.000	-1.9826060	-0.6116808
_cons	-1.6875550	0.0169933	-99.31	0.000	-1.7208610	-1.6542490

Table B.2: Probit Default Model with variables stepwise-selected with significance value of 0.002

Default	Coef.	Std. Err.	z	$P > \|z\|$	[95% Conf.	Interval]
loanapr1	0.3004062	0.0220364	13.63	0.000	0.2572157	0.3435967
cpi	0.6118196	0.0371333	16.48	0.000	0.5390396	0.6845996
wrst46al	0.3580667	0.1086589	3.30	0.001	0.1450991	0.5710342
timebank	0.2679799	0.0545948	4.91	0.000	0.1609760	0.3749838
ssrc4to6	0.3153274	0.0705834	4.47	0.000	0.1769864	0.4536683
socworst	0.2569090	0.0575867	4.46	0.000	0.1440411	0.3697769
loanbal2	-0.7899687	0.1899592	-4.16	0.000	-1.1622820	-0.4176556
loanbal6	-1.4782630	0.4102631	-3.60	0.000	-2.2823640	-0.6741616
spsetld	0.3097557	0.0905687	3.42	0.001	0.1322443	0.4872671
term	0.8046459	0.2570356	3.13	0.002	0.3008653	1.3084260
netincm	-1.3180350	0.3558251	-3.70	0.000	-2.0154390	-0.6206306
alcifdet	0.6065010	0.1958156	3.10	0.002	0.2227094	0.9902925
age	0.1709128	0.0533851	3.20	0.001	0.0662799	0.2755457
worst12	0.3700065	0.1217338	3.04	0.002	0.1314125	0.6086004
spl6m12	0.2763015	0.1052215	2.63	0.009	0.0700711	0.4825319
loan_amt	0.4047468	0.1516572	2.67	0.008	0.1075041	0.7019895
tosettl2	0.7079504	0.3143855	2.25	0.024	0.0917661	1.3241350
socsett	0.2572794	0.1172814	2.19	0.028	0.0274120	0.4871468
ccjgt500	-1.0989980	0.5104424	-2.15	0.031	-2.0994470	-0.0985491
_cons	-1.6859410	0.0171999	-98.02	0.000	-1.7196530	-1.6522300

Table B.3: Probit Default Model with variables stepwise-selected with significance value of 0.05

Table B.4: Indifference Curve Logit

Acceptance	Coef.	Std. Err.	z	$P > \|z\|$	[95% Conf.	Interval]
log(APR)	-2.700994	0.0425410	-63.49	0.0000	-2.784373	-2.617615
insured_in r	-1.039617	0.0280186	-37.10	0.0000	-1.094533	-0.984702
_Inewbus_1	-0.457301	0.0477642	-9.57	0.0000	-0.550917	-0.363685
log(L)	-1.040575	0.0289465	-35.95	0.0000	-1.097309	-0.983841
raw_term	0.012003	0.0009291	12.92	0.0000	0.010182	0.013824
_Itosettl1_3	0.301926	0.0367722	8.21	0.0000	0.229854	0.373998
_Iloanbal3_4	-0.189716	0.0442803	-4.28	0.0000	-0.276504	-0.102928
_Igdscde2_2	0.213876	0.0561095	3.81	0.0000	0.103903	0.323849
_Isnball6m_2	-0.235106	0.0746118	-3.15	0.0020	-0.381343	-0.088870
_Isocsett_2	-0.940652	0.1027899	-9.15	0.0000	-1.142116	-0.739187
_Iloanbal4_2	0.349499	0.0816101	4.28	0.0000	0.189546	0.509452
_Itosettl2_3	0.216715	0.0465052	4.66	0.0000	0.125566	0.307863
_Ispl6m12_5	-0.221187	0.0375577	-5.89	0.0000	-0.294799	-0.147576
_Ibrand_2	0.187637	0.0281253	6.67	0.0000	0.132513	0.242762
_Iinternet_1	-0.217237	0.0295029	-7.36	0.0000	-0.275062	-0.159413
_Imortbal_2	-0.231636	0.0411156	-5.63	0.0000	-0.312221	-0.151051
_Ispsetld_9	-0.322049	0.0395030	-8.15	0.0000	-0.399474	-0.244625
_Iloanbal3_3	0.385128	0.0765311	5.03	0.0000	0.235130	0.535126
_Ispvaldel_4	-1.347730	0.5342820	-2.52	0.0120	-2.394904	-0.300557
_Iloanbal2_2	0.455738	0.0697915	6.53	0.0000	0.318949	0.592527
_Igdscde2_3	-0.250827	0.0441004	-5.69	0.0000	-0.337263	-0.164392
_Itimadd1_4	0.262248	0.0677303	3.87	0.0000	0.129499	0.394997
_Iloanbal1_8	-0.123057	0.0339084	-3.63	0.0000	-0.189516	-0.056598
_Iloanbal4_3	-0.197330	0.0558741	-3.53	0.0000	-0.306841	-0.087818
_Ispvaldel_3	0.417247	0.0915153	4.56	0.0000	0.237880	0.596613
_Igdscde 444	0.620898	0.1342691	4.62	0.0000	0.357735	0.884060
_Inoopen6_2	-0.167125	0.0428147	-3.90	0.0000	-0.251040	-0.083210
_Iloanbal2_4	0.189021	0.0448244	4.22	0.0000	0.101167	0.276875
_Isnball6m_6	-0.452902	0.0766276	-5.91	0.0000	-0.603089	-0.302715
_Iage_9	-0.116865	0.0394048	-2.97	0.0030	-0.194097	-0.039633
_Itosettl3_3	0.221846	0.0596110	3.72	0.0000	0.105010	0.338681
_Iloanbal6_2	-0.312895	0.0752194	-4.16	0.0000	-0.460322	-0.165468
_Iloanbal1_2	0.307311	0.0661903	4.64	0.0000	0.177580	0.437042
_Ispl6m12_4	0.386019	0.0573525	6.73	0.0000	0.273610	0.498428
_Isnrecact_2	-0.534180	0.0828098	-6.45	0.0000	-0.696484	-0.371876
_Ispl6m4_4	0.086398	0.0287109	3.01	0.0030	0.030126	0.142671
_Ino_store_1	0.180021	0.0633093	2.84	0.0040	0.055937	0.304105
_Ispl6m12_3	0.417612	0.1235109	3.38	0.0010	0.175536	0.659689
_Itimadd1_3	-0.175633	0.0427672	-4.11	0.0000	-0.259455	-0.091810

Variable	ln(APR)	ln(AMT)	ln(pc1)	ln(pc2)	ln(pc3)	ln(pc4)
Coefficient	-1.50774	-0.20073	0.026245	0.023633	0.032882	-0.02864
Variable	ln(pc5)	ln(pc6)	ln(pc7)	ln(pc8)	ln(pc9)	
Coefficient	-0.02315	0.00086	0.006294	0.001636	-0.00894	
Variable	ln(pc10)	ln(pc11)	ln(pc12)	ln(pc13)	ln(pc14)	
Coefficient	-0.01506	0.027577	0.008749	0.025294	0.017665	
Variable	ln(pc15)	ln(pc16)	ln(pc17)	ln(pc18)	_cons	
Coefficient	0.016929	0.013523	0.01228	-0.01172	-3.9224	

Table B.5: Parameters estimated from OLS regression

_Itimadd1_9	-0.168349	0.0398130	-4.23	0.0000	-0.246381	-0.090317
_Igdscde 200	0.524374	0.1676055	3.13	0.0020	0.195873	0.852875
_Ispsetld_8	-0.250975	0.0614355	-4.09	0.0000	-0.371386	-0.130564
_Isnbalall_2	0.216802	0.0467015	4.64	0.0000	0.125268	0.308335
_Isnball6m_7	-0.659762	0.0784589	-8.41	0.0000	-0.813538	-0.505985
_Isnball6m_3	-0.725873	0.1294780	-5.61	0.0000	-0.979645	-0.472101
_Isnball6m_4	-0.550400	0.0772546	-7.12	0.0000	-0.701816	-0.398984
_Ispsetld_7	-0.208056	0.0543322	-3.83	0.0000	-0.314545	-0.101567
_Ispsetld_3	-0.134976	0.0389790	-3.46	0.0010	-0.211373	-0.058579
_Itimadd1_5	-0.123718	0.0405609	-3.05	0.0020	-0.203216	-0.044220
_Itimadd1_6	-0.112501	0.0395957	-2.84	0.0040	-0.190107	-0.034895
_Isnbalall_6	0.148504	0.0434600	3.42	0.0010	0.063324	0.233684
_Iloanbal1_7	0.121340	0.0408825	2.97	0.0030	0.041212	0.201469
_Iinc_surp_7	-0.102884	0.0376997	-2.73	0.0060	-0.176774	-0.028994
_Ino_deps_4	0.100590	0.0291208	3.45	0.0010	0.043514	0.157666
_Isocworst_2	-0.680604	0.2735984	-2.49	0.0130	-1.216847	-0.144361
_Ispsetld_6	-0.133687	0.0488140	-2.74	0.0060	-0.229361	-0.038013
_Isnw12tv_2	0.147099	0.0584538	2.52	0.0120	0.032532	0.261667
_Isnbalall_5	0.091368	0.0421038	2.17	0.0300	0.008845	0.173890
_Ialcifdet_2	0.230821	0.1016956	2.27	0.0230	0.031501	0.430141
_Iage_4	0.118337	0.0381368	3.10	0.0020	0.043590	0.193084
_Iage_5	0.118641	0.0408234	2.91	0.0040	0.038629	0.198654
_Iage_6	0.116703	0.0413757	2.82	0.0050	0.035608	0.197798
_Iccjgt500_6	-0.063494	0.0275959	-2.30	0.0210	-0.117581	-0.009407
_Isocworst_4	-0.234336	0.1065833	-2.20	0.0280	-0.443235	-0.025436
_Itosettl4_2	-0.231257	0.0472866	-4.89	0.0000	-0.323937	-0.138577
_Imor_rent_8	-0.079177	0.0323495	-2.45	0.0140	-0.142580	-0.015773
_Imor_rent_4	-0.090851	0.0376629	-2.41	0.0160	-0.164669	-0.017034
_cons	-2.661112	0.2150567	-12.37	0.0000	-3.082615	-2.239609

Appendix C

Appendix for Chapter 5

Table C.1: Use time dependent covariates to test the PH assumption, group 1

_t	Coef.	Std. Err.	z	$P > \|z\|$
rh				
loanapr1	-0.625875	0.099315	-6.30	0.000
cpi	-0.623123	0.113440	-5.49	0.000
term	-0.023298	0.314821	-0.07	0.941
timebank	-0.159357	0.228213	-0.70	0.485
spl6m12	-0.599066	0.185356	-3.23	0.001
ssrc4to6	-0.084699	0.205879	-0.41	0.681
loanbal4	0.336511	0.302102	1.11	0.265
spsetld	-1.171431	0.235313	-4.98	0.000
spl6m4	-0.468656	0.117236	-4.00	0.000
age	-0.406709	0.106335	-3.82	0.000
loanbal1	-0.654438	0.138320	-4.73	0.000
timadd1	-0.662313	0.180766	-3.66	0.000
inc_surp	-0.315840	0.097031	-3.26	0.001
searches	-0.483789	0.212854	-2.27	0.023
spvaldel	-0.437298	0.120251	-3.64	0.000
newbus	19.954610	5.943964	3.36	0.001
loanbal2	-0.427326	0.167069	-2.56	0.011
ccjgt500	-0.571135	0.253296	-2.25	0.024
loanbal6	-0.284608	0.133869	-2.13	0.034
no_amex	-0.834241	0.395129	-2.11	0.035
mortbal	-0.546873	0.272527	-2.01	0.045
snball6m	-0.629072	0.357020	-1.76	0.078
t				
loanapr1	0.048288	0.047978	1.01	0.314
cpi	-0.052760	0.053240	-0.99	0.322
term	-0.459681	0.148602	-3.09	0.002
timebank	-0.163812	0.106824	-1.53	0.125
spl6m12	0.054944	0.088407	0.62	0.534
ssrc4to6	-0.168402	0.096270	-1.75	0.080
loanbal4	-0.413371	0.135136	-3.06	0.002
spsetld	0.331755	0.115104	2.88	0.004

Table C.2: Use time dependent covariates to test the PH assumption, group 2

_t	Coef.	Std. Err.	z	$P > \|z\|$
rh				
loanapr1	-0.534217	0.035139	-15.20	0.000
cpi	-0.727654	0.038817	-18.75	0.000
term	-0.933334	0.109403	-8.53	0.000
timebank	-0.482008	0.084016	-5.74	0.000
spl6m12	-0.487189	0.065384	-7.45	0.000
ssrc4to6	-0.421630	0.070008	-6.02	0.000
loanbal4	-0.511494	0.109147	-4.69	0.000
spsetld	-0.543634	0.087728	-6.20	0.000
spl6m4	0.181920	0.347013	0.52	0.600
age	-0.662490	0.272898	-2.43	0.015
loanbal1	0.011567	0.393104	0.03	0.977
timadd1	-0.181053	0.518125	-0.35	0.727
inc_surp	0.464560	0.278560	1.67	0.095
searches	-0.919852	0.569728	-1.61	0.106
spvaldel	-0.595604	0.316545	-1.88	0.060
newbus	62.724010	18.726890	3.35	0.001
loanbal2	-0.414160	0.166981	-2.48	0.013
ccjgt500	-0.552871	0.253251	-2.18	0.029
loanbal6	-0.297256	0.133713	-2.22	0.026
no_amex	-0.774748	0.398277	-1.95	0.052
mortbal	-0.569620	0.272538	-2.09	0.037
snball6m	-0.630163	0.356810	-1.77	0.077
t				
spl6m4	-0.337476	0.164042	-2.06	0.040
age	0.133352	0.130695	1.02	0.308
loanbal1	-0.343889	0.181671	-1.89	0.058
timadd1	-0.235255	0.244650	-0.96	0.336
inc_surp	-0.396712	0.131744	-3.01	0.003
searches	0.224944	0.266081	0.85	0.398
spvaldel	0.085177	0.157460	0.54	0.589
newbus	-20.898330	8.402471	-2.49	0.013

Table C.3: Use time dependent covariates to test the PH assumption, group 3

_t	Coef.	Std. Err.	z	$P > \|z\|$
rh				
loanapr1	-0.535118	0.035267	-15.17	0.000
cpi	-0.727636	0.038858	-18.73	0.000
term	-0.933955	0.109345	-8.54	0.000
timebank	-0.478003	0.084029	-5.69	0.000
spl6m12	-0.485434	0.065423	-7.42	0.000
ssrc4to6	-0.420806	0.070154	-6.00	0.000
loanbal4	-0.519484	0.109189	-4.76	0.000
spsetld	-0.540068	0.087665	-6.16	0.000
spl6m4	-0.455924	0.117520	-3.88	0.000
age	-0.399045	0.106442	-3.75	0.000
loanbal1	-0.661577	0.138228	-4.79	0.000
timadd1	-0.647672	0.180554	-3.59	0.000
inc_surp	-0.314581	0.096943	-3.25	0.001
searches	-0.492646	0.212988	-2.31	0.021
spvaldel	-0.418671	0.119961	-3.49	0.000
newbus	19.862720	5.939985	3.34	0.001
loanbal2	1.287643	0.462279	2.79	0.005
ccjgt500	-0.377414	0.671179	-0.56	0.574
loanbal6	0.279227	0.422757	0.66	0.509
no_amex	5.281398	2.714625	1.95	0.052
mortbal	-1.645832	0.823706	-2.00	0.046
snball6m	0.266812	1.050864	0.25	0.800
t				
loanbal2	-0.846953	0.209700	-4.04	0.000
ccjgt500	-0.091807	0.312159	-0.29	0.769
loanbal6	-0.270978	0.187536	-1.44	0.148
no_amex	-2.624412	1.096511	-2.39	0.017
mortbal	0.544655	0.383709	1.42	0.156
snball6m	-0.445859	0.496056	-0.90	0.369

Table C.4: Cox with time dependent covariates

_t	Coef.	Std. Err.	z	$P > \|z\|$
rh				
loanapr1	-0.533234	0.035223	-15.14	0.000
cpi	-0.730704	0.038830	-18.82	0.000
term	-0.461500	0.209891	-2.20	0.028
timebank	-0.482071	0.084065	-5.73	0.000
spl6m12	-0.480955	0.065497	-7.34	0.000
ssrc4to6	-0.412186	0.070195	-5.87	0.000
loanbal4	-0.512284	0.109195	-4.69	0.000
spsetld	-1.076827	0.157580	-6.83	0.000
spl6m4	-0.450722	0.117026	-3.85	0.000
age	-0.411820	0.106354	-3.87	0.000
loanbal1	-0.663202	0.138259	-4.80	0.000
timadd1	-0.648787	0.180652	-3.59	0.000
inc_surp	0.146136	0.185226	0.79	0.430
searches	-0.475920	0.212791	-2.24	0.025
spvaldel	-0.402630	0.119798	-3.36	0.001
newbus	37.129000	11.699730	3.17	0.002
loanbal2	-0.413006	0.167004	-2.47	0.013
ccjgt500	-0.569515	0.253463	-2.25	0.025
loanbal6	-0.293090	0.133676	-2.19	0.028
no_amex	-0.824011	0.401275	-2.05	0.040
mortbal	-0.560212	0.272526	-2.06	0.040
snball6m	-0.616047	0.357272	-1.72	0.085
t				
term	-0.052153	0.019884	-2.62	0.009
spsetld	0.063649	0.015800	4.03	0.000
inc_surp	-0.051735	0.017526	-2.95	0.003
newbus	-1.761389	1.009336	-1.75	0.081

Table C.5: Estimates from Weibull Model on paying back early

_t	Haz. Ratio	Std. Err.	z	$P>\|z\|$	[95% Conf.	Interval]
cpi	0.4783145	0.0303514	-11.62	0.00	0.4223773	0.5416596
age	0.5232625	0.0416314	-8.14	0.00	0.4477102	0.6115644
spsetld	0.3415869	0.0267512	-13.72	0.00	0.2929812	0.3982563
raw_loanapr1	1.0314310	0.0043383	7.36	0.00	1.0229630	1.0399690
netincm	0.7619679	0.0579986	-3.57	0.00	0.6563658	0.8845603
timadd1	0.5139337	0.0411061	-8.32	0.00	0.4393648	0.6011585
gdscde2	0.6077858	0.0362376	-8.35	0.00	0.5407542	0.6831265
spl6m4	0.5101184	0.0441318	-7.78	0.00	0.4305575	0.6043811
snbalall	0.6730424	0.0667519	-3.99	0.00	0.5541417	0.8174553
internet	0.6933863	0.0721658	-3.52	0.00	0.5654374	0.8502879
tosettl4	0.2312880	0.0711647	-4.76	0.00	0.1265459	0.4227253
inc_surp	0.6518582	0.0582058	-4.79	0.00	0.5472018	0.7765307
ssrc4to6	0.5848742	0.0602835	-5.20	0.00	0.4778905	0.7158080
L	0.9872349	0.0029864	-4.25	0.00	0.9813991	0.9931054
loanbal1	0.5213038	0.0721717	-4.71	0.00	0.3974172	0.6838096
socsett	0.7251686	0.0535427	-4.35	0.00	0.6274666	0.8380835
timebank	0.6892486	0.0629588	-4.07	0.00	0.5762668	0.8243814
mortbal	0.5907617	0.0928403	-3.35	0.00	0.4341531	0.8038624
tosettl6	0.5324472	0.1087102	-3.09	0.00	0.3568503	0.7944509
term	0.5396387	0.0983520	-3.38	0.00	0.3775436	0.7713280
sncais3m	0.4666930	0.1205424	-2.95	0.00	0.2813027	0.7742633
loanbal2	1.9358770	0.5224799	2.45	0.01	1.1406290	3.2855740
alcifdet	0.4239867	0.1454299	-2.50	0.01	0.2164637	0.8304613
wrstnrev	1.8849790	0.4676688	2.56	0.01	1.1591000	3.0654350
loanbal5	0.6761110	0.1118250	-2.37	0.02	0.4889158	0.9349793
mor_rent	0.7562349	0.0895109	-2.36	0.02	0.5996601	0.9536922
spl6m12	0.7174979	0.1205176	-1.98	0.05	0.5162320	0.9972324
no_amex	0.0080638	0.0199067	-1.95	0.05	0.0000639	1.0182390
tosettl3	0.5974401	0.1691200	-1.82	0.07	0.3430370	1.0405140
searches	0.5731514	0.1830554	-1.74	0.08	0.3064849	1.0718390
ccjgt500	0.7543656	0.1227745	-1.73	0.08	0.5483365	1.0378070
/ln_p	0.6049988	0.0112799	53.64	0.00	0.5828907	0.6271069
p	1.8312500	0.0206562			1.7912090	1.8721860
1/p	0.5460751	0.0061596			0.5341349	0.5582822

Table C.6: Estimates from Exponential Model on paying back early

_t	Haz. Ratio	Std. Err.	z	$P > \|z\|$	[95% Conf.	Interval]
raw_loanapr1	1.0317790	0.0042606	7.58	0.00	1.0234620	1.0401640
spsetld	0.3745639	0.0292505	-12.57	0.00	0.3214059	0.4365139
age	0.5604360	0.0444372	-7.30	0.00	0.4797708	0.6546636
cpi	0.5295871	0.0336188	-10.01	0.00	0.4676298	0.5997533
gdscde2	0.5946140	0.0353844	-8.74	0.00	0.5291534	0.6681725
netincm	0.7441330	0.0565428	-3.89	0.00	0.6411685	0.8636325
timadd1	0.5359900	0.0429057	-7.79	0.00	0.4581614	0.6270396
spl6m4	0.5524892	0.0477138	-6.87	0.00	0.4664583	0.6543872
snbalall	0.6802004	0.0675157	-3.88	0.00	0.5599482	0.8262775
brand	0.7202345	0.0408486	-5.79	0.00	0.6444621	0.8049158
tosettl4	0.2549042	0.0772423	-4.51	0.00	0.1407479	0.4616493
mortbal	0.5928098	0.0940554	-3.30	0.00	0.4343736	0.8090351
L	0.9882287	0.0029605	-3.95	0.00	0.9824433	0.9940481
ssrc4to6	0.6266343	0.0643042	-4.55	0.00	0.5124661	0.7662370
socsett	0.7445302	0.0553493	-3.97	0.00	0.6435806	0.8613143
loanbal1	0.5361562	0.0742230	-4.50	0.00	0.4087472	0.7032792
inc_surp	0.7056412	0.0627786	-3.92	0.00	0.5927279	0.8400643
tosettl6	0.5463024	0.1111075	-2.97	0.00	0.3667034	0.8138631
timebank	0.7141295	0.0651024	-3.69	0.00	0.5972810	0.8538377
term	0.5528161	0.1002790	-3.27	0.00	0.3874142	0.7888343
internet	0.7541911	0.0782761	-2.72	0.01	0.6153707	0.9243278
alcifdet	0.4113670	0.1411012	-2.59	0.01	0.2100209	0.8057428
spl6m12	0.6635417	0.1117109	-2.44	0.02	0.4770499	0.9229383
sncais3m	0.5308436	0.1371757	-2.45	0.01	0.3198947	0.8808992
loanbal2	1.7458990	0.4702486	2.07	0.04	1.0297990	2.9599590
mor_rent	0.7579097	0.0896050	-2.34	0.02	0.6011501	0.9555469
loanbal5	0.6995682	0.1158180	-2.16	0.03	0.5057177	0.9677250
wrstnrev	1.6424840	0.4078645	2.00	0.05	1.0095530	2.6722250
searches	0.5529432	0.1763539	-1.86	0.06	0.2959383	1.0331420
no_amex	0.0138717	0.0335481	-1.77	0.08	0.0001212	1.5875410
ccjgt500	0.7530695	0.1224021	-1.74	0.08	0.5476244	1.0355890

Table C.7: Estimates from PH Cox on paying back early

_t	Coef.	Std. Err.	z	$P > \|z\|$	[95% Conf.	Interval]
raw_loanapr1	0.0331895	0.0041906	7.92	0.00	0.0249760	0.0414030
cpi	-0.7125868	0.0635259	-11.22	0.00	-0.8370953	-0.5880782
spsetld	-1.0471870	0.0782786	-13.38	0.00	-1.2006110	-0.8937642
age	-0.6291370	0.0794940	-7.91	0.00	-0.7849423	-0.4733317
netincm	-0.2752288	0.0761703	-3.61	0.00	-0.4245198	-0.1259377
timadd1	-0.6567771	0.0800183	-8.21	0.00	-0.8136102	-0.4999440
gdscde2	-0.5072920	0.0596893	-8.50	0.00	-0.6242808	-0.3903032
spl6m4	-0.6528448	0.0864778	-7.55	0.00	-0.8223382	-0.4833515
snbalall	-0.3993167	0.0991956	-4.03	0.00	-0.5937365	-0.2048969
tosettl4	-1.4370430	0.3076224	-4.67	0.00	-2.0399720	-0.8341147
internet	-0.3286043	0.1041256	-3.16	0.00	-0.5326867	-0.1245219
L	-0.0128302	0.0030263	-4.24	0.00	-0.0187616	-0.0068989
ssrc4to6	-0.5173966	0.1029967	-5.02	0.00	-0.7192663	-0.3155269
loanbal1	-0.6439720	0.1384485	-4.65	0.00	-0.9153261	-0.3726180
socsett	-0.3146527	0.0744291	-4.23	0.00	-0.4605310	-0.1687743
inc_surp	-0.4015945	0.0893214	-4.50	0.00	-0.5766613	-0.2265277
timebank	-0.3589438	0.0913087	-3.93	0.00	-0.5379055	-0.1799821
tosettl6	-0.6443927	0.2039511	-3.16	0.00	-1.0441300	-0.2446558
mortbal	-0.5263394	0.1589056	-3.31	0.00	-0.8377887	-0.2148902
term	-0.6171462	0.1820455	-3.39	0.00	-0.9739488	-0.2603435
sncais3m	-0.7243424	0.2584550	-2.80	0.01	-1.2309050	-0.2177799
alcifdet	-0.8783158	0.3430178	-2.56	0.01	-1.5506180	-0.2060133
loanbal2	0.6323793	0.2698190	2.34	0.02	0.1035438	1.1612150
mor_rent	-0.2811207	0.1184289	-2.37	0.02	-0.5132372	-0.0490043
wrstnrev	0.5944406	0.2483537	2.39	0.02	0.1076763	1.0812050
loanbal5	-0.3743662	0.1654424	-2.26	0.02	-0.6986274	-0.0501050
spl6m12	-0.3573855	0.1681669	-2.13	0.03	-0.6869865	-0.0277845
brand	-0.1197260	0.0582199	-2.06	0.04	-0.2338348	-0.0056171
no_amex	-4.6559620	2.4409750	-1.91	0.06	-9.4401850	0.1282598
searches	-0.5701334	0.3192294	-1.79	0.07	-1.1958110	0.0555447
ccjgt500	-0.2835569	0.1627574	-1.74	0.08	-0.6025556	0.0354417
tosettl3	-0.4766975	0.2832915	-1.68	0.09	-1.0319390	0.0785437

Table C.8: Lunn and McNeil Method A

_t	Coef.	Robust Std. Err.	z	$P > \|z\|$
type	1.7342980	0.5030923	3.45	0.001
raw_loanapr1	0.1056094	0.0083677	12.62	0.000
_Icpi_2	1.3667740	0.0715040	19.11	0.000

logLXAPR	0.1939021	0.0461542	4.20	0.000
_Ispl6m12_4	0.5227119	0.0991498	5.27	0.000
_Iloanbal4_3	-0.4544236	0.1158050	-3.92	0.000
_Inewbus_1	0.0233963	0.2779477	0.08	0.933
_Ispvaldel_2	-0.5375315	0.1419502	-3.79	0.000
_Iinc_surp_6	0.4307904	0.1221332	3.53	0.000
_Ispl6m4_3	1.3169650	0.4578020	2.88	0.004
_Iinc_surp_4	0.3579988	0.1166776	3.07	0.002
_Issrc4to6_5	0.7878295	0.1760189	4.48	0.000
_Iwrst46al_4	0.3899139	0.1012480	3.85	0.000
raw_term	0.0116839	0.0024762	4.72	0.000
_Iage_7	-0.8047942	0.2078096	-3.87	0.000
_Itimebank_4	-0.4823400	0.1266012	-3.81	0.000
_Itimebank_5	-0.9284479	0.2371845	-3.91	0.000
_Igdscde 999	-0.4156502	0.1392311	-2.99	0.003
_Isnball6m_8	0.3435823	0.1121660	3.06	0.002
_Issrc4to6_4	0.5870922	0.1741699	3.37	0.001
_Isearches_7	-0.1983133	0.0690195	-2.87	0.004
_Iloanbal6_2	-0.2901753	0.1562017	-1.86	0.063
Igdscde2 1	-0.1791319	0.1022374	-1.75	0.080
_Issrc4to6_3	0.2682160	0.1207071	2.22	0.026
_Ispl6m12_5	-0.2523361	0.1345237	-1.88	0.061
_Itimadd1_6	0.3422099	0.1095988	3.12	0.002
_Itimebank_2	-0.3048706	0.1086342	-2.81	0.005
_Itimebank_3	-0.3874635	0.1353873	-2.86	0.004
_Iwrstnrev_3	0.5020158	0.2801753	1.79	0.073
_Iinc_surp_2	-0.1834805	0.1116006	-1.64	0.100
_Iinternet_1	-0.2412311	0.0791869	-3.05	0.002
_Imortbal_3	-0.1948998	0.1088114	-1.79	0.073
_Itimebank_7	-0.2025196	0.1128873	-1.79	0.073
_Iloanbal1_7	-0.2349395	0.1132119	-2.08	0.038
_Itimadd1_7	-0.0889093	0.1253448	-0.71	0.478
_Issrc4to6_2	0.0976311	0.0852790	1.14	0.252
_Ispl6m4_4	-0.0685550	0.0748325	-0.92	0.360
_Ispsetld_9	-0.2775613	0.1153120	-2.41	0.016
_Imortbal_2	-0.1187518	0.1152082	-1.03	0.303
_Iinc_surp_3	-0.2062980	0.1721408	-1.20	0.231
_Iage_6	-0.4211983	0.1573720	-2.68	0.007
_Isocsett_3	-0.0456692	0.1302752	-0.35	0.726
_Iage_5	-0.4991570	0.1533945	-3.25	0.001
_Iage_4	-0.3713094	0.1357683	-2.73	0.006
_Iloanbal5_2	-0.0115215	0.1655331	-0.07	0.945

_Itosettl1_3	-0.0359408	0.0838600	-0.43	0.668
L	-0.0094526	0.0123491	-0.77	0.444
_Ispsetld_7	0.0632302	0.1381898	0.46	0.647
_Ispsetld_8	-0.4357232	0.1868123	-2.33	0.020
Igdscde2 4	0.1183438	0.1193816	0.99	0.322
_Ispsetld_5	-0.2331780	0.1327580	-1.76	0.079
_Iage_3	-0.3544368	0.1409809	-2.51	0.012
_Imor_rent_8	0.0845732	0.0815182	1.04	0.300
_Itosettl4_3	-0.1738918	0.1559771	-1.11	0.265
_Igdscde 333	0.0955900	0.2348779	0.41	0.684
_Iloanbal1_3	-0.0289744	0.1036110	-0.28	0.780
_Iage_2	-0.2965488	0.1314902	-2.26	0.024
_Itimadd1_2	0.0706656	0.1140131	0.62	0.535
_Ispsetld_6	-0.2163748	0.1426030	-1.52	0.129
_Imor_rent_5	-0.0912649	0.1234983	-0.74	0.460
_Ispsetld_4	-0.1386132	0.1101876	-1.26	0.208
_Iloanbal3_4	-0.0206867	0.1079674	-0.19	0.848
_Itimebank_6	-0.0836683	0.1112596	-0.75	0.452
_Isocworst_4	0.0958675	0.2386477	0.40	0.688
_Itimadd1_4	0.2498613	0.1605841	1.56	0.120
_Isnbalall_4	0.1268567	0.1028627	1.23	0.217
_Iloanbal1_2	-0.0647956	0.1077581	-0.60	0.548
_Isocworst_3	0.8537084	0.3504307	2.44	0.015
_Inetincm_6	0.4217224	0.1386253	3.04	0.002
_Inetincm_5	0.4093916	0.1114535	3.67	0.000
_Imor_rent_6	-0.0231949	0.1174985	-0.20	0.844
_Ino_store_1	-0.2885813	0.1492380	-1.93	0.053
_Inoopen6_4	-0.0973047	0.0814409	-1.19	0.232
_Itosettl3_3	0.1378312	0.1226310	1.12	0.261
_Ibrand_2	-0.2237849	0.0781054	-2.87	0.004
_Ispsetld_3	-0.0850318	0.1059072	-0.80	0.422
_Ialcifdet_2	-0.2552834	0.2326306	-1.10	0.272
_Iinc_surp_5	0.1233364	0.1048009	1.18	0.239
_Itimadd1_5	0.1842909	0.1161794	1.59	0.113
_Itimadd1_3	0.2003204	0.1195149	1.68	0.094
_Itimadd1_9	0.1767211	0.1153878	1.53	0.126
_Iccjgt500_3	-0.0339577	0.0805344	-0.42	0.673
_Isocbal_3	-0.7390479	0.2819599	-2.62	0.009
_Itosettl5_3	0.1163873	0.1742490	0.67	0.504
_Isnbalall_7	-0.0528594	0.1559737	-0.34	0.735
_Ino_store_3	0.2855969	0.5128436	0.56	0.578
_Isnball6m_3	0.0272360	0.2279906	0.12	0.905

_Imor_rent_7	-0.1266391	0.2101062	-0.60	0.547
_Inetincm_7	0.8471922	0.2167660	3.91	0.000
_Inetincm_4	0.2593257	0.1107841	2.34	0.019
covt_raw_l 1	-0.0692539	0.0091211	-7.59	0.000
covt_Icpi_2	-1.0237510	0.0774410	-13.22	0.000
covt_logLX R	-0.1686547	0.0503393	-3.35	0.001
covt_Is 2_4	-0.4759050	0.1085605	-4.38	0.000
covt_Il 4_3	0.5989644	0.1262211	4.75	0.000
covt_Inew 1	-0.1366773	0.2896905	-0.47	0.637
covt_Ispv 2	0.5681898	0.1621687	3.50	0.000
covt_Iinc 6	-0.3396654	0.1317687	-2.58	0.010
covt_Ispl 3	-0.9917178	0.5660166	-1.75	0.080
covt_Iinc 4	-0.2104749	0.1253308	-1.68	0.093
covt_Issr 5	-1.0647140	0.2352022	-4.53	0.000
covt_Iwrs 4	-0.4125193	0.1130420	-3.65	0.000
covt_raw_t m	-0.0129296	0.0026912	-4.80	0.000
covt_Iage_7	0.2440877	0.2241867	1.09	0.276
covt_It k_4	0.4268462	0.1373095	3.11	0.002
covt_It k_5	0.7928951	0.2517033	3.15	0.002
covt_Ig 999	0.6940515	0.1488359	4.66	0.000
covt_Isnb 8	-0.3465969	0.1198784	-2.89	0.004
covt_Issr 4	-0.3702454	0.1947617	-1.90	0.057
covt_Isea 7	0.2053970	0.0738522	2.78	0.005
covt_Il 6_2	0.4807359	0.1884277	2.55	0.011
covt_Ig 111	0.4297384	0.1095148	3.92	0.000
covt_Issr 3	-0.0702706	0.1294920	-0.54	0.587
covt_Ispl 5	0.1664340	0.1412285	1.18	0.239
covt_It 1_6	-0.1750487	0.1187132	-1.47	0.140
covt_It k_2	0.3344094	0.1183803	2.82	0.005
covt_It k_3	0.2809307	0.1467279	1.91	0.056
covt_Iwrs 3	-0.5894659	0.3390824	-1.74	0.082
covt_Iinc 2	0.2208191	0.1187255	1.86	0.063
covt_Iint 1	0.1145312	0.0854070	1.34	0.180
covt_Imor 3	0.1580248	0.1168032	1.35	0.176
covt_It k_7	0.1809508	0.1232679	1.47	0.142
covt_Iloa 7	0.2170166	0.1203995	1.80	0.071
covt_It 1_7	0.1487013	0.1326820	1.12	0.262
covt_Issr 2	-0.0859334	0.0919403	-0.93	0.350
covt_Is 4_4	-0.1307550	0.0803077	-1.63	0.103
covt_Isps 9	0.8394370	0.1237026	6.79	0.000
covt_Imor 2	0.2314466	0.1227568	1.89	0.059
covt_Iinc 3	-0.0020990	0.1884621	-0.01	0.991

covt_Iage_6	-0.1127577	0.1722607	-0.65	0.513
covt_I tt_3	-0.1192708	0.1381075	-0.86	0.388
covt_Iage_5	0.0132193	0.1667120	0.08	0.937
covt_Iage_4	-0.0299595	0.1484573	-0.20	0.840
covt_Il 5_2	0.1679286	0.1953128	0.86	0.390
covt_ tl1_3	0.1175862	0.0907403	1.30	0.195
covt_L	-0.0111954	0.0134883	-0.83	0.407
covt_Isps 7	0.3581577	0.1491349	2.40	0.016
covt_Isps 8	0.8449341	0.1976400	4.28	0.000
covt_Ig 444	0.1051353	0.1280935	0.82	0.412
covt_Isps 5	0.5250299	0.1409971	3.72	0.000
covt_Iage_3	0.0415660	0.1534832	0.27	0.787
covt_Imor 8	-0.1698609	0.0885236	-1.92	0.055
covt_It 4_3	0.4655888	0.1734935	2.68	0.007
covt_Ig 333	0.2856564	0.2530528	1.13	0.259
covt_Il 1_3	-0.1269867	0.1123709	-1.13	0.258
covt_Iage_2	0.0769281	0.1435471	0.54	0.592
covt_It 1_2	-0.1472258	0.1242221	-1.19	0.236
covt_Isps 6	0.4725071	0.1513875	3.12	0.002
covt_Imor 5	0.2693393	0.1300914	2.07	0.038
covt_Isps 4	0.3289960	0.1195445	2.75	0.006
covt_Iloa 4	0.1621553	0.1150019	1.41	0.159
covt_It k_6	0.1809682	0.1218586	1.49	0.138
covt_Isoc 4	0.1410744	0.2487152	0.57	0.571
covt_It 1_4	-0.4043452	0.1783746	-2.27	0.023
covt_Isnb 4	0.0222974	0.1089614	0.20	0.838
covt_Il 1_2	-0.0811965	0.1142729	-0.71	0.477
covt_I st_3	-0.5084404	0.3676866	-1.38	0.167
covt_Inet 6	-0.6371011	0.1485811	-4.29	0.000
covt_Inet 5	-0.5771261	0.1196705	-4.82	0.000
covt_Imor 6	0.1935330	0.1256494	1.54	0.123
covt_Ino_ 1	0.1163348	0.1608436	0.72	0.470
covt_Inoo 4	0.0097863	0.0877381	0.11	0.911
covt_It 3_3	0.0074453	0.1348584	0.06	0.956
covt_Ibra 2	0.3225824	0.0846638	3.81	0.000
covt_Isps 3	0.1989782	0.1150580	1.73	0.084
covt_Ialc 2	0.6233770	0.2589525	2.41	0.016
covt_Iinc 5	-0.0166874	0.1118376	-0.15	0.881
covt_It 1_5	0.0104013	0.1241804	0.08	0.933
covt_I d1_3	-0.0199015	0.1283966	-0.16	0.877
covt_Itim 9	-0.0070632	0.1237381	-0.06	0.954
covt_Iccj 3	0.1168845	0.0863685	1.35	0.176

covt_Is 1_3	0.5655174	0.2950834	1.92	0.055
covt_It 5_3	-0.3508563	0.2085312	-1.68	0.092
covt_Isnb 7	-0.1480735	0.1833198	-0.81	0.419
covt_Ino_ 3	-1.4763410	0.7444203	-1.98	0.047
covt_Isnb 3	-0.1979520	0.2433709	-0.81	0.416
covt_Imor 7	0.3564586	0.2266045	1.57	0.116
covt_Inet 7	-1.1392930	0.2439850	-4.67	0.000
covt_Inet 4	-0.3456502	0.1187396	-2.91	0.004

Table C.9: Lunn and McNeil Method B

_t	Coef.	Robust Std. Err.	z	$P > \|z\|$
raw_loanapr1	0.1018492	0.0077413	13.16	0.000
_Icpi_2	1.3401810	0.0699161	19.17	0.000
logLXAPR	0.1954782	0.0433310	4.51	0.000
_Ispl6m12_4	0.4978965	0.0943650	5.28	0.000
_Iloanbal4_3	-0.4581498	0.1113537	-4.11	0.000
_Inewbus_1	0.0036577	0.2699393	0.01	0.989
_Ispvaldel_2	-0.5210826	0.1364245	-3.82	0.000
_Iinc_surp_6	0.4246947	0.1187191	3.58	0.000
_Ispl6m4_3	1.1855470	0.4446374	2.67	0.008
_Iinc_surp_4	0.3515542	0.1134981	3.10	0.002
_Issrc4to6_5	0.7470454	0.1659709	4.50	0.000
_Iwrst46al_4	0.3696304	0.0968179	3.82	0.000
raw_term	0.0112665	0.0023777	4.74	0.000
_Iage_7	-0.7487323	0.2015493	-3.71	0.000
_Itimebank_4	-0.4690129	0.1228060	-3.82	0.000
_Itimebank_5	-0.9040819	0.2332254	-3.88	0.000
_Igdscde 999	-0.4125727	0.1365413	-3.02	0.003
_Isnball6m_8	0.3363813	0.1075902	3.13	0.002
_Issrc4to6_4	0.5602602	0.1654777	3.39	0.001
_Isearches_7	-0.1910413	0.0665926	-2.87	0.004
_Iloanbal6_2	-0.2767611	0.1494603	-1.85	0.064
Igdscde2 1	-0.1837640	0.0988876	-1.86	0.063
_Issrc4to6_3	0.2452768	0.1154466	2.12	0.034
_Ispl6m12_5	-0.2574459	0.1313369	-1.96	0.050
_Itimadd1_6	0.3142171	0.1063882	2.95	0.003
_Itimebank_2	-0.2875642	0.1043480	-2.76	0.006
_Itimebank_3	-0.3754827	0.1311226	-2.86	0.004
_Iwrstnrev_3	0.4528597	0.2651960	1.71	0.088
_Iinc_surp_2	-0.1793502	0.1083827	-1.65	0.098
_Iinternet_1	-0.2320473	0.0764745	-3.03	0.002

_Imortbal_3	-0.1815954	0.1057444	-1.72	0.086
_Itimebank_7	-0.1772049	0.1079026	-1.64	0.101
_Iloanbal1_7	-0.2300851	0.1101203	-2.09	0.037
_Itimadd1_7	-0.0945495	0.1204075	-0.79	0.432
_Issrc4to6_2	0.0978606	0.0824605	1.19	0.235
_Ispl6m4_4	-0.0585307	0.0726850	-0.81	0.421
_Ispsetld_9	-0.3136348	0.1119874	-2.80	0.005
_Imortbal_2	-0.1221811	0.1120054	-1.09	0.275
_Iinc_surp_3	-0.2057809	0.1653342	-1.24	0.213
_Iage_6	-0.3879375	0.1507492	-2.57	0.010
_Isocsett_3	-0.0564283	0.1252100	-0.45	0.652
_Iage_5	-0.4606872	0.1466367	-3.14	0.002
_Iage_4	-0.3325041	0.1290072	-2.58	0.010
_Iloanbal5_2	-0.0278742	0.1580253	-0.18	0.860
_Itosettl1_3	-0.0447790	0.0808796	-0.55	0.580
L	-0.0091777	0.0118207	-0.78	0.438
_Ispsetld_7	0.0271836	0.1338319	0.20	0.839
_Ispsetld_8	-0.4610055	0.1828833	-2.52	0.012
Igdscde2 4	0.1088255	0.1136678	0.96	0.338
_Ispsetld_5	-0.2565591	0.1293131	-1.98	0.047
_Iage_3	-0.3244223	0.1343130	-2.42	0.016
_Imor_rent_8	0.0899324	0.0785240	1.15	0.252
_Itosettl4_3	-0.1887153	0.1506497	-1.25	0.210
_Igdscde 333	0.1036904	0.2246205	0.46	0.644
_Iloanbal1_3	-0.0226262	0.1006130	-0.22	0.822
_Iage_2	-0.2723326	0.1246967	-2.18	0.029
_Itimadd1_2	0.0696557	0.1103795	0.63	0.528
_Ispsetld_6	-0.2210197	0.1381357	-1.60	0.110
_Imor_rent_5	-0.0851579	0.1194901	-0.71	0.476
_Ispsetld_4	-0.1616897	0.1069019	-1.51	0.130
_Iloanbal3_4	-0.0344096	0.1040911	-0.33	0.741
_Itimebank_6	-0.0868331	0.1062525	-0.82	0.414
_Isocworst_4	0.0462197	0.2221337	0.21	0.835
_Itimadd1_4	0.2528039	0.1561137	1.62	0.105
_Isnbalall_4	0.1204128	0.0984050	1.22	0.221
_Iloanbal1_2	-0.0453255	0.1033973	-0.44	0.661
_Isocworst_3	0.8298785	0.3444363	2.41	0.016
_Inetincm_6	0.4237077	0.1328997	3.19	0.001
_Inetincm_5	0.3993098	0.1084849	3.68	0.000
_Imor_rent_6	-0.0289022	0.1140158	-0.25	0.800
_Ino_store_1	-0.2561560	0.1425755	-1.80	0.072
_Inoopen6_4	-0.0885848	0.0783869	-1.13	0.258

_Itosettl3_3	0.1197973	0.1172648	1.02	0.307
_Ibrand_2	-0.1107500	0.0762834	-1.45	0.147
_Ispsetld_3	-0.1048208	0.1019879	-1.03	0.304
_Ialcifdet_2	-0.2701756	0.2184082	-1.24	0.216
_Iinc_surp_5	0.1263974	0.1020875	1.24	0.216
_Itimadd1_5	0.1614716	0.1125719	1.43	0.151
_Itimadd1_3	0.1732643	0.1149247	1.51	0.132
_Itimadd1_9	0.1665754	0.1109827	1.50	0.133
_Iccjgt500_3	-0.0325307	0.0774216	-0.42	0.674
_Isocbal_3	-0.7122386	0.2753442	-2.59	0.010
_Itosettl5_3	0.0955689	0.1699444	0.56	0.574
_Isnbalall_7	-0.0592609	0.1518928	-0.39	0.696
_Ino_store_3	0.2730169	0.4865510	0.56	0.575
_Isnball6m_3	0.0664205	0.2167781	0.31	0.759
_Imor_rent_7	-0.1236541	0.2012783	-0.61	0.539
_Inetincm_7	0.8548974	0.2067028	4.14	0.000
_Inetincm_4	0.2371250	0.1076762	2.20	0.028
covt_raw_l 1	-0.0652286	0.0087551	-7.45	0.000
covt_Icpi_2	-0.9920369	0.0765399	-12.96	0.000
covt_logLX R	-0.1739539	0.0488316	-3.56	0.000
covt_Is 2_4	-0.4481058	0.1054952	-4.25	0.000
covt_Il 4_3	0.6034755	0.1232908	4.89	0.000
covt_Inew 1	-0.1184577	0.2836286	-0.42	0.676
covt_Ispv 2	0.5502057	0.1594124	3.45	0.001
covt_Iinc 6	-0.3302421	0.1296032	-2.55	0.011
covt_Ispl 3	-0.8502487	0.5589157	-1.52	0.128
covt_Iinc 4	-0.2004211	0.1232646	-1.63	0.104
covt_Issr 5	-1.0216380	0.2343810	-4.36	0.000
covt_Iwrs 4	-0.3881525	0.1103690	-3.52	0.000
covt_raw_t m	-0.0124616	0.0026199	-4.76	0.000
covt_Iage_7	0.1782236	0.2199116	0.81	0.418
covt_It k_4	0.4116224	0.1348302	3.05	0.002
covt_It k_5	0.7668584	0.2491529	3.08	0.002
covt_Ig 999	0.6895132	0.1473491	4.68	0.000
covt_Isnb 8	-0.3404504	0.1165280	-2.92	0.003
covt_Issr 4	-0.3379901	0.1897342	-1.78	0.075
covt_Isea 7	0.1975062	0.0721231	2.74	0.006
covt_Il 6_2	0.4672871	0.1845153	2.53	0.011
covt_Ig 111	0.4352901	0.1071350	4.06	0.000
covt_Issr 3	-0.0440702	0.1259814	-0.35	0.726
covt_Ispl 5	0.1738089	0.1387490	1.25	0.210
covt_It 1_6	-0.1451626	0.1166595	-1.24	0.213

covt_It k_2	0.3157187	0.1155460	2.73	0.006
covt_It k_3	0.2660523	0.1439847	1.85	0.065
covt_Iwrs 3	-0.5314546	0.3316357	-1.60	0.109
covt_Iinc 2	0.2171301	0.1163558	1.87	0.062
covt_Iint 1	0.1017433	0.0835493	1.22	0.223
covt_Imor 3	0.1421782	0.1148753	1.24	0.216
covt_It k_7	0.1525187	0.1199076	1.27	0.203
covt_Iloa 7	0.2112519	0.1182785	1.79	0.074
covt_It 1_7	0.1554766	0.1290215	1.21	0.228
covt_Issr 2	-0.0860001	0.0900370	-0.96	0.339
covt_Is 4_4	-0.1425349	0.0788932	-1.81	0.071
covt_Isps 9	0.8810693	0.1214800	7.25	0.000
covt_Imor 2	0.2341004	0.1205643	1.94	0.052
covt_Iinc 3	-0.0034272	0.1833872	-0.02	0.985
covt_Iage_6	-0.1532715	0.1680093	-0.91	0.362
covt_I tt_3	-0.1097381	0.1343637	-0.82	0.414
covt_Iage_5	-0.0326667	0.1622122	-0.20	0.840
covt_Iage_4	-0.0759479	0.1440631	-0.53	0.598
covt_Il 5_2	0.1857377	0.1909685	0.97	0.331
covt_ tl1_3	0.1277234	0.0886999	1.44	0.150
covt_L	-0.0108403	0.0131924	-0.82	0.411
covt_Isps 7	0.3979992	0.1461356	2.72	0.006
covt_Isps 8	0.8737096	0.1949153	4.48	0.000
covt_Ig 444	0.1153000	0.1241542	0.93	0.353
covt_Isps 5	0.5495661	0.1385855	3.97	0.000
covt_Iage_3	0.0064088	0.1490909	0.04	0.966
covt_Imor 8	-0.1758178	0.0865784	-2.03	0.042
covt_It 4_3	0.4840174	0.1707273	2.84	0.005
covt_Ig 333	0.2781709	0.2463569	1.13	0.259
covt_Il 1_3	-0.1348555	0.1103917	-1.22	0.222
covt_Iage_2	0.0482975	0.1392874	0.35	0.729
covt_It 1_2	-0.1470848	0.1218657	-1.21	0.227
covt_Isps 6	0.4786609	0.1483113	3.23	0.001
covt_Imor 5	0.2647781	0.1269798	2.09	0.037
covt_Isps 4	0.3524015	0.1173359	3.00	0.003
covt_Iloa 4	0.1783364	0.1121783	1.59	0.112
covt_It k_6	0.1852914	0.1186932	1.56	0.119
covt_Isoc 4	0.1968687	0.2361874	0.83	0.405
covt_It 1_4	-0.4084757	0.1756138	-2.33	0.020
covt_Isnb 4	0.0295850	0.1055809	0.28	0.779
covt_Il 1_2	-0.1022837	0.1110202	-0.92	0.357
covt_I st_3	-0.4772427	0.3637950	-1.31	0.190

covt_Inet 6	-0.6383882	0.1440872	-4.43	0.000
covt_Inet 5	-0.5673734	0.1176981	-4.82	0.000
covt_Imor 6	0.2017022	0.1231247	1.64	0.101
covt_Ino_ 1	0.0809213	0.1562615	0.52	0.605
covt_Inoo 4	-0.0009676	0.0856405	-0.01	0.991
covt_It 3_3	0.0287035	0.1312643	0.22	0.827
covt_Ibra 2	0.1874752	0.0841314	2.23	0.026
covt_Isps 3	0.2191031	0.1123145	1.95	0.051
covt_Ialc 2	0.6373848	0.2498938	2.55	0.011
covt_Iinc 5	-0.0174252	0.1099802	-0.16	0.874
covt_It 1_5	0.0351622	0.1217490	0.29	0.773
covt_I d1_3	0.0102438	0.1250541	0.08	0.935
covt_Itim 9	0.0037435	0.1206338	0.03	0.975
covt_Iccj 3	0.1159192	0.0841685	1.38	0.168
covt_Is 1_3	0.5342964	0.2900761	1.84	0.065
covt_It 5_3	-0.3301246	0.2070366	-1.59	0.111
covt_Isnb 7	-0.1432147	0.1813719	-0.79	0.430
covt_Ino_ 3	-1.4642190	0.7294338	-2.01	0.045
covt_Isnb 3	-0.2378516	0.2346706	-1.01	0.311
covt_Imor 7	0.3554686	0.2198754	1.62	0.106
covt_Inet 7	-1.1486580	0.2370546	-4.85	0.000
covt_Inet 4	-0.3220521	0.1164902	-2.76	0.006

Table C.10: Stata FAQ Model

_t	Coef.	Robust Std. Err.	z	$P > \|z\|$
raw_loanapr1	0.055683	0.004230	13.16	0.000
_Icpi_2	0.505078	0.027132	18.62	0.000
logLXAPR	0.089326	0.023508	3.80	0.000
_Ispl6m12_4	0.144212	0.044131	3.27	0.001
_Iloanbal4_3	0.048746	0.047381	1.03	0.304
_Inewbus_1	-0.078673	0.088183	-0.89	0.372
_Ispvaldel_2	-0.099098	0.069613	-1.42	0.155
_Iinc_surp_6	0.137913	0.046416	2.97	0.003
_Ispl6m4_3	0.663937	0.240600	2.76	0.006
_Iinc_surp_4	0.170344	0.043592	3.91	0.000
_Issrc4to6_5	0.129881	0.119412	1.09	0.277
_Iwrst46al_4	0.053697	0.046053	1.17	0.244
raw_term	0.000645	0.000968	0.67	0.505
_Iagc_7	-0.592572	0.081371	7.28	0.000
_Itimebank_4	-0.129845	0.050243	-2.58	0.010
_Itimebank_5	-0.258509	0.081006	-3.19	0.001

_Igdscde 999	0.164859	0.048864	3.37	0.001
_Isnball6m_8	0.054399	0.041554	1.31	0.190
_Issrc4to6_4	0.289023	0.079518	3.63	0.000
_Isearches_7	-0.019680	0.025580	-0.77	0.442
_Iloanbal6_2	0.052374	0.079495	0.66	0.510
Igdscde2 1	0.180802	0.036669	4.93	0.000
_Issrc4to6_3	0.199101	0.049965	3.98	0.000
_Ispl6m12_5	-0.092973	0.040843	-2.28	0.023
_Itimadd1_6	0.195756	0.042638	4.59	0.000
_Itimebank_2	-0.030673	0.045132	-0.68	0.497
_Itimebank_3	-0.161976	0.054421	-2.98	0.003
_Iwrstnrev_3	0.133941	0.143957	0.93	0.352
_Iinc_surp_2	0.012753	0.039194	0.33	0.745
_Iinternet_1	-0.143801	0.030911	-4.65	0.000
_Imortbal_3	-0.072925	0.040656	-1.79	0.073
_Itimebank_7	-0.060121	0.047929	-1.25	0.210
_Iloanbal1_7	-0.045913	0.038839	-1.18	0.237
_Itimadd1_7	0.043833	0.045247	0.97	0.333
_Issrc4to6_2	0.030462	0.032015	0.95	0.341
_Ispl6m4_4	-0.178735	0.028433	-6.29	0.000
_Ispsetld_9	0.436150	0.041159	10.60	0.000
_Imortbal_2	0.075838	0.040148	1.89	0.059
_Iinc_surp_3	-0.211094	0.071219	-2.96	0.003
_Iage_6	-0.501902	0.068013	-7.38	0.000
_Isocsett_3	-0.150272	0.043455	-3.46	0.001
_Iage_5	-0.465264	0.065312	-7.12	0.000
_Iage_4	-0.382663	0.059924	-6.39	0.000
_Iloanbal5_2	0.090305	0.083877	1.08	0.282
_Itosettl1_3	0.058035	0.032562	1.78	0.075
L	-0.024798	0.005679	-4.37	0.000
_Ispsetld_7	0.367219	0.052163	7.04	0.000
_Ispsetld_8	0.291699	0.057486	5.07	0.000
Igdscde2 4	0.200880	0.050949	3.94	0.000
_Ispsetld_5	0.216137	0.045760	4.72	0.000
_Iage_3	-0.303268	0.060306	-5.03	0.000
_Imor_rent_8	-0.053732	0.033540	-1.60	0.109
_Itosettl4_3	0.212166	0.071832	2.95	0.003
_Igdscde 333	0.311906	0.089070	3.50	0.000
_Iloanbal1_3	-0.134120	0.040211	-3.34	0.001
_Iage_2	-0.226981	0.059117	-3.84	0.000
_Itimadd1_2	-0.048786	0.046722	-1.04	0.296
_Ispsetld_6	0.184888	0.051168	3.61	0.000

_Imor_rent_5	0.144775	0.039818	3.64	0.000
_Ispsetld_4	0.136617	0.042304	3.23	0.001
_Iloanbal3_4	0.122958	0.038491	3.19	0.001
_Itimebank_6	0.056446	0.048971	1.15	0.249
_Isocworst_4	0.189112	0.088832	2.13	0.033
_Itimadd1_4	-0.074600	0.073023	-1.02	0.307
_Isnbalall_4	0.147958	0.036221	4.08	0.000
_Iloanbal1_2	-0.128585	0.039635	-3.24	0.001
_Isocworst_3	0.383961	0.106376	3.61	0.000
_Inetincm_6	-0.123720	0.051548	-2.40	0.016
_Inetincm_5	-0.074704	0.042307	-1.77	0.077
_Imor_rent_6	0.137815	0.042063	3.28	0.001
_Ino_store_1	-0.194089	0.059541	-3.26	0.001
_Inoopen6_4	-0.091062	0.031776	-2.87	0.004
_Itosettl3_3	0.141220	0.052126	2.71	0.007
_Ibrand_2	0.036917	0.032006	1.15	0.249
_Ispsetld_3	0.076397	0.042349	1.80	0.071
_Ialcifdet_2	0.167155	0.115947	1.44	0.149
_Iinc_surp_5	0.109138	0.037634	2.90	0.004
_Itimadd1_5	0.194161	0.043143	4.50	0.000
_Itimadd1_3	0.182268	0.044109	4.13	0.000
_Itimadd1_9	0.166857	0.043923	3.80	0.000
_Iccjgt500_3	0.065131	0.030326	2.15	0.032
_Isocbal_3	-0.227640	0.082942	-2.74	0.006
_Itosettl5_3	-0.147957	0.091470	-1.62	0.106
_Isnbalall_7	-0.116830	0.079121	-1.48	0.140
_Ino_store_3	-0.494758	0.362736	-1.36	0.173
_Isnball6m_3	-0.145395	0.086918	-1.67	0.094
_Imor_rent_7	0.166009	0.079148	2.10	0.036
_Inetincm_7	-0.042096	0.097994	-0.43	0.668
_Inetincm_4	-0.036089	0.039464	-0.91	0.360

Appendix D

Appendix for Chapter 6

Table D.1: Model 1 acceptance Non-segmentation

paid	Coef.	Std. Err.	z	$P > \|z\|$	[95% Conf.	Interval]
raw_loanapr1	-0.109821	0.004722	-23.26	0.00	-0.119075	-0.100567
logLXAPR	-0.901505	0.026754	-33.70	0.00	-0.953942	-0.849069
_Icpi_2	-1.086233	0.030306	-35.84	0.00	-1.145630	-1.026835
_Inewbus_1	-0.538680	0.055403	-9.72	0.00	-0.647268	-0.430092
L	0.107568	0.006538	16.45	0.00	0.094754	0.120383
_Itosettl1_3	0.316319	0.041300	7.66	0.00	0.235373	0.397266
_Iloanbal3_4	-0.213842	0.057720	-3.70	0.00	-0.326970	-0.100714
_Igdscde2_3	-0.308869	0.049835	-6.20	0.00	-0.406544	-0.211194
raw_term	0.010300	0.001195	8.62	0.00	0.007957	0.012642
_Isnball6m_8	0.006934	0.085466	0.08	0.94	-0.160576	0.174444
_Itosettl2_3	0.266951	0.053825	4.96	0.00	0.161456	0.372446
_Ibrand_2	0.208700	0.030334	6.88	0.00	0.149246	0.268153
_Iloanbal4_2	0.359453	0.098042	3.67	0.00	0.167295	0.551612
_Ispl6m12_5	-0.208903	0.040323	-5.18	0.00	-0.287934	-0.129872
_Isocsett_2	-0.991852	0.120655	-8.22	0.00	-1.228331	-0.755374
_Ispsetld_9	-0.359439	0.043466	-8.27	0.00	-0.444632	-0.274247
_Itimadd1_4	0.238913	0.071136	3.36	0.00	0.099489	0.378337
_Ispvaldel_4	0.213947	0.888467	0.24	0.81	-1.527416	1.955310
_Itosettl4_3	0.013415	0.112321	0.12	0.91	-0.206731	0.233560
_Igdscde 444	0.699600	0.159527	4.39	0.00	0.386934	1.012266
_Iloanbal2_2	0.425217	0.083841	5.07	0.00	0.260892	0.589541
_Imortbal_2	-0.227225	0.043435	-5.23	0.00	-0.312356	-0.142093
_Iinternet_1	-0.158126	0.031639	-5.00	0.00	-0.220138	-0.096115
_Ispvaldel_2	-0.384439	0.084311	-4.56	0.00	-0.549687	-0.219192
_Iloanbal6_2	-0.328136	0.087967	-3.73	0.00	-0.500548	-0.155724
_Iloanbal1_8	-0.102668	0.035766	-2.87	0.00	-0.172768	-0.032568

_Igdscde 200	0.887350	0.216241	4.10	0.00	0.463526	1.311174
_Isnball6m_5	0.012270	0.114352	0.11	0.92	-0.211856	0.236395
_Inoopen6_2	-0.142901	0.047722	-2.99	0.00	-0.236435	-0.049368
_Iloanbal3_3	0.357621	0.093247	3.84	0.00	0.174861	0.540382
_Ispl6m12_3	0.589850	0.137728	4.28	0.00	0.319907	0.859792
_Igdscde2_2	0.174039	0.063780	2.73	0.01	0.049033	0.299046
_Ino_store_1	0.224606	0.069598	3.23	0.00	0.088196	0.361016
_Iage_9	-0.144874	0.040451	-3.58	0.00	-0.224157	-0.065591
_Itosettl3_2	-0.054681	0.051128	-1.07	0.29	-0.154889	0.045527
_Iloanbal2_4	0.165822	0.050754	3.27	0.00	0.066346	0.265298
_Isocworst_2	-0.368389	0.352973	-1.04	0.30	-1.060204	0.323426
_Isnball6m_7	-0.645688	0.119019	-5.43	0.00	-0.878960	-0.412415
_Isnbalall_2	0.233814	0.053345	4.38	0.00	0.129259	0.338369
_Isnbalall_6	0.187906	0.047879	3.92	0.00	0.094066	0.281747
_Ispl6m12_4	0.324437	0.060326	5.38	0.00	0.206201	0.442673
_Isnrecact_2	-3.472373	0.872165	-3.98	0.00	-5.181784	-1.762961
_Ispl6mact_4	3.025700	0.876523	3.45	0.00	1.307746	4.743653
_Itimadd1_9	-0.184539	0.041895	-4.40	0.00	-0.266652	-0.102427
_Isnw12tv_2	0.171833	0.063417	2.71	0.01	0.047538	0.296127
_Itimadd1_3	-0.161754	0.045558	-3.55	0.00	-0.251045	-0.072462
_Ispsetld_8	-0.239048	0.067212	-3.56	0.00	-0.370780	-0.107315
_Igdscde 111	0.438051	0.168992	2.59	0.01	0.106834	0.769269
_Ispl6m4_5	-0.083613	0.032643	-2.56	0.01	-0.147591	-0.019635
_Itimadd1_6	-0.128004	0.041991	-3.05	0.00	-0.210304	-0.045704
_Itimadd1_5	-0.107161	0.042895	-2.50	0.01	-0.191235	-0.023088
_Isnball6m_4	-0.505948	0.115709	-4.37	0.00	-0.732733	-0.279163
_Isnball6m_3	-0.641071	0.169271	-3.79	0.00	-0.972835	-0.309307
_Ino_deps_4	0.076175	0.030561	2.49	0.01	0.016278	0.136073
_Imor_rent_5	0.129076	0.044054	2.93	0.00	0.042732	0.215420
_Issrc4to6_5	0.323415	0.128197	2.52	0.01	0.072154	0.574676
_Ispsetld_7	-0.178632	0.059096	-3.02	0.00	-0.294457	-0.062807
_Ispsetld_6	-0.158857	0.052476	-3.03	0.00	-0.261708	-0.056005
_Ispsetld_3	-0.123689	0.041576	-2.98	0.00	-0.205176	-0.042203
_Ino_other_1	0.360534	0.160025	2.25	0.02	0.046892	0.674177
_Imor_rent_2	0.087906	0.037534	2.34	0.02	0.014341	0.161471
_Isnbalall_5	0.113028	0.045306	2.49	0.01	0.024230	0.201827
_Isnball6m_6	-0.349162	0.116654	-2.99	0.00	-0.577800	-0.120524
_Iccjgt500_6	-0.064776	0.029624	-2.19	0.03	-0.122838	-0.006714
_Inetincm_9	0.082680	0.040325	2.05	0.04	0.003644	0.161716
_Ino_amex_1	0.300576	0.144681	2.08	0.04	0.017007	0.584145
_Iloanbal1_2	0.223906	0.082826	2.70	0.01	0.061571	0.386241
_Iloanbal1_7	0.137829	0.048719	2.83	0.01	0.042341	0.233317

_Isocbal_2	0.241939	0.098483	2.46	0.01	0.048916	0.434962
_Itosettl4_2	-0.237381	0.064534	-3.68	0.00	-0.363866	-0.110897
_Iloanbal4_3	-0.238809	0.072886	-3.28	0.00	-0.381663	-0.095956
_Itosettl3_3	0.187901	0.087359	2.15	0.03	0.016681	0.359122
_Isnball6m_2	-0.232633	0.113193	-2.06	0.04	-0.454487	-0.010778
_cons	6.419856	0.253751	25.30	0.00	5.922513	6.917200

Table D.2: Model 1 acceptance on Internet segment

paid	Coef.	Std. Err.	z	$P > \|z\|$	[95% Conf.	Interval]
raw_loanapr1	-0.133445	0.009172	-14.55	0.00	-0.151423	-0.115468
logLXAPR	-0.980643	0.046289	-21.19	0.00	-1.071368	-0.889918
L	0.126448	0.010074	12.55	0.00	0.106705	0.146192
_Icpi_2	-0.618278	0.055028	-11.24	0.00	-0.726130	-0.510425
_Igdscde2_3	-0.408150	0.063024	-6.48	0.00	-0.531675	-0.284626
_Iloanbal4_3	-0.202304	0.081425	-2.48	0.01	-0.361894	-0.042714
_Itosettl1_3	0.246044	0.068622	3.59	0.00	0.111546	0.380541
raw_term	0.009433	0.001877	5.03	0.00	0.005754	0.013112
_Isnball6m_8	0.148997	0.067384	2.21	0.03	0.016928	0.281066
_Inewbus_1	-0.340683	0.103804	-3.28	0.00	-0.544136	-0.137230
_Iloanbal3_3	0.481369	0.137941	3.49	0.00	0.211010	0.751729
_Itimadd1_8	0.208307	0.066722	3.12	0.00	0.077535	0.339079
_Itimebank_4	0.191424	0.074992	2.55	0.01	0.044444	0.338405
_Ispsetld_9	-0.218998	0.063103	-3.47	0.00	-0.342678	-0.095318
_Iage_4	0.212976	0.063138	3.37	0.00	0.089228	0.336725
_Ino_visa_3	0.287657	0.129558	2.22	0.03	0.033728	0.541585
_Igdscde 200	0.868778	0.284551	3.05	0.00	0.311068	1.426489
_Itimebank_7	-0.193754	0.059924	-3.23	0.00	-0.311204	-0.076305
_Itimebank_9	-0.196182	0.066680	-2.94	0.00	-0.326873	-0.065492
_Isnball6m_7	-0.371057	0.124067	-2.99	0.00	-0.614225	-0.127890
_Isnball6m_4	-0.354434	0.131047	-2.70	0.01	-0.611282	-0.097586
_Iworst12_3	-0.348842	0.130584	-2.67	0.01	-0.604781	-0.092902
_Itosettl2_3	0.166396	0.085755	1.94	0.05	-0.001681	0.334473
_Ispl6m12_5	-0.142044	0.060607	-2.34	0.02	-0.260831	-0.023256
_Imor_rent_2	0.150257	0.062403	2.41	0.02	0.027950	0.272564
_Ispl6mact_4	-0.459703	0.142294	-3.23	0.00	-0.738595	-0.180812
_Ibrand_2	-0.126908	0.049863	-2.55	0.01	-0.224638	-0.029179
_Ino_amex_1	0.628345	0.282256	2.23	0.03	0.075134	1.181555
_Itimadd1_4	0.273923	0.124562	2.20	0.03	0.029785	0.518060
_Iage_6	0.163256	0.072163	2.26	0.02	0.021820	0.304691
_Isocsett_2	-0.725949	0.255001	-2.85	0.00	-1.225741	-0.226157
_Isocworst_2	-2.222121	0.684282	-3.25	0.00	-3.563290	-0.880952

_Inoopen6_2	-0.153640	0.073767	-2.08	0.04	-0.298222	-0.009059
_Ispl6mact_3	0.432669	0.206197	2.10	0.04	0.028531	0.836807
_Ispl6m12_4	0.223090	0.100295	2.22	0.03	0.026515	0.419665
_Iloanbal3_4	-0.144386	0.065598	-2.20	0.03	-0.272956	-0.015816
_Ispsetld_8	-0.212566	0.102629	-2.07	0.04	-0.413715	-0.011417
_Ispvaldel_2	-0.257955	0.129977	-1.98	0.05	-0.512705	-0.003204
_cons	6.117873	0.346584	17.65	0.00	5.438581	6.797165

Table D.3: Model 1 acceptance on Non-Internet segment

paid	Coef.	Std. Err.	z	$P > \|z\|$	[95% Conf.	Interval]
_Icpi_2	-1.312067	0.036837	-35.62	0.00	-1.384266	-1.239868
raw_loanapr1	-0.093883	0.005602	-16.76	0.00	-0.104862	-0.082903
_Inewbus_1	-0.480023	0.065438	-7.34	0.00	-0.608280	-0.351767
logLXAPR	-0.763473	0.034492	-22.13	0.00	-0.831076	-0.695869
_Itosettl2_3	0.302263	0.069450	4.35	0.00	0.166144	0.438383
_Igdscde2_3	-0.298025	0.062438	-4.77	0.00	-0.420401	-0.175649
_Iloanbal3_4	-0.306247	0.086849	-3.53	0.00	-0.476468	-0.136025
_Ibrand_2	0.381302	0.039405	9.68	0.00	0.304070	0.458534
L	0.074213	0.009311	7.97	0.00	0.055964	0.092462
_Itosettl1_3	0.357294	0.051957	6.88	0.00	0.255461	0.459128
raw_term	0.010530	0.001565	6.73	0.00	0.007462	0.013598
_Ispl6mact_5	-0.081579	0.045686	-1.79	0.07	-0.171123	0.007964
_Isocsett_2	-0.996011	0.133288	-7.47	0.00	-1.257251	-0.734771
_Imortbal_2	-0.204936	0.044154	-4.64	0.00	-0.291476	-0.118395
_Iloanbal4_2	0.426054	0.131156	3.25	0.00	0.168992	0.683115
_Ispsetld_9	-0.324407	0.055824	-5.81	0.00	-0.433819	-0.214995
_Isnball6m_8	0.294851	0.059973	4.92	0.00	0.177306	0.412396
_Iloanbal2_2	0.690494	0.115789	5.96	0.00	0.463551	0.917437
_Ispvaldel_4	-2.815235	0.787105	-3.58	0.00	-4.357932	-1.272537
_Itosettl4_3	0.105045	0.140587	0.75	0.46	-0.170501	0.380591
_Iloanbal2_4	0.328215	0.070846	4.63	0.00	0.189360	0.467069
_Igdscde 444	0.824142	0.186634	4.42	0.00	0.458346	1.189937
_Itimadd1_4	0.314060	0.085555	3.67	0.00	0.146375	0.481744
_Iloanbal1_8	-0.126945	0.045853	-2.77	0.01	-0.216816	-0.037074
_Ispvaldel_3	0.496752	0.131822	3.77	0.00	0.238386	0.755119
_Ispl6m12_5	-0.229952	0.062574	-3.67	0.00	-0.352596	-0.107309
_Iloanbal6_2	-0.429358	0.118047	-3.64	0.00	-0.660726	-0.197990
_Ino_deps_4	0.120105	0.039606	3.03	0.00	0.042478	0.197732
_Itosettl3_3	0.250815	0.096383	2.60	0.01	0.061909	0.439722
_Ino_store_1	0.252688	0.073291	3.45	0.00	0.109040	0.396336
_Ispl6m12_4	0.400589	0.075642	5.30	0.00	0.252334	0.548845

_Ispl6mact_4	-0.463244	0.106327	-4.36	0.00	-0.671640	-0.254847
_Iloanbal2_3	0.160235	0.064855	2.47	0.01	0.033121	0.287348
_Ispsetld_1	0.197055	0.052663	3.74	0.00	0.093837	0.300273
_Isearches_7	-0.091071	0.036218	-2.51	0.01	-0.162058	-0.020084
_Ialcifdet_2	0.349434	0.142317	2.46	0.01	0.070499	0.628370
_Igdscde 200	1.006511	0.355948	2.83	0.01	0.308865	1.704157
_Iloanbal3_2	-0.133162	0.082787	-1.61	0.11	-0.295421	0.029097
_Ispsetld_2	0.160399	0.052598	3.05	0.00	0.057310	0.263489
_Inoopen6_2	-0.170372	0.062789	-2.71	0.01	-0.293436	-0.047308
_Igdscde2_2	0.181057	0.077835	2.33	0.02	0.028503	0.333612
_Ispl6m12_3	0.504421	0.173686	2.90	0.00	0.164004	0.844839
_Isnball6m_5	0.240316	0.111398	2.16	0.03	0.021979	0.458653
_Isocworst_2	-0.909867	0.325712	-2.79	0.01	-1.548250	-0.271484
_Ino_other_1	0.487456	0.196279	2.48	0.01	0.102756	0.872157
_Itimadd1_9	-0.144725	0.052529	-2.76	0.01	-0.247679	-0.041771
_Itimadd1_3	-0.146237	0.058476	-2.50	0.01	-0.260847	-0.031626
_Iloanbal1_2	0.374997	0.113202	3.31	0.00	0.153124	0.596869
_Igdscde 111	0.469818	0.220995	2.13	0.03	0.036676	0.902960
_Itosettl4_2	-0.278986	0.081377	-3.43	0.00	-0.438482	-0.119491
_Iloanbal4_3	-0.194389	0.097530	-1.99	0.05	-0.385544	-0.003235
_Inetincm_6	-0.150235	0.064226	-2.34	0.02	-0.276116	-0.024354
_Ispl6m4_2	0.201220	0.100117	2.01	0.04	0.004993	0.397447
_Issrc4to6_5	0.362567	0.162964	2.22	0.03	0.043164	0.681971
_Isnball6m_3	-0.656838	0.243934	-2.69	0.01	-1.134939	-0.178736
_Isocworst_4	-0.238795	0.132806	-1.80	0.07	-0.499089	0.021499
_Isnball6m_7	-0.372230	0.136276	-2.73	0.01	-0.639327	-0.105134
_Isnbalall_2	0.242895	0.075466	3.22	0.00	0.094984	0.390806
_Isnbalall_6	0.184583	0.063112	2.92	0.00	0.060887	0.308280
_Iloanbal3_3	0.270665	0.127982	2.11	0.03	0.019824	0.521506
_Iloanbal1_7	0.178155	0.066555	2.68	0.01	0.047709	0.308600
_Isocbal_2	0.249041	0.113525	2.19	0.03	0.026536	0.471546
_cons	5.056947	0.299245	16.90	0.00	4.470438	5.643456

Table D.4: Model 1 default Non-Segmentation

_t	Coef.	Std. Err.	z	$P > \lvert z \rvert$	[95% Conf.	Interval]
L	-0.002947	0.013425	-0.22	0.83	-0.029258	0.023365
raw_loanapr1	0.111734	0.007427	15.05	0.00	0.097178	0.126290
_Icpi_2	1.379818	0.076488	18.04	0.00	1.229904	1.529733
logLXAPR	0.205549	0.045748	4.49	0.00	0.115884	0.295214
_Ispl6m12_4	0.714983	0.095007	7.53	0.00	0.528772	0.901194
_Iloanbal4_3	-0.526297	0.092395	-5.70	0.00	-0.707387	-0.345207

_Ispvaldel_2	-0.460430	0.136247	-3.38	0.00	-0.727468	-0.193391
_Ispl6m4_3	1.561862	0.378576	4.13	0.00	0.819866	2.303858
_Isocworst_3	1.009246	0.282471	3.57	0.00	0.455613	1.562879
_Iinc_surp_6	0.246995	0.093423	2.64	0.01	0.063890	0.430100
_Iwrst46al_4	0.392416	0.099135	3.96	0.00	0.198116	0.586717
_Itimebank_4	-0.528957	0.113617	-4.66	0.00	-0.751642	-0.306272
_Itimebank_5	-0.955438	0.251306	-3.80	0.00	-1.447988	-0.462887
_Isocbal_3	-0.894538	0.291042	-3.07	0.00	-1.464971	-0.324106
_Igdscde 999	-0.514257	0.157840	-3.26	0.00	-0.823618	-0.204896
raw_term	0.010933	0.003219	3.40	0.00	0.004623	0.017242
_Iinc_surp_2	-0.224989	0.098300	-2.29	0.02	-0.417654	-0.032325
_Issrc4to6_4	0.493364	0.162702	3.03	0.00	0.174474	0.812254
_Isnball6m_8	0.280184	0.100503	2.79	0.01	0.083202	0.477166
_Issrc4to6_5	0.571990	0.179724	3.18	0.00	0.219739	0.924242
_Isearches_7	-0.182298	0.071432	-2.55	0.01	-0.322303	-0.042294
_Itimadd1_6	0.201436	0.094994	2.12	0.03	0.015251	0.387621
_Itimebank_2	-0.289882	0.084325	-3.44	0.00	-0.455155	-0.124609
_Itimebank_3	-0.390592	0.122258	-3.19	0.00	-0.630212	-0.150971
_Imor_rent_7	0.427748	0.174159	2.46	0.01	0.086403	0.769093
_Ispl6m12_3	0.443712	0.190808	2.33	0.02	0.069735	0.817688
_Iinternet_1	-0.188297	0.078667	-2.39	0.02	-0.342481	-0.034112
Igdscde2 1	-0.266788	0.112795	-2.37	0.02	-0.487863	-0.045714
_Ispsetld_8	-0.504540	0.212706	-2.37	0.02	-0.921435	-0.087644
_Isnw12tv_2	-0.355917	0.158162	-2.25	0.02	-0.665910	-0.045925
_Iinc_surp_4	0.197508	0.093938	2.10	0.04	0.013393	0.381623
_Iloanbal4_2	0.347772	0.146529	2.37	0.02	0.060580	0.634965
_Iage_7	-0.392830	0.184179	-2.13	0.03	-0.753814	-0.031847
_Itimadd1_7	-0.250545	0.115221	-2.17	0.03	-0.476375	-0.024715
_Iloanbal2_3	0.200047	0.095267	2.10	0.04	0.013328	0.386766

Table D.5: Model 1 default on Internet segment

_t	Coef.	Std. Err.	z	$P > \|z\|$	[95% Conf.	Interval]
L	-0.017683	0.024936	-0.71	0.48	-0.066557	0.031190
raw_loanapr1	0.103387	0.016790	6.16	0.00	0.070480	0.136294
_Icpi_2	1.463070	0.135094	10.83	0.00	1.198292	1.727849
logLXAPR	0.270360	0.096584	2.80	0.01	0.081059	0.459662
_Ispl6m12_4	0.550518	0.197135	2.79	0.01	0.164140	0.936895
_Iloanbal4_3	-0.338257	0.180530	-1.87	0.06	-0.692089	0.015576
_Ispvaldel_2	-0.700212	0.248600	-2.82	0.01	-1.187459	-0.212965
_Ispl6m4_3	0.141860	1.141846	0.12	0.90	-2.096116	2.379836
_Isocworst_3	0.623484	0.526380	1.18	0.24	-0.408203	1.655170

_Iinc_surp_6	0.397846	0.184960	2.15	0.03	0.035330	0.760361
_Iwrst46al_4	0.497919	0.194110	2.57	0.01	0.117471	0.878367
_Itimebank_4	-0.612971	0.217710	-2.82	0.01	-1.039674	-0.186268
_Itimebank_5	-0.849079	0.468607	-1.81	0.07	-1.767531	0.069373
_Isocbal_3	-0.686154	0.588508	-1.17	0.24	-1.839609	0.467301
_Igdscde 999	-1.134135	0.463348	-2.45	0.01	-2.042282	-0.225989
raw_term	0.012880	0.006383	2.02	0.04	0.000370	0.025390
_Iinc_surp_2	-0.372294	0.178941	-2.08	0.04	-0.723012	-0.021575
_Issrc4to6_4	0.835570	0.322602	2.59	0.01	0.203282	1.467858
_Isnball6m_8	0.327405	0.188383	1.74	0.08	-0.041820	0.696629
_Issrc4to6_5	0.627965	0.321488	1.95	0.05	-0.002140	1.258069
_Isearches_7	-0.005990	0.131743	-0.05	0.96	-0.264202	0.252222
_Itimadd1_6	0.369117	0.169671	2.18	0.03	0.036569	0.701665
_Itimebank_2	-0.276456	0.156095	-1.77	0.08	-0.582395	0.029484
_Itimebank_3	-0.434553	0.276785	-1.57	0.12	-0.977041	0.107935
_Imor_rent_7	0.328145	0.276418	1.19	0.24	-0.213624	0.869914
_Ispl6m12_3	0.262085	0.385126	0.68	0.50	-0.492748	1.016917
Igdscde2 1	-0.240032	0.237658	-1.01	0.31	-0.705833	0.225769
_Ispsetld_8	-0.700896	0.420451	-1.67	0.10	-1.524964	0.123172
_Isnw12tv_2	-0.549218	0.274623	-2.00	0.05	-1.087470	-0.010967
_Iinc_surp_4	-0.154026	0.203021	-0.76	0.45	-0.551940	0.243889
_Iloanbal4_2	0.490375	0.277231	1.77	0.08	-0.052987	1.033737
_Iage_7	-0.072276	0.352488	-0.21	0.84	-0.763139	0.618588
_Itimadd1_7	-0.149420	0.227560	-0.66	0.51	-0.595430	0.296590
_Iloanbal2_3	0.305914	0.173949	1.76	0.08	-0.035020	0.646847

Table D.6: Model 1 default on Non-Internet segment

_t	Coef.	Std. Err.	z	$P > \|z\|$	[95% Conf.	Interval]
L	0.002549	0.016507	0.15	0.88	-0.029803	0.034901
raw_loanapr1	0.114501	0.008506	13.46	0.00	0.097830	0.131171
_Icpi_2	1.345353	0.093405	14.40	0.00	1.162282	1.528424
logLXAPR	0.185011	0.053818	3.44	0.00	0.079530	0.290493
_Ispl6m12_4	0.786350	0.109424	7.19	0.00	0.571883	1.000816
_Iloanbal4_3	-0.613384	0.108442	-5.66	0.00	-0.825927	-0.400841
_Ispvaldel_2	-0.343246	0.165265	-2.08	0.04	-0.667158	-0.019333
_Ispl6m4_3	1.893873	0.396858	4.77	0.00	1.116046	2.671701
_Isocworst_3	1.221027	0.340138	3.59	0.00	0.554369	1.887686
_Iinc_surp_6	0.227496	0.109085	2.09	0.04	0.013693	0.441300
_Iwrst46al_4	0.374790	0.116376	3.22	0.00	0.146697	0.602884
_Itimebank_4	-0.514158	0.133755	-3.84	0.00	-0.776312	-0.252004
_Itimebank_5	-1.005323	0.298981	-3.36	0.00	-1.591315	-0.419330

_Isocbal_3	-1.085209	0.347168	-3.13	0.00	-1.765646	-0.404772
_Igdscde 999	-0.402630	0.172063	-2.34	0.02	-0.739868	-0.065392
raw_term	0.010143	0.003768	2.69	0.01	0.002758	0.017528
_Iinc_surp_2	-0.146555	0.118584	-1.24	0.22	-0.378975	0.085865
_Issrc4to6_4	0.406334	0.190456	2.13	0.03	0.033047	0.779622
_Isnball6m_8	0.313416	0.121229	2.59	0.01	0.075812	0.551021
_Issrc4to6_5	0.587793	0.219503	2.68	0.01	0.157576	1.018010
_Isearches_7	-0.266020	0.086324	-3.08	0.00	-0.435212	-0.096827
_Itimadd1_6	0.146279	0.116046	1.26	0.21	-0.081168	0.373725
_Itimebank_2	-0.305476	0.100639	-3.04	0.00	-0.502724	-0.108228
_Itimebank_3	-0.398373	0.136972	-2.91	0.00	-0.666834	-0.129912
_Imor_rent_7	0.520420	0.227093	2.29	0.02	0.075327	0.965513
_Ispl6m12_3	0.493322	0.221029	2.23	0.03	0.060113	0.926531
Igdscde2 1	-0.247913	0.129109	-1.92	0.06	-0.500962	0.005136
_Ispsetld_8	-0.413158	0.248055	-1.67	0.10	-0.899336	0.073020
_Isnw12tv_2	-0.282342	0.195769	-1.44	0.15	-0.666043	0.101358
_Iinc_surp_4	0.320907	0.107834	2.98	0.00	0.109558	0.532257
_Iloanbal4_2	0.247772	0.175740	1.41	0.16	-0.096673	0.592216
_Iage_7	-0.502447	0.217669	-2.31	0.02	-0.929070	-0.075823
_Itimadd1_7	-0.281407	0.134393	-2.09	0.04	-0.544814	-0.018001
_Iloanbal2_3	0.164989	0.114829	1.44	0.15	-0.060072	0.390050

Table D.7: Model 1 paying back early Non-Segmentation

_t	Coef.	Std. Err.	z	$P > \|z\|$	[95% Conf.	Interval]
logLXAPR	-0.000024	0.027349	0.00	1.00	-0.053627	0.053578
raw_loanapr1	0.040366	0.004645	8.69	0.00	0.031262	0.049470
_Icpi_2	0.366036	0.032138	11.39	0.00	0.303047	0.429026
_Ispsetld_9	0.582632	0.048080	12.12	0.00	0.488397	0.676868
_Ispl6m4_4	-0.193319	0.033077	-5.84	0.00	-0.258149	-0.128488
_Itosettl4_2	0.018049	0.062166	0.29	0.77	-0.103795	0.139893
_Iinc_surp_3	-0.265471	0.082465	-3.22	0.00	-0.427100	-0.103843
_Iage_6	-0.614028	0.075599	-8.12	0.00	-0.762200	-0.465857
_Iage_7	-0.697223	0.091989	-7.58	0.00	-0.877519	-0.516928
_Isocsett_3	-0.119095	0.063585	-1.87	0.06	-0.243720	0.005530
_Iage_5	-0.527521	0.072971	-7.23	0.00	-0.670542	-0.384500
_Iage_4	-0.447398	0.067182	-6.66	0.00	-0.579073	-0.315724
_Imortbal_2	0.121908	0.045837	2.66	0.01	0.032070	0.211746
_Itosettl1_3	0.111819	0.038929	2.87	0.00	0.035519	0.188119
_Iloanbal5_2	0.261557	0.115284	2.27	0.02	0.035604	0.487510
_Ispsetld_7	0.422466	0.061306	6.89	0.00	0.302308	0.542624
_Ispsetld_8	0.430874	0.066529	6.48	0.00	0.300479	0.561269

L	-0.017017	0.007308	-2.33	0.02	-0.031341	-0.002694
_Issrc4to6_3	0.177868	0.055021	3.23	0.00	0.070029	0.285708
_Imor_rent_5	0.254899	0.046212	5.52	0.00	0.164325	0.345473
_Igdscde 999	0.309159	0.057878	5.34	0.00	0.195720	0.422599
Igdscde2 1	0.245662	0.043882	5.60	0.00	0.159656	0.331669
Igdscde2 4	0.233134	0.057811	4.03	0.00	0.119826	0.346442
_Isnbalall_4	0.145024	0.040462	3.58	0.00	0.065721	0.224328
_Iage_3	-0.303641	0.068180	-4.45	0.00	-0.437271	-0.170011
_Inoopen6_4	-0.095923	0.032242	-2.98	0.00	-0.159117	-0.032729
_Imor_rent_6	0.253668	0.049173	5.16	0.00	0.157291	0.350046
_Inetincm_5	-0.209360	0.049091	-4.26	0.00	-0.305577	-0.113143
_Inetincm_6	-0.220895	0.058299	-3.79	0.00	-0.335159	-0.106630
_Igdscde 333	0.352904	0.104600	3.37	0.00	0.147891	0.557917
_Iloanbal1_3	-0.139578	0.045578	-3.06	0.00	-0.228910	-0.050247
_Iloanbal1_2	-0.139254	0.044157	-3.15	0.00	-0.225800	-0.052708
_Ispsetld_5	0.254810	0.053780	4.74	0.00	0.149403	0.360217
_Ispsetld_6	0.272157	0.058415	4.66	0.00	0.157665	0.386648
_Ispsetld_4	0.200906	0.049757	4.04	0.00	0.103385	0.298427
_Itimadd1_4	-0.296150	0.084623	-3.50	0.00	-0.462007	-0.130292
_Itimadd1_2	-0.166998	0.050186	-3.33	0.00	-0.265361	-0.068636
_Iage_2	-0.200534	0.067267	-2.98	0.00	-0.332374	-0.068694
_Iloanbal6_2	0.230300	0.110892	2.08	0.04	0.012956	0.447643
_Itimebank_6	0.138349	0.044199	3.13	0.00	0.051721	0.224977
_Iinternet_1	-0.106486	0.034607	-3.08	0.00	-0.174316	-0.038657
_Itimebank_2	0.084104	0.032901	2.56	0.01	0.019619	0.148588
_Itosettl3_2	-0.090896	0.039121	-2.32	0.02	-0.167572	-0.014220
_Issrc4to6_5	-0.422647	0.170375	-2.48	0.01	-0.756576	-0.088718
_Isocworst_4	0.266780	0.093047	2.87	0.00	0.084411	0.449149
_Isocworst_3	0.307496	0.085530	3.60	0.00	0.139861	0.475131
_Ibrand_2	0.095833	0.038458	2.49	0.01	0.020457	0.171209
_Iinc_surp_4	0.115157	0.041086	2.80	0.01	0.034630	0.195683
_Iinc_surp_5	0.085930	0.033125	2.59	0.01	0.021007	0.150854
_Imor_rent_7	0.325197	0.092857	3.50	0.00	0.143200	0.507194
_Ialcifdet_2	0.374900	0.150137	2.50	0.01	0.080637	0.669164
_Isocbal_3	-0.234455	0.088774	-2.64	0.01	-0.408448	-0.060461
_Imor_rent_4	0.107339	0.044623	2.41	0.02	0.019879	0.194798
_Inetincm_7	-0.322168	0.122298	-2.63	0.01	-0.561869	-0.082468
_Inetincm_4	-0.107155	0.045540	-2.35	0.02	-0.196412	-0.017897
_Ino_store_1	-0.167949	0.069014	-2.43	0.02	-0.303213	-0.032685
_Ispsetld_3	0.112417	0.049119	2.29	0.02	0.016145	0.208689
_Itosettl4_3	0.422810	0.096828	4.37	0.00	0.233030	0.612590
_Iloanbal4_3	0.224973	0.068570	3.28	0.00	0.090579	0.359367

_Itimadd1_8	-0.097285	0.045660	-2.13	0.03	-0.186776	-0.007793
_Iccjgt500_3	0.075616	0.035110	2.15	0.03	0.006802	0.144430
_Issrc4to6_4	0.205185	0.096120	2.13	0.03	0.016793	0.393577
_Itosettl5_3	-0.275430	0.131459	-2.10	0.04	-0.533085	-0.017775
_Ispl6m12_5	-0.096507	0.046489	-2.08	0.04	-0.187622	-0.005391
_Isocnoact_2	0.258502	0.119641	2.16	0.03	0.024010	0.492994
_Igdscde3_2	-0.413088	0.207918	-1.99	0.05	-0.820600	-0.005577
_Ispl6m12_3	-0.313816	0.156468	-2.01	0.05	-0.620487	-0.007145
_Isocsett_2	0.170080	0.086439	1.97	0.05	0.000663	0.339496

Table D.8: Model 1 paying back early Internet segment

_t	Coef.	Std. Err.	z	$P > \|z\|$	[95% Conf.	Interval]
logLXAPR	0.023693	0.047634	0.50	0.62	-0.069667	0.117054
raw_loanapr1	0.054035	0.008816	6.13	0.00	0.036756	0.071313
_Icpi_2	0.332499	0.066272	5.02	0.00	0.202608	0.462389
_Ispsetld_9	0.617580	0.079404	7.78	0.00	0.461952	0.773208
_Ispl6m4_4	-0.286388	0.056862	-5.04	0.00	-0.397835	-0.174941
_Itosettl4_2	0.005251	0.105341	0.05	0.96	-0.201213	0.211716
_Iinc_surp_3	-0.257968	0.112914	-2.28	0.02	-0.479276	-0.036660
_Iage_6	-0.841356	0.132373	-6.36	0.00	-1.100802	-0.581909
_Iage_7	-0.773945	0.165683	-4.67	0.00	-1.098678	-0.449212
_Isocsett_3	-0.276923	0.149297	-1.85	0.06	-0.569539	0.015694
_Iage_5	-0.617040	0.120445	-5.12	0.00	-0.853107	-0.380972
_Iage_4	-0.568565	0.106124	-5.36	0.00	-0.776565	-0.360565
_Imortbal_2	0.283546	0.263619	1.08	0.28	-0.233138	0.800230
_Itosettl1_3	0.154911	0.069969	2.21	0.03	0.017774	0.292048
_Iloanbal5_2	0.509999	0.205572	2.48	0.01	0.107086	0.912913
_Ispsetld_7	0.439562	0.101029	4.35	0.00	0.241550	0.637574
_Ispsetld_8	0.445578	0.114013	3.91	0.00	0.222116	0.669039
L	-0.017997	0.011427	-1.58	0.12	-0.040393	0.004399
_Issrc4to6_3	0.144949	0.094669	1.53	0.13	-0.040600	0.330497
_Imor_rent_5	0.313544	0.072608	4.32	0.00	0.171236	0.455853
_Igdscde 999	0.148624	0.119724	1.24	0.21	-0.086032	0.383279
Igdscde2 1	0.256887	0.081003	3.17	0.00	0.098124	0.415650
Igdscde2 4	0.359118	0.121031	2.97	0.00	0.121902	0.596334
_Isnbalall_4	0.150493	0.067770	2.22	0.03	0.017666	0.283319
_Iage_3	-0.464874	0.106652	-4.36	0.00	-0.673907	-0.255841
_Inoopen6_4	-0.118744	0.054657	-2.17	0.03	-0.225870	-0.011618
_Imor_rent_6	0.304849	0.077166	3.95	0.00	0.153607	0.456091
_Inetincm_5	-0.225261	0.077621	-2.90	0.00	-0.377395	-0.073128
_Inetincm_6	-0.150346	0.086082	-1.75	0.08	-0.319064	0.018372
_Igdscde 333	0.271526	0.165990	1.64	0.10	-0.053808	0.596860
_Iloanbal1_3	-0.209886	0.075763	-2.77	0.01	-0.358379	-0.061393
_Iloanbal1_2	-0.171325	0.076166	-2.25	0.02	-0.320607	-0.022043
_Ispsetld_5	0.337497	0.091641	3.68	0.00	0.157884	0.517111
_Ispsetld_6	0.225623	0.102987	2.19	0.03	0.023772	0.427474
_Ispsetld_4	0.135554	0.084688	1.60	0.11	-0.030431	0.301539
_Itimadd1_4	-0.226954	0.166822	-1.36	0.17	-0.553918	0.100010
_Itimadd1_2	-0.191470	0.090720	-2.11	0.04	-0.369278	-0.013663
_Iage_2	-0.329367	0.104061	-3.17	0.00	-0.533323	-0.125410
_Iloanbal6_2	0.163633	0.183741	0.89	0.37	-0.196493	0.523760

_Itimebank_6	0.028617	0.072765	0.39	0.69	-0.114000	0.171235
_Itimebank_2	0.088835	0.054840	1.62	0.11	-0.018649	0.196320
_Itosettl3_2	-0.024790	0.065053	-0.38	0.70	-0.152292	0.102711
_Issrc4to6_5	-0.506939	0.294751	-1.72	0.09	-1.084641	0.070763
_Isocworst_4	0.039629	0.263158	0.15	0.88	-0.476152	0.555410
_Isocworst_3	0.474396	0.244466	1.94	0.05	-0.004749	0.953541
_Ibrand_2	0.119087	0.064939	1.83	0.07	-0.008191	0.246364
_Iinc_surp_4	0.173210	0.074252	2.33	0.02	0.027679	0.318740
_Iinc_surp_5	0.120514	0.058102	2.07	0.04	0.006635	0.234392
_Imor_rent_7	0.315154	0.132749	2.37	0.02	0.054972	0.575336
_Ialcifdet_2	0.302875	0.238555	1.27	0.20	-0.164685	0.770434
_Isocbal_3	-0.500347	0.264208	-1.89	0.06	-1.018184	0.017491
_Imor_rent_4	0.068922	0.076446	0.90	0.37	-0.080909	0.218754
_Inetincm_7	-0.330149	0.172484	-1.91	0.06	-0.668211	0.007913
_Inetincm_4	-0.040886	0.073860	-0.55	0.58	-0.185648	0.103876
_Ino_store_1	-0.109241	0.227030	-0.48	0.63	-0.554212	0.335730
_Ispsetld_3	0.086128	0.083837	1.03	0.30	-0.078189	0.250445
_Itosettl4_3	0.639678	0.182985	3.50	0.00	0.281035	0.998321
_Iloanbal4_3	0.247648	0.117180	2.11	0.04	0.017979	0.477317
_Itimadd1_8	-0.140317	0.082266	-1.71	0.09	-0.301556	0.020922
_Iccjgt500_3	0.081505	0.061901	1.32	0.19	-0.039818	0.202828
_Issrc4to6_4	0.160136	0.172282	0.93	0.35	-0.177531	0.497803
_Itosettl5_3	-0.549697	0.306408	-1.79	0.07	-1.150245	0.050852
_Ispl6m12_5	-0.106917	0.072305	-1.48	0.14	-0.248631	0.034798
_Isocnoact_2	0.574077	0.428333	1.34	0.18	-0.265440	1.413594
_Igdscde3_2	-0.600977	0.385613	-1.56	0.12	-1.356763	0.154810
_Ispl6m12_3	-0.347583	0.307906	-1.13	0.26	-0.951068	0.255902
_Isocsett_2	0.221144	0.224640	0.98	0.33	-0.219142	0.661430

Table D.9: Model 1 paying back early Non-Internet segment

_t	Coef.	Std. Err.	z	$P > \lvert z \rvert$	[95% Conf.	Interval]
logLXAPR	-0.013317	0.035131	-0.38	0.71	-0.082173	0.055540
raw_loanapr1	0.032299	0.005625	5.74	0.00	0.021274	0.043324
_Icpi_2	0.376484	0.037072	10.16	0.00	0.303824	0.449145
_Ispsetld_9	0.556231	0.060624	9.18	0.00	0.437411	0.675051
_Ispl6m4_4	-0.150434	0.040840	-3.68	0.00	-0.230480	-0.070389
_Itosettl4_2	0.017340	0.077662	0.22	0.82	-0.134874	0.169554
_Iinc_surp_3	-0.266633	0.122080	-2.18	0.03	-0.505906	-0.027360
_Iage_6	-0.495145	0.095292	-5.20	0.00	-0.681914	-0.308376
_Iage_7	-0.624804	0.113514	-5.50	0.00	-0.847288	-0.402320
_Isocsett_3	-0.084992	0.070776	-1.20	0.23	-0.223710	0.053727

_Iage_5	-0.456859	0.093490	-4.89	0.00	-0.640097	-0.273622
_Iage_4	-0.365401	0.087828	-4.16	0.00	-0.537541	-0.193262
_Imortbal_2	0.127119	0.048433	2.62	0.01	0.032192	0.222046
_Itosettl1_3	0.084515	0.047207	1.79	0.07	-0.008009	0.177039
_Iloanbal5_2	0.128378	0.140808	0.91	0.36	-0.147601	0.404357
_Ispsetld_7	0.416424	0.077443	5.38	0.00	0.264638	0.568210
_Ispsetld_8	0.420344	0.082238	5.11	0.00	0.259161	0.581526
L	-0.017575	0.009977	-1.76	0.08	-0.037129	0.001979
_Issrc4to6_3	0.202225	0.067979	2.97	0.00	0.068989	0.335461
_Imor_rent_5	0.213661	0.060531	3.53	0.00	0.095023	0.332300
_Igdscde 999	0.360825	0.068056	5.30	0.00	0.227438	0.494211
Igdscde2 1	0.241490	0.052800	4.57	0.00	0.138003	0.344977
Igdscde2 4	0.210886	0.066221	3.18	0.00	0.081096	0.340677
_Isnbalall_4	0.147556	0.050701	2.91	0.00	0.048185	0.246928
_Iage_3	-0.199216	0.089520	-2.23	0.03	-0.374673	-0.023760
_Inoopen6_4	-0.079717	0.040150	-1.99	0.05	-0.158409	-0.001025
_Imor_rent_6	0.218267	0.064481	3.39	0.00	0.091888	0.344647
_Inetincm_5	-0.198993	0.063944	-3.11	0.00	-0.324320	-0.073665
_Inetincm_6	-0.290316	0.081712	-3.55	0.00	-0.450468	-0.130164
_Igdscde 333	0.434455	0.135384	3.21	0.00	0.169108	0.699801
_Iloanbal1_3	-0.097932	0.057349	-1.71	0.09	-0.210335	0.014470
_Iloanbal1_2	-0.131879	0.054437	-2.42	0.02	-0.238575	-0.025184
_Ispsetld_5	0.215265	0.066764	3.22	0.00	0.084409	0.346120
_Ispsetld_6	0.299342	0.071264	4.20	0.00	0.159668	0.439016
_Ispsetld_4	0.240434	0.061726	3.90	0.00	0.119453	0.361415
_Itimadd1_4	-0.322452	0.098646	-3.27	0.00	-0.515795	-0.129109
_Itimadd1_2	-0.157077	0.060504	-2.60	0.01	-0.275663	-0.038491
_Iage_2	-0.115869	0.089020	-1.30	0.19	-0.290345	0.058606
_Iloanbal6_2	0.279944	0.139792	2.00	0.05	0.005956	0.553932
_Itimebank_6	0.206602	0.056082	3.68	0.00	0.096684	0.316521
_Itimebank_2	0.084814	0.041327	2.05	0.04	0.003815	0.165813
_Itosettl3_2	-0.124774	0.049320	-2.53	0.01	-0.221440	-0.028108
_Issrc4to6_5	-0.379489	0.209718	-1.81	0.07	-0.790529	0.031552
_Isocworst_4	0.322821	0.100728	3.20	0.00	0.125397	0.520244
_Isocworst_3	0.294586	0.093262	3.16	0.00	0.111797	0.477376
_Ibrand_2	0.095639	0.048413	1.98	0.05	0.000752	0.190526
_Iinc_surp_4	0.086388	0.049543	1.74	0.08	-0.010713	0.183490
_Iinc_surp_5	0.068937	0.040514	1.70	0.09	-0.010469	0.148342
_Imor_rent_7	0.335205	0.131765	2.54	0.01	0.076950	0.593459
_Ialcifdet_2	0.405830	0.193923	2.09	0.04	0.025748	0.785912
_Isocbal_3	-0.233263	0.095517	-2.44	0.02	-0.420473	-0.046054
_Imor_rent_4	0.132816	0.055347	2.40	0.02	0.024338	0.241295

_Inetincm_7	-0.305949	0.175952	-1.74	0.08	-0.650807	0.038910
_Inetincm_4	-0.144659	0.058305	-2.48	0.01	-0.258936	-0.030383
_Ino_store_1	-0.163857	0.072828	-2.25	0.02	-0.306598	-0.021117
_Ispsetld_3	0.121432	0.060836	2.00	0.05	0.002196	0.240668
_Itosettl4_3	0.328819	0.115510	2.85	0.00	0.102423	0.555214
_Iloanbal4_3	0.214643	0.085232	2.52	0.01	0.047591	0.381695
_Itimadd1_8	-0.074040	0.055095	-1.34	0.18	-0.182025	0.033945
_Iccjgt500_3	0.068557	0.042818	1.60	0.11	-0.015365	0.152479
_Issrc4to6_4	0.230956	0.116585	1.98	0.05	0.002454	0.459458
_Itosettl5_3	-0.185113	0.146685	-1.26	0.21	-0.472610	0.102385
_Ispl6m12_5	-0.086341	0.061080	-1.41	0.16	-0.206056	0.033375
_Isocnoact_2	0.228274	0.125666	1.82	0.07	-0.018027	0.474575
Igdscde3 2	-0.353070	0.248203	-1.42	0.16	-0.839540	0.133399
_Ispl6m12_3	-0.309199	0.182254	-1.70	0.09	-0.666411	0.048013
_Isocsett_2	0.189098	0.094391	2.00	0.05	0.004096	0.374100

Table D.10: Model 2 acceptance Non-Segmentation

paid	Coef.	Std. Err.	z	$P > \|z\|$	[95% Conf.	Interval]
raw_loanapr1	-0.186747	0.004125	-45.27	0.00	-0.194833	-0.178662
_Icpi_2	-1.028874	0.029599	-34.76	0.00	-1.086888	-0.970860
_Inewbus_1	-0.465375	0.053941	-8.63	0.00	-0.571097	-0.359652
L	-0.082166	0.003141	-26.16	0.00	-0.088322	-0.076011
_Igdscde2_3	-0.257687	0.047973	-5.37	0.00	-0.351712	-0.163661
_Iloanbal3_4	-0.174507	0.046875	-3.72	0.00	-0.266382	-0.082633
_Isnball6m_8	0.000685	0.083729	0.01	0.99	-0.163421	0.164791
_Itosettl1_3	0.240704	0.039894	6.03	0.00	0.162514	0.318895
_Isocsett_2	-0.993333	0.117533	-8.45	0.00	-1.223693	-0.762972
_Itosettl2_3	0.219455	0.051789	4.24	0.00	0.117951	0.320959
_Iloanbal4_2	0.352175	0.092720	3.80	0.00	0.170448	0.533902
_Iage_9	-0.136106	0.043770	-3.11	0.00	-0.221893	-0.050318
_Ispsetld_9	-0.247431	0.040667	-6.08	0.00	-0.327136	-0.167725
_Inetincm_9	0.165102	0.040752	4.05	0.00	0.085230	0.244974
_Ibrand_2	0.153272	0.029752	5.15	0.00	0.094959	0.211584
_Iinternet_1	-0.202607	0.031190	-6.50	0.00	-0.263739	-0.141476
_Imortbal_2	-0.223320	0.043562	-5.13	0.00	-0.308699	-0.137941
_Iloanbal2_2	0.443349	0.080535	5.51	0.00	0.285504	0.601195
_Isnrecact_2	-2.125331	0.510226	-4.17	0.00	-3.125356	-1.125307
_Ispl6m12_4	0.319347	0.059428	5.37	0.00	0.202870	0.435824
_Ispl6mact_5	-0.090989	0.035659	-2.55	0.01	-0.160878	-0.021099
_Itimadd1_9	-0.239596	0.042125	-5.69	0.00	-0.322160	-0.157033
_Ispl6mact_4	1.533386	0.517730	2.96	0.00	0.518654	2.548117

_Isnball6m_5	-0.015443	0.111702	-0.14	0.89	-0.234374	0.203488
_Itimadd1_3	-0.216363	0.045605	-4.74	0.00	-0.305747	-0.126979
_Ispl6mact_2	-0.344037	0.056091	-6.13	0.00	-0.453973	-0.234100
_Ispvaldel_2	-0.359988	0.081412	-4.42	0.00	-0.519552	-0.200423
_Itosettl3_3	0.202780	0.067947	2.98	0.00	0.069607	0.335954
_Itimadd1_4	0.261120	0.071754	3.64	0.00	0.120486	0.401755
_Iloanbal6_2	-0.265935	0.083727	-3.18	0.00	-0.430036	-0.101834
_Igdscde 444	0.570504	0.151782	3.76	0.00	0.273018	0.867990
_Igdscde 200	0.791283	0.206501	3.83	0.00	0.386548	1.196017
_Iloanbal1_8	-0.097923	0.034725	-2.82	0.01	-0.165982	-0.029865
_Iloanbal3_3	0.344621	0.088707	3.88	0.00	0.170759	0.518482
_Inoopen6_2	-0.208043	0.047553	-4.37	0.00	-0.301245	-0.114840
_Ino_store_1	0.226586	0.068345	3.32	0.00	0.092633	0.360539
_Ispl6m12_5	-0.160724	0.045385	-3.54	0.00	-0.249677	-0.071771
_Inoopen6_3	-0.253670	0.076606	-3.31	0.00	-0.403815	-0.103524
_Igdscde2_2	0.181232	0.061517	2.95	0.00	0.060662	0.301802
_Iloanbal1_2	0.339243	0.080208	4.23	0.00	0.182037	0.496449
_Iloanbal2_4	0.153267	0.048929	3.13	0.00	0.057368	0.249166
_Itimadd1_5	-0.134123	0.043002	-3.12	0.00	-0.218404	-0.049841
_Itimadd1_6	-0.131883	0.042032	-3.14	0.00	-0.214263	-0.049502
_Imor_rent_8	-0.132444	0.034174	-3.88	0.00	-0.199424	-0.065464
_Isnball6m_7	-0.702971	0.115334	-6.10	0.00	-0.929021	-0.476922
_Isnball6m_4	-0.577425	0.110956	-5.20	0.00	-0.794895	-0.359956
_Itosettl4_2	-0.252892	0.051773	-4.88	0.00	-0.354365	-0.151420
_Iloanbal4_3	-0.221737	0.061383	-3.61	0.00	-0.342045	-0.101428
_Itimebank_9	-0.155843	0.041775	-3.73	0.00	-0.237720	-0.073966
_Itimebank_7	-0.132608	0.041519	-3.19	0.00	-0.213985	-0.051232
_Isnball6m_3	-0.709800	0.162020	-4.38	0.00	-1.027354	-0.392247
_Igdscde 111	0.428192	0.161639	2.65	0.01	0.111385	0.744999
_Iwrst46al_4	0.232124	0.068555	3.39	0.00	0.097758	0.366490
_Isocbal_2	0.312375	0.093994	3.32	0.00	0.128150	0.496600
_Iloanbal1_7	0.131736	0.047146	2.79	0.01	0.039331	0.224141
_Isnbalall_2	0.154777	0.050156	3.09	0.00	0.056473	0.253081
_Isnball6m_6	-0.427998	0.112505	-3.80	0.00	-0.648503	-0.207493
_Isnball6m_2	-0.274680	0.110548	-2.48	0.01	-0.491350	-0.058009
_Iinc_surp_7	-0.084863	0.040444	-2.10	0.04	-0.164131	-0.005594
_Isnbalall_6	0.105910	0.044817	2.36	0.02	0.018071	0.193749
_Itimebank_8	-0.083388	0.041634	-2.00	0.05	-0.164988	-0.001787
_Isnw12tv_2	0.152156	0.062146	2.45	0.01	0.030352	0.273961
_Iccjgt500_6	-0.068556	0.029032	-2.36	0.02	-0.125458	-0.011654
_Ino_amex_1	0.330360	0.139008	2.38	0.02	0.057909	0.602811
_Ino_other_1	0.352490	0.156159	2.26	0.02	0.046425	0.658555

_Issrc4to6_2	-0.072103	0.031691	-2.28	0.02	-0.134217	-0.009989
_Iwrst46al_5	0.298393	0.103709	2.88	0.00	0.095127	0.501658
_Ispl6m4_3	-0.812068	0.297248	-2.73	0.01	-1.394664	-0.229472
_Imor_rent_4	-0.085613	0.039731	-2.15	0.03	-0.163484	-0.007743
_Ino_deps_4	0.072504	0.031247	2.32	0.02	0.011262	0.133747
_Ispsetld_8	-0.138877	0.065050	-2.13	0.03	-0.266373	-0.011380
_Iage_5	0.151833	0.044982	3.38	0.00	0.063671	0.239996
_Iage_4	0.128983	0.041708	3.09	0.00	0.047236	0.210730
_Iage_6	0.142541	0.046053	3.10	0.00	0.052278	0.232804
_Iinc_surp_9	0.089274	0.042237	2.11	0.04	0.006490	0.172058
_Iage_7	0.096151	0.046576	2.06	0.04	0.004864	0.187438
_cons	5.356435	0.235670	22.73	0.00	4.894530	5.818341

Table D.11: Model 2 acceptance on Internet segment

paid	Coef.	Std. Err.	z	$P > \|z\|$	[95% Conf.	Interval]
raw_loanapr1	-0.252919	0.007259	-34.84	0.00	-0.267145	-0.238692
L	-0.061755	0.004303	-14.35	0.00	-0.070190	-0.053321
_Icpi_2	-0.583965	0.054170	-10.78	0.00	-0.690135	-0.477794
_Igdscde2_3	-0.357327	0.060686	-5.89	0.00	-0.476269	-0.238385
_Isnball6m_8	0.318231	0.071952	4.42	0.00	0.177207	0.459255
_Iloanbal3_4	-0.191535	0.054500	-3.51	0.00	-0.298353	-0.084717
_Iage_9	-0.163269	0.062353	-2.62	0.01	-0.285480	-0.041059
_Itosettl1_3	0.207082	0.063611	3.26	0.00	0.082406	0.331757
_Iloanbal3_3	0.432219	0.136846	3.16	0.00	0.164006	0.700433
_Itimebank_9	-0.307694	0.065036	-4.73	0.00	-0.435163	-0.180225
_Itimebank_7	-0.238275	0.060206	-3.96	0.00	-0.356276	-0.120274
_Ibrand_2	-0.229492	0.048462	-4.74	0.00	-0.324475	-0.134509
_Itimadd1_9	-0.153331	0.060824	-2.52	0.01	-0.272544	-0.034118
_Ispl6mact_2	-0.459336	0.087444	-5.25	0.00	-0.630723	-0.287950
_Itimadd1_8	0.240554	0.067437	3.57	0.00	0.108380	0.372727
_Ispl6mact_4	-0.674999	0.141005	-4.79	0.00	-0.951363	-0.398635
_Ispl6mact_5	-0.158189	0.048082	-3.29	0.00	-0.252427	-0.063951
_Isnball6m_5	0.288931	0.145599	1.98	0.05	0.003563	0.574299
_Itimebank_4	0.167045	0.075494	2.21	0.03	0.019079	0.315011
_Iage_4	0.271567	0.064493	4.21	0.00	0.145163	0.397970
_Inetincm_9	0.217295	0.076138	2.85	0.00	0.068067	0.366524
_Igdscde 200	0.828426	0.276131	3.00	0.00	0.287219	1.369633
_Isocsett_2	-0.769603	0.243284	-3.16	0.00	-1.246431	-0.292776
_Ispl6m4_3	-1.651193	0.464851	-3.55	0.00	-2.562283	-0.740103
_Ispsetld_9	-0.145787	0.061976	-2.35	0.02	-0.267258	-0.024315
_Iloanbal4_2	0.333513	0.132538	2.52	0.01	0.073743	0.593283

_Ino_amex_1	0.699701	0.265324	2.64	0.01	0.179675	1.219727
_Isnball6m_7	-0.384338	0.123164	-3.12	0.00	-0.625734	-0.142941
_Isnball6m_4	-0.376168	0.129523	-2.90	0.00	-0.630027	-0.122308
_Inoopen6_2	-0.171729	0.072663	-2.36	0.02	-0.314146	-0.029313
_Ispl6m12_4	0.237750	0.099069	2.40	0.02	0.043578	0.431922
_Isocworst_3	-0.233986	0.099013	-2.36	0.02	-0.428049	-0.039924
_Imor_rent_2	0.162981	0.061934	2.63	0.01	0.041593	0.284369
_Itimadd1_2	0.173683	0.071149	2.44	0.02	0.034234	0.313133
_Iinc_surp_7	-0.149591	0.065082	-2.30	0.02	-0.277150	-0.022033
_Ino_visa_3	0.272387	0.121621	2.24	0.03	0.034014	0.510760
_Ispsetld_5	0.167931	0.076328	2.20	0.03	0.018332	0.317531
_Iwrst46al_4	0.224770	0.107318	2.09	0.04	0.014430	0.435110
_Itimadd1_4	0.288863	0.124708	2.32	0.02	0.044441	0.533286
_Imor_rent_5	0.134038	0.064104	2.09	0.04	0.008396	0.259680
_Iage_6	0.179094	0.073652	2.43	0.02	0.034738	0.323450
_Iage_5	0.159727	0.070122	2.28	0.02	0.022290	0.297163
_cons	4.600866	0.292018	15.76	0.00	4.028521	5.173211

Table D.12: Model 2 acceptance on Non-Internet segment

paid	Coef.	Std. Err.	z	$P > \|z\|$	[95% Conf.	Interval]
_Icpi_2	-1.296857	0.036304	-35.72	0.00	-1.368012	-1.225702
raw_loanapr1	-0.148856	0.004981	-29.89	0.00	-0.158617	-0.139094
_Inewbus_1	-0.419175	0.060831	-6.89	0.00	-0.538401	-0.299948
L	-0.101040	0.004471	-22.60	0.00	-0.109803	-0.092278
_Itosettl2_3	0.284297	0.065974	4.31	0.00	0.154990	0.413604
_Iloanbal3_4	-0.416144	0.081290	-5.12	0.00	-0.575469	-0.256819
_Igdscde2_3	-0.275953	0.060799	-4.54	0.00	-0.395117	-0.156788
_Ibrand_2	0.329204	0.038931	8.46	0.00	0.252900	0.405508
_Isnball6m_8	0.005983	0.103523	0.06	0.95	-0.196918	0.208883
_Itosettl1_3	0.284460	0.050695	5.61	0.00	0.185100	0.383820
_Isocsett_2	-0.999528	0.130091	-7.68	0.00	-1.254502	-0.744553
_Iloanbal4_2	0.401786	0.131427	3.06	0.00	0.144194	0.659378
_Imortbal_2	-0.207635	0.045252	-4.59	0.00	-0.296326	-0.118943
_Ispsetld_9	-0.276418	0.055164	-5.01	0.00	-0.384537	-0.168299
_Iloanbal2_2	0.730403	0.113760	6.42	0.00	0.507437	0.953369
_Isocworst_2	-1.060231	0.308972	-3.43	0.00	-1.665805	-0.454657
_Ispl6m12_5	-0.254204	0.054245	-4.69	0.00	-0.360522	-0.147887
_Itimadd1_4	0.326463	0.086703	3.77	0.00	0.156529	0.496397
_Itosettl4_3	0.108212	0.130256	0.83	0.41	-0.147085	0.363509
_Iloanbal2_4	0.328251	0.069291	4.74	0.00	0.192443	0.464060
_Igdscde 444	0.747054	0.181041	4.13	0.00	0.392219	1.101888

_Isncais3m_3	-0.334102	0.105146	-3.18	0.00	-0.540184	-0.128020
_Iloanbal1_8	-0.168157	0.045586	-3.69	0.00	-0.257504	-0.078810
_Ispvaldel_3	0.488763	0.128436	3.81	0.00	0.237034	0.740492
_Iloanbal6_2	-0.463886	0.138156	-3.36	0.00	-0.734668	-0.193105
_Inetincm_9	0.160887	0.047361	3.40	0.00	0.068062	0.253713
_Iloanbal1_2	0.426398	0.105931	4.03	0.00	0.218777	0.634019
_Ispl6m4_5	-0.106425	0.042884	-2.48	0.01	-0.190476	-0.022375
_Isearches_7	-0.105385	0.035807	-2.94	0.00	-0.175565	-0.035205
_Iloanbal3_2	-0.141454	0.080578	-1.76	0.08	-0.299384	0.016476
_Igdscde 200	0.963059	0.341995	2.82	0.01	0.292762	1.633356
_Ialcifdet_2	0.300993	0.140151	2.15	0.03	0.026302	0.575684
_Ino_deps_4	0.136974	0.039829	3.44	0.00	0.058910	0.215038
_Iage_9	-0.158092	0.054300	-2.91	0.00	-0.264518	-0.051666
_Ino_store_1	0.232273	0.072796	3.19	0.00	0.089597	0.374950
_Itosettl3_2	-0.122339	0.056910	-2.15	0.03	-0.233880	-0.010798
_Iloanbal2_3	0.171810	0.064044	2.68	0.01	0.046286	0.297333
_Isnball6m_5	-0.075644	0.138492	-0.55	0.59	-0.347083	0.195795
_Igdscde 111	0.523840	0.214349	2.44	0.02	0.103725	0.943956
_Itimadd1_3	-0.214798	0.058782	-3.65	0.00	-0.330009	-0.099588
_Itimadd1_9	-0.212015	0.053243	-3.98	0.00	-0.316370	-0.107661
_Ispl6mact_4	-0.522538	0.104464	-5.00	0.00	-0.727283	-0.317792
_Ispl6m12_4	0.385739	0.074645	5.17	0.00	0.239437	0.532041
_Isocworst_4	-0.354192	0.127356	-2.78	0.01	-0.603805	-0.104579
_Ispsetld_1	0.171913	0.052111	3.30	0.00	0.069777	0.274049
_Ispsetld_2	0.161104	0.051946	3.10	0.00	0.059292	0.262916
_Ispvaldel_4	-2.098463	0.765812	-2.74	0.01	-3.599427	-0.597499
_Ino_other_1	0.469742	0.193015	2.43	0.02	0.091440	0.848044
_Igdscde2_2	0.172161	0.075629	2.28	0.02	0.023932	0.320391
_Ispl6m12_3	0.367650	0.168334	2.18	0.03	0.037721	0.697579
_Imor_rent_8	-0.113617	0.043297	-2.62	0.01	-0.198479	-0.028756
_Inetincm_6	-0.149335	0.063590	-2.35	0.02	-0.273969	-0.024701
_Itosettl6_2	-0.272999	0.111713	-2.44	0.02	-0.491951	-0.054046
_Isnball6m_3	-0.956257	0.250601	-3.82	0.00	-1.447426	-0.465087
_Isnball6m_7	-0.700510	0.159388	-4.39	0.00	-1.012905	-0.388115
_Isnball6m_4	-0.524130	0.144576	-3.63	0.00	-0.807494	-0.240766
_Inoopen6_2	-0.127085	0.063155	-2.01	0.04	-0.250867	-0.003303
_Itimadd1_5	-0.116877	0.056797	-2.06	0.04	-0.228198	-0.005556
_Iloanbal3_3	0.258979	0.124216	2.08	0.04	0.015521	0.502438
_Isnbalall_2	0.199339	0.073677	2.71	0.01	0.054934	0.343744
_Isnbalall_6	0.171298	0.062652	2.73	0.01	0.048503	0.294093
_Isnball6m_6	-0.499885	0.150542	-3.32	0.00	-0.794942	-0.204827
_Isnball6m_2	-0.327515	0.141546	-2.31	0.02	-0.604940	-0.050091

_Itosettl4_2	-0.256427	0.084455	-3.04	0.00	-0.421956	-0.090897
_Iloanbal4_3	-0.222456	0.096919	-2.30	0.02	-0.412413	-0.032499
_Iloanbal5_2	-0.298005	0.135494	-2.20	0.03	-0.563569	-0.032442
_cons	5.537923	0.314941	17.58	0.00	4.920650	6.155196

Table D.13: Model 2 default Non-Segmentation

_t	Coef.	Std. Err.	z	$P > \|z\|$	[95% Conf.	Interval]
raw_loanapr1	0.123655	0.006579	18.80	0.00	0.110761	0.136548
_Icpi_2	1.372879	0.077416	17.73	0.00	1.221147	1.524610
L	0.039693	0.007385	5.37	0.00	0.025219	0.054168
_Ispl6m12_4	0.715035	0.095187	7.51	0.00	0.528472	0.901599
_Iloanbal4_3	-0.536693	0.092167	-5.82	0.00	-0.717336	-0.356049
raw_term	0.016223	0.003020	5.37	0.00	0.010305	0.022142
_Ispvaldel_2	-0.495112	0.135739	-3.65	0.00	-0.761155	-0.229069
_Ispl6m4_3	1.658914	0.380107	4.36	0.00	0.913919	2.403909
_Isocworst_3	1.050004	0.282522	3.72	0.00	0.496271	1.603737
_Itimebank_4	-0.571067	0.113130	-5.05	0.00	-0.792797	-0.349336
_Iwrst46al_4	0.394551	0.099653	3.96	0.00	0.199234	0.589868
_Itimebank_5	-1.015976	0.251137	-4.05	0.00	-1.508195	-0.523757
_Isocbal_3	-0.926586	0.290915	-3.19	0.00	-1.496768	-0.356404
_Iinc_surp_2	-0.285370	0.093495	-3.05	0.00	-0.468617	-0.102124
_Igdscde 999	-0.520624	0.161119	-3.23	0.00	-0.836412	-0.204837
_Issrc4to6_4	0.465124	0.164418	2.83	0.01	0.142871	0.787378
_Itimebank_2	-0.310455	0.084056	-3.69	0.00	-0.475201	-0.145708
_Itimebank_3	-0.433393	0.121776	-3.56	0.00	-0.672069	-0.194718
_Ispl6m12_3	0.531265	0.190994	2.78	0.01	0.156925	0.905606
_Isearches_7	-0.187989	0.071392	-2.63	0.01	-0.327915	-0.048063
_Itimadd1_6	0.223362	0.095004	2.35	0.02	0.037158	0.409565
_Issrc4to6_5	0.601267	0.181064	3.32	0.00	0.246388	0.956147
_Isnball6m_8	0.348931	0.108924	3.20	0.00	0.135443	0.562418
_Iloanbal5_2	-0.348930	0.141964	-2.46	0.01	-0.627174	-0.070686
_Ispsetld_8	-0.512172	0.213014	-2.40	0.02	-0.929671	-0.094673
_Itimadd1_7	-0.269574	0.115593	-2.33	0.02	-0.496132	-0.043015
_Iage_7	-0.397718	0.183814	-2.16	0.03	-0.757987	-0.037448
_Iinternet_1	-0.250119	0.084364	-2.96	0.00	-0.415469	-0.084769
Igdscde2 1	-0.270488	0.113525	-2.38	0.02	-0.492993	-0.047982
_Iloanbal2_3	0.213506	0.095165	2.24	0.03	0.026986	0.400027
_Ispsetld_1	0.207529	0.091410	2.27	0.02	0.028368	0.386689
_Isnw12tv_2	-0.341937	0.158222	-2.16	0.03	-0.652046	-0.031827
_Inoopen6_4	-0.160652	0.076655	-2.10	0.04	-0.310894	-0.010410
_Imortbal_4	0.189382	0.083779	2.26	0.02	0.025178	0.353587

_Imor_rent_7	0.349061	0.174336	2.00	0.05	0.007368	0.690754
Igdscde3 4	0.482910	0.243127	1.99	0.05	0.006391	0.959429

Table D.14: Model 2 default on Internet segment

_t	Coef.	Std. Err.	z	$P > \|z\|$	[95% Conf.	Interval]
raw_loanapr1	0.139803	0.012730	10.98	0.00	0.114853	0.164754
_Icpi_2	1.489309	0.135541	10.99	0.00	1.223655	1.754963
raw_term	0.016919	0.006021	2.81	0.01	0.005117	0.028720
_Isnbalall_7	0.436572	0.228427	1.91	0.06	-0.011137	0.884281
_Ispvaldel_2	-0.873972	0.234021	-3.73	0.00	-1.332645	-0.415298
_Ibrand_2	-0.449351	0.140855	-3.19	0.00	-0.725423	-0.173280
_Isncais3m_3	0.622983	0.258648	2.41	0.02	0.116042	1.129924
_Itimebank_4	-0.495144	0.204565	-2.42	0.02	-0.896084	-0.094204
_Iwrst46al_4	0.444881	0.191883	2.32	0.02	0.068796	0.820965
L	0.040553	0.013223	3.07	0.00	0.014637	0.066469
_Iinc_surp_6	0.375270	0.174395	2.15	0.03	0.033462	0.717079
_Issrc4to6_4	0.649260	0.315400	2.06	0.04	0.031087	1.267433
_Ispl6m12_4	0.610427	0.210700	2.90	0.00	0.197462	1.023392
_Itimadd1_6	0.410436	0.166195	2.47	0.01	0.084700	0.736172
_Iwrstnrev_2	0.375824	0.192607	1.95	0.05	-0.001678	0.753327
Igdscde3 4	1.455844	0.527746	2.76	0.01	0.421481	2.490207
_Ino_visa_3	-1.412567	0.483670	-2.92	0.00	-2.360543	-0.464591
_Iloanbal3_4	-0.594260	0.201468	-2.95	0.00	-0.989129	-0.199390
_Igdscde 999	-0.947874	0.453800	-2.09	0.04	-1.837304	-0.058443
_Itosettl3_2	-0.389409	0.192987	-2.02	0.04	-0.767655	-0.011162
_Iinc_surp_2	-0.353737	0.173586	-2.04	0.04	-0.693959	-0.013515
_Ispl6m12_5	-0.522644	0.263391	-1.98	0.05	-1.038880	-0.006408

Table D.15: Model 2 default on Non-Internet segment

_t	Coef.	Std. Err.	z	$P > \|z\|$	[95% Conf.	Interval]
raw_loanapr1	0.124267	0.007729	16.08	0.00	0.109120	0.139415
_Icpi_2	1.332967	0.093873	14.20	0.00	1.148981	1.516954
L	0.042188	0.009095	4.64	0.00	0.024362	0.060015
_Ispl6m12_4	0.775229	0.110522	7.01	0.00	0.558610	0.991849
_Iloanbal4_3	-0.606028	0.109081	-5.56	0.00	-0.819823	-0.392233
_Ispl6m4_3	2.034859	0.395641	5.14	0.00	1.259417	2.810300
_Isocworst_3	1.264048	0.338940	3.73	0.00	0.599737	1.928359
raw_term	0.013788	0.003525	3.91	0.00	0.006880	0.020696
_Iage_7	-0.664115	0.219215	-3.03	0.00	-1.093768	-0.234462

_Ispvaldel_2	-0.373276	0.164735	-2.27	0.02	-0.696150	-0.050401
_Isocbal_3	-1.063016	0.346464	-3.07	0.00	-1.742074	-0.383959
_Isearches_7	-0.266996	0.086105	-3.10	0.00	-0.435758	-0.098234
_Iage_6	-0.368638	0.134010	-2.75	0.01	-0.631294	-0.105983
_Ispl6m12_2	0.333582	0.106074	3.14	0.00	0.125682	0.541483
_Itimebank_6	0.270381	0.106079	2.55	0.01	0.062470	0.478292
_Ispsetld_5	-0.462065	0.178273	-2.59	0.01	-0.811474	-0.112656
_Ispl6m4_2	0.360381	0.155725	2.31	0.02	0.055166	0.665596
_Issrc4to6_5	0.621142	0.220375	2.82	0.01	0.189216	1.053069
_Issrc4to6_4	0.470180	0.191656	2.45	0.01	0.094542	0.845818
_Isnball6m_8	0.284289	0.119964	2.37	0.02	0.049164	0.519414
_Iinc_surp_4	0.235631	0.098352	2.40	0.02	0.042866	0.428397
_Imor_rent_7	0.515629	0.228994	2.25	0.02	0.066809	0.964450
_Iage_5	-0.193118	0.117064	-1.65	0.10	-0.422559	0.036323
_Ispsetld_2	-0.301928	0.125942	-2.40	0.02	-0.548769	-0.055087
_Imortbal_4	0.198182	0.085844	2.31	0.02	0.029931	0.366433
_Iwrst46al_3	-0.568206	0.263792	-2.15	0.03	-1.085228	-0.051184
_Igdscde 888	1.093775	0.505921	2.16	0.03	0.102188	2.085361
_Itimebank_5	-0.736537	0.300775	-2.45	0.01	-1.326045	-0.147029
_Itimebank_4	-0.299999	0.129528	-2.32	0.02	-0.553870	-0.046128
_Igdscde 999	-0.376723	0.169581	-2.22	0.03	-0.709095	-0.044351
_Ispl6m12_3	0.452383	0.225101	2.01	0.04	0.011193	0.893572
_Itimadd1_7	-0.271861	0.133302	-2.04	0.04	-0.533128	-0.010594
_Iloanbal5_2	-0.332372	0.168280	-1.98	0.05	-0.662194	-0.002549

Table D.16: Model 2 paying back early Non-Segmentation

_t	Coef.	Std. Err.	z	$P > \|z\|$	[95% Conf.	Interval]
raw_loanapr1	0.040364	0.004504	8.96	0.00	0.031537	0.049192
_Icpi_2	0.366037	0.032130	11.39	0.00	0.303063	0.429011
_Ispsetld_9	0.582635	0.047979	12.14	0.00	0.488598	0.676672
_Ispl6m4_4	-0.193319	0.033077	-5.84	0.00	-0.258149	-0.128488
_Itosettl4_2	0.018048	0.062141	0.29	0.77	-0.103747	0.139842
_Iinc_surp_3	-0.265470	0.082459	-3.22	0.00	-0.427086	-0.103854
_Iage_6	-0.614025	0.075495	-8.13	0.00	-0.761991	-0.466058
_Iage_7	-0.697220	0.091905	-7.59	0.00	-0.877350	-0.517090
_Isocsett_3	-0.119095	0.063583	-1.87	0.06	-0.243715	0.005525
_Iage_5	-0.527518	0.072900	-7.24	0.00	-0.670400	-0.384636
_Iage_4	-0.447396	0.067143	-6.66	0.00	-0.578993	-0.315799
_Imortbal_2	0.121907	0.045832	2.66	0.01	0.032079	0.211736
_Itosettl1_3	0.111816	0.038787	2.88	0.00	0.035794	0.187837
_Iloanbal5_2	0.261558	0.115282	2.27	0.02	0.035608	0.487507

_Ispsetld_7	0.422468	0.061260	6.90	0.00	0.302401	0.542535
_Ispsetld_8	0.430878	0.066417	6.49	0.00	0.300702	0.561053
L	-0.017023	0.003807	-4.47	0.00	-0.024484	-0.009562
_Issrc4to6_3	0.177867	0.054987	3.23	0.00	0.070095	0.285639
_Imor_rent_5	0.254899	0.046212	5.52	0.00	0.164325	0.345472
_Igdscde 999	0.309158	0.057844	5.34	0.00	0.195785	0.422530
Igdscde2 1	0.245661	0.043866	5.60	0.00	0.159685	0.331637
Igdscde2 4	0.233133	0.057805	4.03	0.00	0.119838	0.346429
_Isnbalall_4	0.145024	0.040461	3.58	0.00	0.065721	0.224327
_Iage_3	-0.303639	0.068145	-4.46	0.00	-0.437201	-0.170077
_Inoopen6_4	-0.095922	0.032179	-2.98	0.00	-0.158990	-0.032853
_Imor_rent_6	0.253668	0.049172	5.16	0.00	0.157294	0.350043
_Inetincm_5	-0.209360	0.049091	-4.26	0.00	-0.305576	-0.113144
_Inetincm_6	-0.220894	0.058297	-3.79	0.00	-0.335154	-0.106634
_Igdscde 333	0.352900	0.104540	3.38	0.00	0.148007	0.557794
_Iloanbal1_3	-0.139578	0.045577	-3.06	0.00	-0.228908	-0.050248
_Iloanbal1_2	-0.139254	0.044155	-3.15	0.00	-0.225796	-0.052712
_Ispsetld_5	0.254812	0.053722	4.74	0.00	0.149520	0.360104
_Ispsetld_6	0.272160	0.058307	4.67	0.00	0.157879	0.386440
_Ispsetld_4	0.200907	0.049740	4.04	0.00	0.103419	0.298394
_Itimadd1_4	-0.296149	0.084618	-3.50	0.00	-0.461996	-0.130302
_Itimadd1_2	-0.166997	0.050174	-3.33	0.00	-0.265336	-0.068659
_Iage_2	-0.200532	0.067235	-2.98	0.00	-0.332310	-0.068755
_Iloanbal6_2	0.230300	0.110892	2.08	0.04	0.012956	0.447643
_Itimebank_6	0.138348	0.044191	3.13	0.00	0.051737	0.224960
_Iinternet_1	-0.106486	0.034607	-3.08	0.00	-0.174315	-0.038657
_Itimebank_2	0.084104	0.032901	2.56	0.01	0.019620	0.148588
_Itosettl3_2	-0.090896	0.039121	-2.32	0.02	-0.167572	-0.014220
_Issrc4to6_5	-0.422647	0.170374	-2.48	0.01	-0.756575	-0.088720
_Isocworst_4	0.266782	0.093020	2.87	0.00	0.084467	0.449097
_Isocworst_3	0.307491	0.085371	3.60	0.00	0.140168	0.474815
_Ibrand_2	0.095831	0.038387	2.50	0.01	0.020594	0.171067
_Iinc_surp_4	0.115156	0.041076	2.80	0.01	0.034648	0.195663
_Iinc_surp_5	0.085930	0.033121	2.59	0.01	0.021014	0.150846
_Imor_rent_7	0.325200	0.092805	3.50	0.00	0.143306	0.507094
_Ialcifdet_2	0.374901	0.150134	2.50	0.01	0.080644	0.669158
_Isocbal_3	-0.234449	0.088531	-2.65	0.01	-0.407966	-0.060932
_Imor_rent_4	0.107340	0.044615	2.41	0.02	0.019896	0.194783
_Inetincm_7	-0.322164	0.122215	-2.64	0.01	-0.561701	-0.082628
_Inetincm_4	-0.107154	0.045538	-2.35	0.02	-0.196407	-0.017902
_Ino_store_1	-0.167950	0.069008	-2.43	0.02	-0.303204	-0.032696
_Ispsetld_3	0.112418	0.049099	2.29	0.02	0.016185	0.208651

_Itosettl4_3	0.422811	0.096824	4.37	0.00	0.233040	0.612582
_Iloanbal4_3	0.224974	0.068563	3.28	0.00	0.090592	0.359355
_Itimadd1_8	-0.097284	0.045640	-2.13	0.03	-0.186735	-0.007832
_Iccjgt500_3	0.075616	0.035108	2.15	0.03	0.006805	0.144426
_Issrc4to6_4	0.205182	0.096074	2.14	0.03	0.016880	0.393485
_Itosettl5_3	-0.275432	0.131437	-2.10	0.04	-0.533044	-0.017819
_Ispl6m12_5	-0.096506	0.046484	-2.08	0.04	-0.187614	-0.005398
_Isocnoact_2	0.258502	0.119640	2.16	0.03	0.024011	0.492992
_Igdscde3_2	-0.413087	0.207911	-1.99	0.05	-0.820585	-0.005589
_Ispl6m12_3	-0.313824	0.156231	-2.01	0.05	-0.620031	-0.007616
_Isocsett_2	0.170080	0.086439	1.97	0.05	0.000663	0.339496

Table D.17: Model 2 paying back early on Internet segment

_t	Coef.	Std. Err.	z	$P>\|z\|$	[95% Conf.	Interval]
raw_loanapr1	0.050680	0.007842	6.46	0.00	0.035310	0.066050
_Ispl6m4_4	-0.306019	0.056283	-5.44	0.00	-0.416332	-0.195706
_Ispsetld_9	0.552519	0.068427	8.07	0.00	0.418405	0.686634
_Iinc_surp_3	-0.323689	0.102306	-3.16	0.00	-0.524206	-0.123173
_Iloanbal5_2	0.675272	0.188290	3.59	0.00	0.306231	1.044313
_Icpi_2	0.325011	0.065698	4.95	0.00	0.196246	0.453775
_Iage_6	-0.754534	0.130744	-5.77	0.00	-1.010789	-0.498280
_Itimadd1_9	0.277242	0.072478	3.83	0.00	0.135187	0.419297
_Itimadd1_5	0.313188	0.074870	4.18	0.00	0.166446	0.459930
_Isocsett_3	-0.396365	0.114792	-3.45	0.00	-0.621354	-0.171376
_Ispsetld_7	0.374640	0.094385	3.97	0.00	0.189649	0.559630
_Itimadd1_3	0.216929	0.076805	2.82	0.01	0.066394	0.367464
_Itimadd1_6	0.238031	0.075427	3.16	0.00	0.090196	0.385866
_Itosettl1_3	0.160507	0.065183	2.46	0.01	0.032751	0.288264
_Iinc_surp_4	0.260030	0.075550	3.44	0.00	0.111956	0.408105
_Igdscde3_1	0.652203	0.193959	3.36	0.00	0.272051	1.032355
_Igdscde2_4	0.317455	0.119956	2.65	0.01	0.082345	0.552565
_Ispsetld_8	0.366509	0.107257	3.42	0.00	0.156290	0.576729
_Iloanbal3_4	0.126134	0.071907	1.75	0.08	-0.014802	0.267070
_Itosettl4_3	0.487484	0.151694	3.21	0.00	0.190169	0.784799
_Igdscde2_1	0.255295	0.079930	3.19	0.00	0.098635	0.411954
_Ispsetld_5	0.266116	0.084429	3.15	0.00	0.100637	0.431594
L	-0.012332	0.005648	-2.18	0.03	-0.023402	-0.001262
_Imor_rent_5	0.244225	0.068622	3.56	0.00	0.109729	0.378720
_Isnbalall_4	0.197882	0.067792	2.92	0.00	0.065012	0.330751
_Iinc_surp_6	0.196659	0.085282	2.31	0.02	0.029509	0.363808
_Iinc_surp_5	0.171190	0.059690	2.87	0.00	0.054199	0.288180

_Inoopen6_4	-0.157311	0.052964	-2.97	0.00	-0.261119	-0.053504
_Isnball6m_3	-0.376221	0.162825	-2.31	0.02	-0.695352	-0.057090
_Iage_7	-0.720826	0.163839	-4.40	0.00	-1.041943	-0.399708
_Iloanbal1_3	-0.204607	0.075258	-2.72	0.01	-0.352109	-0.057104
_Iloanbal1_2	-0.160795	0.076062	-2.11	0.04	-0.309874	-0.011717
_Iloanbal4_3	0.176276	0.098146	1.80	0.07	-0.016086	0.368638
_Imor_rent_6	0.189496	0.071170	2.66	0.01	0.050006	0.328986
_Inetincm_5	-0.138751	0.071708	-1.93	0.05	-0.279297	0.001794
_Iage_4	-0.493420	0.103218	-4.78	0.00	-0.695723	-0.291117
_Iage_5	-0.540051	0.118462	-4.56	0.00	-0.772233	-0.307869
_Iage_3	-0.405970	0.104037	-3.90	0.00	-0.609878	-0.202061
_Iage_2	-0.303761	0.103262	-2.94	0.00	-0.506151	-0.101371
Igdscde2 0	-0.727619	0.355528	-2.05	0.04	-1.424442	-0.030797
_Isearches_3	-0.151376	0.075608	-2.00	0.05	-0.299565	-0.003186

Table D.18: Model 2 paying back early on Non-Internet segment

_t	Coef.	Std. Err.	z	$P > \|z\|$	[95% Conf.	Interval]
_Icpi_2	0.382095	0.036945	10.34	0.00	0.309685	0.454505
raw_loanapr1	0.026656	0.005424	4.91	0.00	0.016025	0.037287
_Ispsetld_9	0.565946	0.059444	9.52	0.00	0.449437	0.682455
_Ispl6m4_4	-0.159976	0.041370	-3.87	0.00	-0.241059	-0.078892
_Itosettl4_2	-0.138628	0.059942	-2.31	0.02	-0.256113	-0.021143
_Iage_2	-0.120700	0.088718	-1.36	0.17	-0.294585	0.053185
_Iage_3	-0.210818	0.089318	-2.36	0.02	-0.385879	-0.035757
_Isocsett_2	0.274819	0.074237	3.70	0.00	0.129318	0.420320
_Imortbal_2	0.124466	0.048430	2.57	0.01	0.029546	0.219387
_Iinc_surp_3	-0.402240	0.111049	-3.62	0.00	-0.619892	-0.184588
_Iloanbal6_2	0.424850	0.119903	3.54	0.00	0.189846	0.659855
_Igdscde 999	0.298989	0.062927	4.75	0.00	0.175656	0.422323
_Itimebank_6	0.212495	0.055938	3.80	0.00	0.102858	0.322132
_Itimadd1_4	-0.307114	0.098253	-3.13	0.00	-0.499687	-0.114540
_Isocworst_4	0.333176	0.096801	3.44	0.00	0.143449	0.522902
_Itimadd1_2	-0.146388	0.060099	-2.44	0.02	-0.264180	-0.028596
L	-0.020839	0.004952	-4.21	0.00	-0.030546	-0.011133
_Itosettl2_3	0.155605	0.052357	2.97	0.00	0.052986	0.258223
Igdscde2 1	0.228163	0.050581	4.51	0.00	0.129026	0.327299
_Ispsetld_7	0.422333	0.076852	5.50	0.00	0.271706	0.572960
_Ispsetld_8	0.418580	0.081737	5.12	0.00	0.258379	0.578781
Igdscde2 4	0.214787	0.065202	3.29	0.00	0.086994	0.342580
_Issrc4to6_3	0.203021	0.067625	3.00	0.00	0.070479	0.335564
_Igdscde 333	0.427274	0.135151	3.16	0.00	0.162384	0.692165

_Iage_7	-0.642436	0.113175	-5.68	0.00	-0.864254	-0.420618
_Iage_6	-0.515010	0.094825	-5.43	0.00	-0.700864	-0.329157
_Ispsetld_6	0.305977	0.070670	4.33	0.00	0.167466	0.444489
_Inetincm_6	-0.290031	0.078289	-3.70	0.00	-0.443475	-0.136588
_Ispsetld_4	0.237757	0.061715	3.85	0.00	0.116797	0.358716
_Iage_5	-0.469589	0.093130	-5.04	0.00	-0.652120	-0.287057
_Iage_4	-0.377751	0.087484	-4.32	0.00	-0.549216	-0.206285
_Ispsetld_5	0.218588	0.066689	3.28	0.00	0.087880	0.349295
_Itosettl3_2	-0.133351	0.048649	-2.74	0.01	-0.228700	-0.038001
_Isnbalall_4	0.133125	0.050073	2.66	0.01	0.034985	0.231266
_Iloanbal1_5	0.147953	0.053223	2.78	0.01	0.043637	0.252269
_Issrc4to6_4	0.260601	0.116231	2.24	0.03	0.032793	0.488409
_Itimebank_2	0.093708	0.041231	2.27	0.02	0.012897	0.174520
_Itimadd1_3	0.120416	0.054665	2.20	0.03	0.013275	0.227557
_Inetincm_5	-0.206017	0.062879	-3.28	0.00	-0.329257	-0.082777
_Imor_rent_5	0.207245	0.060226	3.44	0.00	0.089205	0.325285
_Isocnoact_2	0.299761	0.124990	2.40	0.02	0.054786	0.544736
_Iwrstnrev_2	-0.152886	0.056140	-2.72	0.01	-0.262917	-0.042855
_Ispsetld_3	0.123062	0.060740	2.03	0.04	0.004014	0.242109
_Ino_store_1	-0.162763	0.072470	-2.25	0.03	-0.304803	-0.020724
_Imor_rent_6	0.193629	0.063595	3.04	0.00	0.068985	0.318274
_Inetincm_4	-0.136085	0.057742	-2.36	0.02	-0.249258	-0.022912
_Imor_rent_4	0.135223	0.055100	2.45	0.01	0.027228	0.243217
_Iloanbal1_4	0.112162	0.051015	2.20	0.03	0.012175	0.212149
_Inoopen6_4	-0.090290	0.040111	-2.25	0.02	-0.168907	-0.011674
_Isnball6m_2	-0.668508	0.319297	-2.09	0.04	-1.294319	-0.042697
_Imor_rent_7	0.264458	0.125344	2.11	0.04	0.018788	0.510129
_Ibrand_2	0.094824	0.046555	2.04	0.04	0.003578	0.186071
_Ialcifdet_2	0.391284	0.193796	2.02	0.04	0.011451	0.771118

Table D.19: Model 3 Acceptance on Internet segment

_t	Coef.	Std. Err.	z	$P > \lvert z \rvert$	[95% Conf.	Interval]
raw_loanapr1	-0.133517	0.009173	-14.56	0.00	-0.151496	-0.115538
logLXAPR	-0.981518	0.046292	-21.20	0.00	-1.072249	-0.890787
L	0.126633	0.010074	12.57	0.00	0.106889	0.146377
_Icpi_2	-0.617480	0.055009	-11.23	0.00	-0.725296	-0.509664
_Igdscde2_3	-0.408587	0.063017	-6.48	0.00	-0.532097	-0.285076
_Iloanbal4_3	-0.205962	0.081389	-2.53	0.01	-0.365482	-0.046443
_Itosettl1_3	0.245406	0.068616	3.58	0.00	0.110920	0.379892
raw_term	0.009455	0.001877	5.04	0.00	0.005777	0.013134
_Isnball6m_8	0.148790	0.067382	2.21	0.03	0.016724	0.280856

_Inewbus_1	-0.340044	0.103813	-3.28	0.00	-0.543513	-0.136575
_Iloanbal3_3	0.478583	0.137937	3.47	0.00	0.208232	0.748934
_Itimadd1_8	0.207503	0.066722	3.11	0.00	0.076730	0.338277
_Itimebank_4	0.191083	0.074994	2.55	0.01	0.044098	0.338069
_Ispsetld_9	-0.220432	0.063097	-3.49	0.00	-0.344099	-0.096764
_Iage_4	0.214345	0.063122	3.40	0.00	0.090629	0.338061
_Ino_visa_3	0.285653	0.129553	2.20	0.03	0.031734	0.539573
_Igdscde 200	0.868026	0.284561	3.05	0.00	0.310297	1.425756
_Itimebank_7	-0.193162	0.059919	-3.22	0.00	-0.310601	-0.075724
_Itimebank_9	-0.195015	0.066662	-2.93	0.00	-0.325670	-0.064360
_Isnball6m_7	-0.372247	0.124069	-3.00	0.00	-0.615418	-0.129076
_Isnball6m_4	-0.355494	0.131052	-2.71	0.01	-0.612351	-0.098636
_Iworst12_3	-0.349450	0.130593	-2.68	0.01	-0.605408	-0.093492
_Itosettl2_3	0.168083	0.085714	1.96	0.05	0.000086	0.336080
_Ispl6m12_5	-0.141325	0.060593	-2.33	0.02	-0.260085	-0.022565
_Imor_rent_2	0.149741	0.062406	2.40	0.02	0.027429	0.272054
_Ispl6mact_4	-0.459521	0.142304	-3.23	0.00	-0.738431	-0.180611
_Ibrand_2	-0.126119	0.049864	-2.53	0.01	-0.223851	-0.028388
_Ino_amex_1	0.626536	0.282282	2.22	0.03	0.073274	1.179797
_Itimadd1_4	0.273189	0.124560	2.19	0.03	0.029056	0.517322
_Iage_6	0.162680	0.072164	2.25	0.02	0.021242	0.304119
_Isocsett_2	-0.726378	0.255020	-2.85	0.00	-1.226207	-0.226548
_Isocworst_2	-2.225803	0.684320	-3.25	0.00	-3.567045	-0.884561
_Inoopen6_2	-0.152731	0.073755	-2.07	0.04	-0.297288	-0.008174
_Ispl6mact_3	0.432539	0.206229	2.10	0.04	0.028337	0.836740
_Ispl6m12_4	0.223060	0.100301	2.22	0.03	0.026474	0.419647
_Iloanbal3_4	-0.144898	0.065605	-2.21	0.03	-0.273481	-0.016315
_Ispsetld_8	-0.213785	0.102629	-2.08	0.04	-0.414933	-0.012636
_Ispvaldel_2	-0.261749	0.129832	-2.02	0.04	-0.516214	-0.007283
_cons	6.126946	0.346538	17.68	0.00	5.447745	6.806147

Table D.20: Model 3 Acceptance on Non-Internet segment

_t	Coef.	Std. Err.	z	$P > \|z\|$	[95% Conf.	Interval]
_Icpi_2	-1.312067	0.036837	-35.62	0.00	-1.384266	-1.239868
raw_loanapr1	-0.093883	0.005602	-16.76	0.00	-0.104862	-0.082903
_Inewbus_1	-0.480023	0.065438	-7.34	0.00	-0.608280	-0.351767
logLXAPR	-0.763473	0.034492	-22.13	0.00	-0.831076	-0.695869
_Itosettl2_3	0.302263	0.069450	4.35	0.00	0.166144	0.438383
_Igdscde2_3	-0.298025	0.062438	-4.77	0.00	-0.420401	-0.175649
_Iloanbal3_4	-0.306247	0.086849	-3.53	0.00	-0.476468	-0.136025
_Ibrand_2	0.381302	0.039405	9.68	0.00	0.304070	0.458534

L	0.074213	0.009311	7.97	0.00	0.055964	0.092462
_Itosettl1_3	0.357294	0.051957	6.88	0.00	0.255461	0.459128
raw_term	0.010530	0.001565	6.73	0.00	0.007462	0.013598
_Ispl6mact_5	-0.081579	0.045686	-1.79	0.07	-0.171123	0.007964
_Isocsett_2	-0.996011	0.133288	-7.47	0.00	-1.257251	-0.734771
_Imortbal_2	-0.204936	0.044154	-4.64	0.00	-0.291476	-0.118395
_Iloanbal4_2	0.426054	0.131156	3.25	0.00	0.168992	0.683115
_Ispsetld_9	-0.324407	0.055824	-5.81	0.00	-0.433819	-0.214995
_Isnball6m_8	0.294851	0.059973	4.92	0.00	0.177306	0.412396
_Iloanbal2_2	0.690494	0.115789	5.96	0.00	0.463551	0.917437
_Ispvaldel_4	-2.815235	0.787105	-3.58	0.00	-4.357932	-1.272537
_Itosettl4_3	0.105045	0.140587	0.75	0.46	-0.170501	0.380591
_Iloanbal2_4	0.328215	0.070846	4.63	0.00	0.189360	0.467069
_Igdscde 444	0.824142	0.186634	4.42	0.00	0.458346	1.189937
_Itimadd1_4	0.314060	0.085555	3.67	0.00	0.146375	0.481744
_Iloanbal1_8	-0.126945	0.045853	-2.77	0.01	-0.216816	-0.037074
_Ispvaldel_3	0.496752	0.131822	3.77	0.00	0.238386	0.755119
_Ispl6m12_5	-0.229952	0.062574	-3.67	0.00	-0.352596	-0.107309
_Iloanbal6_2	-0.429358	0.118047	-3.64	0.00	-0.660726	-0.197990
_Ino_deps_4	0.120105	0.039606	3.03	0.00	0.042478	0.197732
_Itosettl3_3	0.250815	0.096383	2.60	0.01	0.061909	0.439722
_Ino_store_1	0.252688	0.073291	3.45	0.00	0.109040	0.396336
_Ispl6m12_4	0.400589	0.075642	5.30	0.00	0.252334	0.548845
_Ispl6mact_4	-0.463244	0.106327	-4.36	0.00	-0.671640	-0.254847
_Iloanbal2_3	0.160235	0.064855	2.47	0.01	0.033121	0.287348
_Ispsetld_1	0.197055	0.052663	3.74	0.00	0.093837	0.300273
_Isearches_7	-0.091071	0.036218	-2.51	0.01	-0.162058	-0.020084
_Ialcifdet_2	0.349434	0.142317	2.46	0.01	0.070499	0.628370
_Igdscde 200	1.006511	0.355948	2.83	0.01	0.308865	1.704157
_Iloanbal3_2	-0.133162	0.082787	-1.61	0.11	-0.295421	0.029097
_Ispsetld_2	0.160399	0.052598	3.05	0.00	0.057310	0.263489
_Inoopen6_2	-0.170372	0.062789	-2.71	0.01	-0.293436	-0.047308
_Igdscde2_2	0.181057	0.077835	2.33	0.02	0.028503	0.333612
_Ispl6m12_3	0.504421	0.173686	2.90	0.00	0.164004	0.844839
_Isnball6m_5	0.240316	0.111398	2.16	0.03	0.021979	0.458653
_Isocworst_2	-0.909867	0.325712	-2.79	0.01	-1.548250	-0.271484
_Ino_other_1	0.487456	0.196279	2.48	0.01	0.102756	0.872157
_Itimadd1_9	-0.144725	0.052529	-2.76	0.01	-0.247679	-0.041771
_Itimadd1_3	-0.146237	0.058476	-2.50	0.01	-0.260847	-0.031626
_Iloanbal1_2	0.374997	0.113202	3.31	0.00	0.153124	0.596869
_Igdscde 111	0.469818	0.220995	2.13	0.03	0.036676	0.902960
_Itosettl4_2	-0.278986	0.081377	-3.43	0.00	-0.438482	-0.119491

_Iloanbal4_3	-0.194389	0.097530	-1.99	0.05	-0.385544	-0.003235
_Inetincm_6	-0.150235	0.064226	-2.34	0.02	-0.276116	-0.024354
_Ispl6m4_2	0.201220	0.100117	2.01	0.04	0.004993	0.397447
_Issrc4to6_5	0.362567	0.162964	2.22	0.03	0.043164	0.681971
_Isnball6m_3	-0.656838	0.243934	-2.69	0.01	-1.134939	-0.178736
_Isocworst_4	-0.238795	0.132806	-1.80	0.07	-0.499089	0.021499
_Isnball6m_7	-0.372230	0.136276	-2.73	0.01	-0.639327	-0.105134
_Isnbalall_2	0.242895	0.075466	3.22	0.00	0.094984	0.390806
_Isnbalall_6	0.184583	0.063112	2.92	0.00	0.060887	0.308280
_Iloanbal3_3	0.270665	0.127982	2.11	0.03	0.019824	0.521506
_Iloanbal1_7	0.178155	0.066555	2.68	0.01	0.047709	0.308600
_Isocbal_2	0.249041	0.113525	2.19	0.03	0.026536	0.471546
_cons	5.056947	0.299245	16.90	0.00	4.470438	5.643456

Table D.21: Model 3 default on Internet segment

_t	Coef.	Std. Err.	z	$P > \|z\|$	[95% Conf.	Interval]
L	-0.021079	0.024094	-0.87	0.38	-0.068302	0.026144
raw_loanapr1	0.106788	0.016148	6.61	0.00	0.075139	0.138436
_Icpi_2	1.496162	0.134666	11.11	0.00	1.232221	1.760103
logLXAPR	0.361478	0.089314	4.05	0.00	0.186426	0.536530
_Ibrand_2	-0.518876	0.141040	-3.68	0.00	-0.795310	-0.242441
_Ispvaldel_2	-0.840232	0.236445	-3.55	0.00	-1.303656	-0.376808
_Iinc_surp_6	0.587514	0.169818	3.46	0.00	0.254677	0.920350
_Ino_visa_4	0.743148	0.370407	2.01	0.05	0.017163	1.469133
_Iwrst46al_4	0.469294	0.189254	2.48	0.01	0.098363	0.840224
Igdscde3 4	1.533292	0.524284	2.92	0.00	0.505715	2.560869
_Itimadd1_6	0.399572	0.165500	2.41	0.02	0.075198	0.723946
_Ino_visa_3	-1.263757	0.471880	-2.68	0.01	-2.188624	-0.338890
_Iloanbal3_4	-0.430360	0.143537	-3.00	0.00	-0.711686	-0.149033
_Ispl6m12_4	0.591284	0.194393	3.04	0.00	0.210281	0.972286
_Itimebank_4	-0.449019	0.204577	-2.19	0.03	-0.849983	-0.048055
_Igdscde 999	-1.005572	0.453667	-2.22	0.03	-1.894743	-0.116400
_Issrc4to6_4	0.599881	0.315654	1.90	0.06	-0.018790	1.218552

Table D.22: Model 3 default on Non-Internet segment

_t	Coef.	Std. Err.	z	$P > \|z\|$	[95% Conf.	Interval]
L	0.001599	0.016398	0.10	0.92	-0.030541	0.033739
raw_loanapr1	0.120355	0.008324	14.46	0.00	0.104041	0.136670
_Icpi_2	1.354850	0.093655	14.47	0.00	1.171290	1.538410

logLXAPR	0.179764	0.052347	3.43	0.00	0.077165	0.282363
_Ispl6m12_4	0.688068	0.108847	6.32	0.00	0.474731	0.901404
_Iloanbal4_3	-0.691368	0.100326	-6.89	0.00	-0.888003	-0.494733
_Ispl6m4_3	1.933147	0.398092	4.86	0.00	1.152902	2.713393
_Isocworst_3	1.261844	0.339156	3.72	0.00	0.597111	1.926577
_Iage_7	-0.584655	0.218170	-2.68	0.01	-1.012260	-0.157049
_Ispvaldel_2	-0.425868	0.163649	-2.60	0.01	-0.746614	-0.105122
_Isocbal_3	-1.043541	0.346644	-3.01	0.00	-1.722950	-0.364132
_Isearches_7	-0.272728	0.086237	-3.16	0.00	-0.441748	-0.103707
_Iwrst46al_4	0.352244	0.116681	3.02	0.00	0.123552	0.580935
_Iage_6	-0.295713	0.130775	-2.26	0.02	-0.552028	-0.039398
raw_term	0.010251	0.003680	2.79	0.01	0.003037	0.017464
_Iwrstnrev_3	0.587874	0.253431	2.32	0.02	0.091159	1.084588
_Itimebank_6	0.254988	0.106067	2.40	0.02	0.047100	0.462875
_Imor_rent_7	0.579757	0.227447	2.55	0.01	0.133969	1.025546
_Ispsetld_5	-0.486962	0.178532	-2.73	0.01	-0.836878	-0.137047
_Iinc_surp_2	-0.242150	0.112453	-2.15	0.03	-0.462554	-0.021746
_Itimebank_5	-0.748668	0.299020	-2.50	0.01	-1.334737	-0.162599
_Isnball6m_8	0.288124	0.120229	2.40	0.02	0.052480	0.523769
_Issrc4to6_5	0.540856	0.219979	2.46	0.01	0.109706	0.972006
_Itimebank_4	-0.293287	0.128645	-2.28	0.02	-0.545427	-0.041148
_Igdscde 888	1.130709	0.505529	2.24	0.03	0.139890	2.121528
_Iwrst46al_3	-0.557558	0.263418	-2.12	0.03	-1.073848	-0.041269
_Ispsetld_2	-0.282796	0.125557	-2.25	0.02	-0.528884	-0.036709
_Igdscde 999	-0.362563	0.169278	-2.14	0.03	-0.694342	-0.030783
_Issrc4to6_4	0.442982	0.190737	2.32	0.02	0.069144	0.816820
_Imor_rent_8	0.175816	0.085080	2.07	0.04	0.009062	0.342570
_Ino_mastr_1	-0.248112	0.125559	-1.98	0.05	-0.494204	-0.002021

Table D.23: Model 3 paying back early on Internet segment

_t	Coef.	Std. Err.	z	$P > \|z\|$	[95% Conf.	Interval]
logLXAPR	0.024548	0.046884	0.52	0.60	-0.067342	0.116438
raw_loanapr1	0.048841	0.008653	5.64	0.00	0.031882	0.065800
_Ispl6m4_4	-0.306129	0.056285	-5.44	0.00	-0.416446	-0.195811
_Ispsetld_9	0.553844	0.068476	8.09	0.00	0.419633	0.688055
_Iinc_surp_3	-0.320325	0.102485	-3.13	0.00	-0.521192	-0.119458
_Iloanbal5_2	0.677806	0.188347	3.60	0.00	0.308652	1.046960
_Icpi_2	0.324146	0.065700	4.93	0.00	0.195376	0.452916
_Iage_6	-0.752116	0.130774	-5.75	0.00	-1.008429	-0.495803
_Itimadd1_9	0.275035	0.072600	3.79	0.00	0.132741	0.417329
_Itimadd1_5	0.312150	0.074888	4.17	0.00	0.165371	0.458928

_Isocsett_3	-0.400003	0.114977	-3.48	0.00	-0.625354	-0.174652
_Ispsetld_7	0.376461	0.094449	3.99	0.00	0.191345	0.561578
_Itimadd1_3	0.215000	0.076889	2.80	0.01	0.064301	0.365699
_Itimadd1_6	0.237168	0.075441	3.14	0.00	0.089306	0.385031
_Itosettl1_3	0.158217	0.065316	2.42	0.02	0.030200	0.286234
_Iinc_surp_4	0.259728	0.075547	3.44	0.00	0.111659	0.407797
Igdscde3 1	0.650810	0.193988	3.35	0.00	0.270600	1.031020
Igdscde2 4	0.314638	0.120091	2.62	0.01	0.079265	0.550012
_Ispsetld_8	0.368990	0.107346	3.44	0.00	0.158596	0.579384
_Iloanbal3_4	0.126488	0.071895	1.76	0.08	-0.014423	0.267399
_Itosettl4_3	0.487501	0.151693	3.21	0.00	0.190188	0.784814
Igdscde2 1	0.254127	0.079958	3.18	0.00	0.097413	0.410841
_Ispsetld_5	0.267672	0.084476	3.17	0.00	0.102103	0.433242
L	-0.017337	0.011125	-1.56	0.12	-0.039141	0.004468
_Imor_rent_5	0.243877	0.068623	3.55	0.00	0.109378	0.378376
_Isnbalall_4	0.198686	0.067800	2.93	0.00	0.065799	0.331572
_Iinc_surp_6	0.199383	0.085397	2.33	0.02	0.032009	0.366758
_Iinc_surp_5	0.171051	0.059692	2.87	0.00	0.054057	0.288044
_Inoopen6_4	-0.155911	0.053037	-2.94	0.00	-0.259862	-0.051960
_Isnball6m_3	-0.379491	0.162964	-2.33	0.02	-0.698894	-0.060088
_Iage_7	-0.717821	0.163914	-4.38	0.00	-1.039087	-0.396555
_Iloanbal1_3	-0.204037	0.075264	-2.71	0.01	-0.351552	-0.056522
_Iloanbal1_2	-0.161329	0.076064	-2.12	0.03	-0.310412	-0.012245
_Iloanbal4_3	0.178239	0.098204	1.81	0.07	-0.014238	0.370716
_Imor_rent_6	0.190717	0.071206	2.68	0.01	0.051155	0.330278
_Inetincm_5	-0.139522	0.071722	-1.95	0.05	-0.280094	0.001050
_Iage_4	-0.491473	0.103213	-4.76	0.00	-0.693767	-0.289179
_Iage_5	-0.537734	0.118487	-4.54	0.00	-0.769963	-0.305504
_Iage_3	-0.404384	0.104016	-3.89	0.00	-0.608252	-0.200516
_Iage_2	-0.303818	0.103201	-2.94	0.00	-0.506088	-0.101548
Igdscde2 0	-0.726914	0.355530	-2.04	0.04	-1.423739	-0.030088
_Isearches_3	-0.152417	0.075633	-2.02	0.04	-0.300656	-0.004178

Table D.24: Model 3 paying back early on Non-Internet segment

_t	Coef.	Std. Err.	z	$P > \|z\|$	[95% Conf.	Interval]
logLXAPR	-0.016163	0.034528	-0.47	0.64	-0.083836	0.051511
_Icpi_2	0.381626	0.036965	10.32	0.00	0.309177	0.454076
raw_loanapr1	0.027194	0.005535	4.91	0.00	0.016346	0.038041
_Ispsetld_9	0.564048	0.059584	9.47	0.00	0.447266	0.680831
_Ispl6m4_4	-0.160041	0.041371	-3.87	0.00	-0.241126	-0.078955
_Itosettl4_2	-0.137298	0.060010	-2.29	0.02	-0.254915	-0.019682

_Iage_2	-0.122558	0.088811	-1.38	0.17	-0.296624	0.051508
_Iage_3	-0.212179	0.089370	-2.37	0.02	-0.387341	-0.037017
_Isocsett_2	0.275157	0.074237	3.71	0.00	0.129655	0.420659
_Imortbal_2	0.124587	0.048430	2.57	0.01	0.029666	0.219507
_Iinc_surp_3	-0.403503	0.111105	-3.63	0.00	-0.621264	-0.185742
_Iloanbal6_2	0.424934	0.119909	3.54	0.00	0.189918	0.659951
_Igdscde 999	0.300115	0.062982	4.77	0.00	0.176674	0.423557
_Itimebank_6	0.213128	0.055958	3.81	0.00	0.103451	0.322804
_Itimadd1_4	-0.307426	0.098253	-3.13	0.00	-0.499998	-0.114854
_Isocworst_4	0.332521	0.096818	3.43	0.00	0.142761	0.522281
_Itimadd1_2	-0.146731	0.060105	-2.44	0.02	-0.264534	-0.028928
L	-0.016848	0.009845	-1.71	0.09	-0.036144	0.002448
_Itosettl2_3	0.156676	0.052415	2.99	0.00	0.053944	0.259407
Igdscde2 1	0.228971	0.050613	4.52	0.00	0.129772	0.328171
_Ispsetld_7	0.421281	0.076885	5.48	0.00	0.270590	0.571972
_Ispsetld_8	0.416492	0.081863	5.09	0.00	0.256043	0.576940
Igdscde2 4	0.215440	0.065216	3.30	0.00	0.087619	0.343262
_Issrc4to6_3	0.203889	0.067646	3.01	0.00	0.071305	0.336473
_Igdscde 333	0.429579	0.135234	3.18	0.00	0.164526	0.694632
_Iage_7	-0.644983	0.113305	-5.69	0.00	-0.867057	-0.422909
_Iage_6	-0.517459	0.094972	-5.45	0.00	-0.703602	-0.331317
_Ispsetld_6	0.304223	0.070767	4.30	0.00	0.165522	0.442924
_Inetincm_6	-0.289539	0.078298	-3.70	0.00	-0.443000	-0.136078
_Ispsetld_4	0.236906	0.061743	3.84	0.00	0.115893	0.357919
_Iage_5	-0.471449	0.093219	-5.06	0.00	-0.654154	-0.288743
_Iage_4	-0.379059	0.087534	-4.33	0.00	-0.550622	-0.207496
_Ispsetld_5	0.217520	0.066729	3.26	0.00	0.086733	0.348307
_Itosettl3_2	-0.133808	0.048662	-2.75	0.01	-0.229184	-0.038431
_Isnbalall_4	0.133196	0.050077	2.66	0.01	0.035047	0.231345
_Iloanbal1_5	0.149586	0.053333	2.80	0.01	0.045055	0.254116
_Issrc4to6_4	0.263036	0.116344	2.26	0.02	0.035007	0.491066
_Itimebank_2	0.093939	0.041235	2.28	0.02	0.013121	0.174757
_Itimadd1_3	0.121093	0.054685	2.21	0.03	0.013911	0.228274
_Inetincm_5	-0.206326	0.062886	-3.28	0.00	-0.329580	-0.083072
_Imor_rent_5	0.206930	0.060225	3.44	0.00	0.088892	0.324968
_Isocnoact_2	0.299289	0.124999	2.39	0.02	0.054296	0.544283
_Iwrstnrev_2	-0.153658	0.056164	-2.74	0.01	-0.263738	-0.043579
_Ispsetld_3	0.122249	0.060764	2.01	0.04	0.003155	0.241344
_Ino_store_1	-0.162026	0.072486	-2.24	0.03	-0.304096	-0.019956
_Imor_rent_6	0.193063	0.063603	3.04	0.00	0.068403	0.317723
_Inetincm_4	-0.136683	0.057758	-2.37	0.02	-0.249886	-0.023479
_Imor_rent_4	0.134657	0.055113	2.44	0.02	0.026638	0.242676

_Iloanbal1_4	0.113544	0.051102	2.22	0.03	0.013387	0.213701
_Inoopen6_4	-0.091364	0.040175	-2.27	0.02	-0.170104	-0.012623
_Isnball6m_2	-0.668339	0.319299	-2.09	0.04	-1.294153	-0.042524
_Imor_rent_7	0.261830	0.125485	2.09	0.04	0.015884	0.507776
_Ibrand_2	0.096208	0.046653	2.06	0.04	0.004770	0.187647
_Ialcifdet_2	0.390993	0.193796	2.02	0.04	0.011160	0.770826

Table D.25: Model 4 Acceptance on Internet segment

_t	Coef.	Std. Err.	z	$P > \|z\|$	[95% Conf.	Interval]
raw_loanapr1	-0.254640	0.007491	-33.99	0.00	-0.269323	-0.239958
_Icpi_2	-0.614224	0.054554	-11.26	0.00	-0.721149	-0.507299
_Inewbus_1	-0.347534	0.114356	-3.04	0.00	-0.571668	-0.123400
L	-0.064418	0.004474	-14.40	0.00	-0.073187	-0.055649
_Igdscde2_3	-0.289881	0.081567	-3.55	0.00	-0.449748	-0.130014
_Iloanbal3_4	-0.140200	0.072949	-1.92	0.06	-0.283176	0.002777
_Isnball6m_8	0.055378	0.145931	0.38	0.70	-0.230641	0.341398
_Itosettl1_3	0.154207	0.067971	2.27	0.02	0.020986	0.287428
_Isocsett_2	-0.850096	0.256579	-3.31	0.00	-1.352981	-0.347211
_Itosettl2_3	0.094379	0.087647	1.08	0.28	-0.077406	0.266164
_Iloanbal4_2	0.249436	0.144140	1.73	0.08	-0.033075	0.531946
_Iage_9	-0.132706	0.065461	-2.03	0.04	-0.261007	-0.004404
_Ispsetld_9	-0.192510	0.062879	-3.06	0.00	-0.315751	-0.069269
_Inetincm_9	0.218626	0.079954	2.73	0.01	0.061919	0.375334
_Ibrand_2	-0.222160	0.049213	-4.51	0.00	-0.318615	-0.125705
_Imortbal_2	0.395094	0.265098	1.49	0.14	-0.124489	0.914676
_Iloanbal2_2	0.173438	0.123219	1.41	0.16	-0.068068	0.414944
_Isnrecact_2	-0.077173	0.940674	-0.08	0.94	-1.920861	1.766514
_Ispl6m12_4	0.217551	0.100397	2.17	0.03	0.020775	0.414326
_Ispl6mact_5	-0.085811	0.056529	-1.52	0.13	-0.196605	0.024983
_Itimadd1_9	-0.312264	0.067045	-4.66	0.00	-0.443670	-0.180859
_Ispl6mact_4	-0.527298	0.952074	-0.55	0.58	-2.393329	1.338733
_Isnball6m_5	0.057622	0.193846	0.30	0.77	-0.322308	0.437553
_Itimadd1_3	-0.216025	0.071493	-3.02	0.00	-0.356149	-0.075901
_Ispl6mact_2	-0.503977	0.091246	-5.52	0.00	-0.682815	-0.325138
_Ispvaldel_2	-0.232262	0.126706	-1.83	0.07	-0.480602	0.016078
_Itosettl3_3	0.099824	0.113449	0.88	0.38	-0.122531	0.322179
_Itimadd1_4	0.164018	0.127762	1.28	0.20	-0.086390	0.414427
_Iloanbal6_2	-0.188372	0.130240	-1.45	0.15	-0.443638	0.066895
_Igdscde 444	0.078296	0.324611	0.24	0.81	-0.557930	0.714522
_Igdscde 200	0.823129	0.275108	2.99	0.00	0.283927	1.362331
_Iloanbal1_8	-0.063103	0.055147	-1.14	0.25	-0.171190	0.044984

_Iloanbal3_3	0.376816	0.139961	2.69	0.01	0.102499	0.651134
_Inoopen6_2	-0.202081	0.075071	-2.69	0.01	-0.349217	-0.054944
_Ino_store_1	0.184065	0.222781	0.83	0.41	-0.252579	0.620708
_Ispl6m12_5	-0.102859	0.069481	-1.48	0.14	-0.239039	0.033321
_Inoopen6_3	-0.148003	0.119331	-1.24	0.22	-0.381887	0.085881
_Igdscde2_2	0.116029	0.109091	1.06	0.29	-0.097785	0.329843
_Iloanbal1_2	0.223123	0.121324	1.84	0.07	-0.014668	0.460913
_Iloanbal2_4	-0.035290	0.076279	-0.46	0.64	-0.184794	0.114214
_Itimadd1_5	-0.174481	0.065441	-2.67	0.01	-0.302744	-0.046219
_Itimadd1_6	-0.227128	0.065651	-3.46	0.00	-0.355802	-0.098454
_Imor_rent_8	-0.172536	0.055984	-3.08	0.00	-0.282263	-0.062808
_Isnball6m_7	-0.660607	0.182616	-3.62	0.00	-1.018527	-0.302687
_Isnball6m_4	-0.616171	0.182276	-3.38	0.00	-0.973426	-0.258916
_Itosettl4_2	-0.155133	0.080532	-1.93	0.05	-0.312974	0.002708
_Iloanbal4_3	-0.156092	0.096828	-1.61	0.11	-0.345872	0.033688
_Itimebank_9	-0.330771	0.066331	-4.99	0.00	-0.460776	-0.200765
_Itimebank_7	-0.268634	0.061909	-4.34	0.00	-0.389972	-0.147295
_Isnball6m_3	-0.424392	0.232953	-1.82	0.07	-0.880972	0.032187
_Igdscde 111	0.353347	0.252943	1.40	0.16	-0.142412	0.849106
_Iwrst46al_4	0.276228	0.111266	2.48	0.01	0.058152	0.494304
_Isocbal_2	0.522910	0.212510	2.46	0.01	0.106398	0.939422
_Iloanbal1_7	0.119134	0.070702	1.69	0.09	-0.019439	0.257706
_Isnbalall_2	0.090119	0.072733	1.24	0.22	-0.052434	0.232672
_Isnball6m_6	-0.354342	0.181823	-1.95	0.05	-0.710708	0.002024
_Isnball6m_2	-0.186447	0.185474	-1.01	0.32	-0.549969	0.177075
_Iinc_surp_7	-0.150983	0.065668	-2.30	0.02	-0.279690	-0.022276
_Isnbalall_6	0.018393	0.067400	0.27	0.79	-0.113709	0.150496
_Itimebank_8	-0.091846	0.066500	-1.38	0.17	-0.222183	0.038491
_Isnw12tv_2	0.117013	0.097687	1.20	0.23	-0.074450	0.308475
_Iccjgt500_6	-0.064982	0.045087	-1.44	0.15	-0.153351	0.023387
_Ino_amex_1	0.629394	0.268342	2.35	0.02	0.103454	1.155334
_Ino_other_1	0.251210	0.271379	0.93	0.36	-0.280684	0.783104
_Issrc4to6_2	-0.065877	0.049865	-1.32	0.19	-0.163611	0.031857
_Iwrst46al_5	0.278939	0.168512	1.66	0.10	-0.051339	0.609216
_Ispl6m4_3	-1.648736	0.633871	-2.60	0.01	-2.891100	-0.406372
_Imor_rent_4	-0.105389	0.061184	-1.72	0.09	-0.225307	0.014530
_Ino_deps_4	0.021881	0.049189	0.44	0.66	-0.074526	0.118289
_Ispsetld_8	-0.186939	0.101497	-1.84	0.07	-0.385870	0.011992
_Iage_5	0.167457	0.072601	2.31	0.02	0.025161	0.309753
_Iage_4	0.265553	0.067460	3.94	0.00	0.133334	0.397772
_Iage_6	0.195262	0.075662	2.58	0.01	0.046967	0.343556
_Iinc_surp_9	0.036645	0.078835	0.46	0.64	-0.117868	0.191159

_Iage_7	0.069690	0.080455	0.87	0.39	-0.088000	0.227379
_cons	5.203634	0.428880	12.13	0.00	4.363045	6.044223

Table D.26: Model 4 Acceptance on Non-Internet segment

_t	Coef.	Std. Err.	z	$P > \|z\|$	[95% Conf.	Interval]
raw_loanaprl	-0.148112	0.005116	-28.95	0.00	-0.158139	-0.138085
_Icpi_2	-1.288077	0.036348	-35.44	0.00	-1.359318	-1.216836
_Inewbus_1	-0.415036	0.063486	-6.54	0.00	-0.539466	-0.290605
L	-0.100407	0.004575	-21.95	0.00	-0.109373	-0.091441
_Igdscde2_3	-0.278908	0.060800	-4.59	0.00	-0.398074	-0.159741
_Iloanbal3_4	-0.238483	0.062761	-3.80	0.00	-0.361492	-0.115473
_Isnball6m_8	-0.015294	0.103901	-0.15	0.88	-0.218936	0.188349
_Itosettl1_3	0.281968	0.050739	5.56	0.00	0.182521	0.381415
_Isocsett_2	-1.015261	0.132755	-7.65	0.00	-1.275457	-0.755066
_Itosettl2_3	0.303854	0.066034	4.60	0.00	0.174430	0.433278
_Iloanbal4_2	0.440524	0.126177	3.49	0.00	0.193222	0.687826
_Iage_9	-0.112384	0.060845	-1.85	0.07	-0.231637	0.006869
_Ispsetld_9	-0.332013	0.054846	-6.05	0.00	-0.439509	-0.224517
_Inetincm_9	0.146572	0.047920	3.06	0.00	0.052649	0.240494
_Ibrand_2	0.323813	0.038982	8.31	0.00	0.247409	0.400217
_Imortbal_2	-0.186045	0.046123	-4.03	0.00	-0.276444	-0.095645
_Iloanbal2_2	0.669819	0.111545	6.00	0.00	0.451195	0.888444
_Isnrecact_2	-3.176604	0.669603	-4.74	0.00	-4.489001	-1.864206
_Ispl6m12_4	0.378220	0.074749	5.06	0.00	0.231715	0.524724
_Ispl6mact_5	-0.088585	0.046958	-1.89	0.06	-0.180621	0.003452
_Itimadd1_9	-0.203636	0.055572	-3.66	0.00	-0.312554	-0.094718
_Ispl6mact_4	2.622550	0.678540	3.86	0.00	1.292635	3.952465
_Isnball6m_5	-0.075287	0.138534	-0.54	0.59	-0.346809	0.196235
_Itimadd1_3	-0.217859	0.060918	-3.58	0.00	-0.337256	-0.098462
_Ispl6mact_2	-0.222808	0.073239	-3.04	0.00	-0.366353	-0.079263
_Ispvaldel_2	-0.436648	0.107962	-4.04	0.00	-0.648249	-0.225048
_Itosettl3_3	0.246957	0.088253	2.80	0.01	0.073984	0.419929
_Itimadd1_4	0.333111	0.087721	3.80	0.00	0.161181	0.505041
_Iloanbal6_2	-0.378285	0.113831	-3.32	0.00	-0.601390	-0.155180
_Igdscde 444	0.772789	0.180967	4.27	0.00	0.418101	1.127477
_Igdscde 200	0.971510	0.339643	2.86	0.00	0.305823	1.637198
_Iloanbal1_8	-0.144240	0.045779	-3.15	0.00	-0.233966	-0.054515
_Iloanbal3_3	0.342946	0.118169	2.90	0.00	0.111340	0.574553
_Inoopen6_2	-0.216416	0.062961	-3.44	0.00	-0.339818	-0.093014
_Ino_store_1	0.206363	0.072327	2.85	0.00	0.064604	0.348121
_Ispl6m12_5	-0.219866	0.061645	-3.57	0.00	-0.340689	-0.099044

_Inoopen6_3	-0.305604	0.103320	-2.96	0.00	-0.508107	-0.103101
_Igdscde2_2	0.168698	0.075716	2.23	0.03	0.020297	0.317100
_Iloanbal1_2	0.490642	0.110930	4.42	0.00	0.273223	0.708061
_Iloanbal2_4	0.279077	0.065561	4.26	0.00	0.150580	0.407573
_Itimadd1_5	-0.110920	0.058890	-1.88	0.06	-0.226342	0.004502
_Itimadd1_6	-0.064137	0.056053	-1.14	0.25	-0.173997	0.045724
_Imor_rent_8	-0.111362	0.044013	-2.53	0.01	-0.197625	-0.025099
_Isnball6m_7	-0.660460	0.159221	-4.15	0.00	-0.972527	-0.348393
_Isnball6m_4	-0.500300	0.144591	-3.46	0.00	-0.783693	-0.216906
_Itosettl4_2	-0.311039	0.069672	-4.46	0.00	-0.447594	-0.174483
_Iloanbal4_3	-0.231015	0.081564	-2.83	0.01	-0.390877	-0.071153
_Itimebank_9	-0.028959	0.055224	-0.52	0.60	-0.137195	0.079278
_Itimebank_7	0.002228	0.057827	0.04	0.97	-0.111111	0.115566
_Isnball6m_3	-0.931458	0.250414	-3.72	0.00	-1.422260	-0.440657
_Igdscde 111	0.489132	0.213728	2.29	0.02	0.070233	0.908031
_Iwrst46al_4	0.216627	0.089740	2.41	0.02	0.040741	0.392514
_Isocbal_2	0.256581	0.109653	2.34	0.02	0.041664	0.471497
_Iloanbal1_7	0.168870	0.065241	2.59	0.01	0.040999	0.296740
_Isnbalall_2	0.243426	0.072406	3.36	0.00	0.101513	0.385339
_Isnball6m_6	-0.433124	0.149801	-2.89	0.00	-0.726729	-0.139520
_Isnball6m_2	-0.326548	0.141425	-2.31	0.02	-0.603735	-0.049361
_Iinc_surp_7	-0.025391	0.052595	-0.48	0.63	-0.128476	0.077694
_Isnbalall_6	0.192122	0.061877	3.10	0.00	0.070845	0.313399
_Itimebank_8	-0.065583	0.054564	-1.20	0.23	-0.172526	0.041360
_Isnw12tv_2	0.154682	0.082738	1.87	0.06	-0.007481	0.316846
_Iccjgt500_6	-0.082615	0.038917	-2.12	0.03	-0.158891	-0.006340
_Ino_amex_1	0.215555	0.166320	1.30	0.20	-0.110427	0.541537
_Ino_other_1	0.435893	0.192808	2.26	0.02	0.057997	0.813789
_Issrc4to6_2	-0.079337	0.042021	-1.89	0.06	-0.161698	0.003023
_Iwrst46al_5	0.297321	0.134490	2.21	0.03	0.033725	0.560917
_Ispl6m4_3	-0.553984	0.357421	-1.55	0.12	-1.254515	0.146548
_Imor_rent_4	-0.086901	0.053500	-1.62	0.10	-0.191759	0.017957
_Ino_deps_4	0.127808	0.041544	3.08	0.00	0.046383	0.209232
_Ispsetld_8	-0.129429	0.086600	-1.49	0.14	-0.299162	0.040304
_Iage_5	0.092388	0.058482	1.58	0.11	-0.022236	0.207011
_Iage_4	0.032058	0.054338	0.59	0.56	-0.074443	0.138559
_Iage_6	0.090479	0.059344	1.52	0.13	-0.025833	0.206790
_Iinc_surp_9	0.068322	0.050795	1.35	0.18	-0.031235	0.167878
_Iage_7	0.065728	0.058356	1.13	0.26	-0.048649	0.180104
_cons	5.360419	0.297923	17.99	0.00	4.776501	5.944337

Table D.27: Model 4 default on Internet segment

_t	Coef.	Std. Err.	z	$P > \|z\|$	[95% Conf.	Interval]
raw_loanapr1	0.123936	0.013656	9.08	0.00	0.097172	0.150700
_Icpi_2	1.488281	0.135566	10.98	0.00	1.222576	1.753986
L	0.035435	0.013006	2.72	0.01	0.009943	0.060928
_Ispl6m12_4	0.567339	0.197965	2.87	0.00	0.179335	0.955342
_Iloanbal4_3	-0.352716	0.175408	-2.01	0.04	-0.696510	-0.008923
raw_term	0.019082	0.006007	3.18	0.00	0.007309	0.030855
_Ispvaldel_2	-0.712389	0.251289	-2.83	0.01	-1.204907	-0.219872
_Ispl6m4_3	-0.026417	1.143924	-0.02	0.98	-2.268467	2.215633
_Isocworst_3	0.478129	0.521272	0.92	0.36	-0.543546	1.499804
_Itimebank_4	-0.673665	0.216474	-3.11	0.00	-1.097945	-0.249384
_Iwrst46al_4	0.479822	0.196891	2.44	0.02	0.093923	0.865722
_Itimebank_5	-0.896321	0.467179	-1.92	0.06	-1.811975	0.019332
_Isocbal_3	-0.554029	0.586005	-0.95	0.34	-1.702578	0.594520
_Iinc_surp_2	-0.397275	0.171420	-2.32	0.02	-0.733252	-0.061298
_Igdscde 999	-1.035133	0.462025	-2.24	0.03	-1.940687	-0.129580
_Issrc4to6_4	0.753783	0.322470	2.34	0.02	0.121754	1.385811
_Itimebank_2	-0.311506	0.154381	-2.02	0.04	-0.614087	-0.008925
_Itimebank_3	-0.586266	0.285501	-2.05	0.04	-1.145838	-0.026694
_Ispl6m12_3	0.760222	0.379992	2.00	0.05	0.015452	1.504993
_Isearches_7	-0.018869	0.131530	-0.14	0.89	-0.276664	0.238926
_Itimadd1_6	0.389828	0.169895	2.29	0.02	0.056839	0.722817
_Issrc4to6_5	0.871025	0.318303	2.74	0.01	0.247163	1.494887
_Isnball6m_8	0.396031	0.203959	1.94	0.05	-0.003721	0.795783
_Iloanbal5_2	-0.312743	0.271993	-1.15	0.25	-0.845840	0.220355
_Ispsetld_8	-0.771282	0.423840	-1.82	0.07	-1.601993	0.059428
_Itimadd1_7	-0.145889	0.228202	-0.64	0.52	-0.593157	0.301378
_Iage_7	-0.107800	0.351576	-0.31	0.76	-0.796876	0.581276
Igdscde2 1	-0.251489	0.238721	-1.05	0.29	-0.719373	0.216395
_Iloanbal2_3	0.291530	0.173928	1.68	0.09	-0.049362	0.632423
_Ispsetld_1	0.206549	0.173234	1.19	0.23	-0.132984	0.546082
_Isnw12tv_2	-0.517527	0.273841	-1.89	0.06	-1.054245	0.019192
_Inoopen6_4	-0.264340	0.145888	-1.81	0.07	-0.550274	0.021594
_Imortbal_4	1.095349	0.758045	1.44	0.15	-0.390391	2.581090
_Imor_rent_7	0.273244	0.273993	1.00	0.32	-0.263773	0.810261
Igdscde3 4	1.297023	0.548753	2.36	0.02	0.221487	2.372560

Table D.28: Model 4 default on Non-Internet segment

_t	Coef.	Std. Err.	z	$P > \|z\|$	[95% Conf.	Interval]

raw_loanapr1	0.123843	0.007698	16.09	0.00	0.108755	0.138931
_Icpi_2	1.320721	0.094054	14.04	0.00	1.136379	1.505064
L	0.041120	0.009097	4.52	0.00	0.023291	0.058949
_Ispl6m12_4	0.777575	0.109567	7.10	0.00	0.562827	0.992322
_Iloanbal4_3	-0.609707	0.109348	-5.58	0.00	-0.824024	-0.395389
raw_term	0.015040	0.003543	4.25	0.00	0.008096	0.021983
_Ispvaldel_2	-0.407116	0.164356	-2.48	0.01	-0.729247	-0.084984
_Ispl6m4_3	2.002906	0.395268	5.07	0.00	1.228194	2.777617
_Isocworst_3	1.264901	0.339899	3.72	0.00	0.598711	1.931090
_Itimebank_4	-0.542912	0.133410	-4.07	0.00	-0.804391	-0.281433
_Iwrst46al_4	0.366698	0.117154	3.13	0.00	0.137081	0.596316
_Itimebank_5	-1.061217	0.298802	-3.55	0.00	-1.646857	-0.475576
_Isocbal_3	-1.123392	0.346895	-3.24	0.00	-1.803293	-0.443491
_Iinc_surp_2	-0.233512	0.112098	-2.08	0.04	-0.453220	-0.013803
_Igdscde 999	-0.414522	0.175683	-2.36	0.02	-0.758854	-0.070189
_Issrc4to6_4	0.385773	0.193086	2.00	0.05	0.007331	0.764215
_Itimebank_2	-0.308917	0.100642	-3.07	0.00	-0.506171	-0.111663
_Itimebank_3	-0.430660	0.136402	-3.16	0.00	-0.698003	-0.163318
_Ispl6m12_3	0.502831	0.221400	2.27	0.02	0.068895	0.936767
_Isearches_7	-0.266937	0.086375	-3.09	0.00	-0.436228	-0.097646
_Itimadd1_6	0.165209	0.116058	1.42	0.16	-0.062260	0.392678
_Issrc4to6_5	0.569614	0.221471	2.57	0.01	0.135539	1.003690
_Isnball6m_8	0.354738	0.130857	2.71	0.01	0.098263	0.611214
_Iloanbal5_2	-0.334261	0.168282	-1.99	0.05	-0.664088	-0.004434
_Ispsetld_8	-0.413513	0.248406	-1.66	0.10	-0.900379	0.073353
_Itimadd1_7	-0.297089	0.134978	-2.20	0.03	-0.561641	-0.032536
_Iage_7	-0.512910	0.216928	-2.36	0.02	-0.938082	-0.087738
Igdscde2 1	-0.239360	0.129709	-1.85	0.07	-0.493586	0.014866
_Iloanbal2_3	0.193613	0.114756	1.69	0.09	-0.031305	0.418530
_Ispsetld_1	0.208221	0.108363	1.92	0.06	-0.004165	0.420608
_Isnw12tv_2	-0.258445	0.195902	-1.32	0.19	-0.642407	0.125516
_Inoopen6_4	-0.114703	0.090614	-1.27	0.21	-0.292303	0.062898
_Imortbal_4	0.179472	0.085848	2.09	0.04	0.011213	0.347731
_Imor_rent_7	0.442967	0.228957	1.93	0.05	-0.005781	0.891714
Igdscde3 4	0.306764	0.275048	1.12	0.27	-0.232319	0.845847

Table D.29: Model 4 paying back early on Internet segment

_t	Coef.	Std. Err.	z	$P > \|z\|$	[95% Conf.	Interval]
raw_loanapr1	0.055628	0.008158	6.82	0.00	0.039639	0.071616
_Icpi_2	0.333386	0.066264	5.03	0.00	0.203510	0.463261

_Ispsetld_9	0.615343	0.079271	7.76	0.00	0.459976	0.770711
_Ispl6m4_4	-0.286531	0.056860	-5.04	0.00	-0.397975	-0.175088
_Itosettl4_2	0.006099	0.105333	0.06	0.95	-0.200351	0.212548
_Iinc_surp_3	-0.258632	0.112927	-2.29	0.02	-0.479964	-0.037300
_Iage_6	-0.843473	0.132363	-6.37	0.00	-1.102900	-0.584046
_Iage_7	-0.776320	0.165643	-4.69	0.00	-1.100974	-0.451665
_Isocsett_3	-0.276914	0.149313	-1.85	0.06	-0.569563	0.015735
_Iage_5	-0.618593	0.120472	-5.13	0.00	-0.854713	-0.382472
_Iage_4	-0.569434	0.106189	-5.36	0.00	-0.777560	-0.361308
_Imortbal_2	0.284348	0.263612	1.08	0.28	-0.232322	0.801017
_Itosettl1_3	0.157040	0.069843	2.25	0.03	0.020151	0.293930
_Iloanbal5_2	0.506712	0.205380	2.47	0.01	0.104175	0.909248
_Ispsetld_7	0.436983	0.100889	4.33	0.00	0.239244	0.634722
_Ispsetld_8	0.442091	0.113811	3.88	0.00	0.219026	0.665155
L	-0.013104	0.005771	-2.27	0.02	-0.024414	-0.001793
_Issrc4to6_3	0.147553	0.094520	1.56	0.12	-0.037703	0.332808
_Imor_rent_5	0.314449	0.072591	4.33	0.00	0.172174	0.456725
_Igdscde 999	0.150868	0.119622	1.26	0.21	-0.083587	0.385324
Igdscde2 1	0.258441	0.080944	3.19	0.00	0.099794	0.417087
Igdscde2 4	0.361686	0.120909	2.99	0.00	0.124709	0.598662
_Isnbalall_4	0.149796	0.067774	2.21	0.03	0.016961	0.282630
_Iage_3	-0.465274	0.106725	-4.36	0.00	-0.674451	-0.256096
_Inoopen6_4	-0.120112	0.054586	-2.20	0.03	-0.227098	-0.013125
_Imor_rent_6	0.304391	0.077166	3.94	0.00	0.153148	0.455633
_Inetincm_5	-0.225180	0.077626	-2.90	0.00	-0.377325	-0.073035
_Inetincm_6	-0.151910	0.086035	-1.77	0.08	-0.320535	0.016715
_Igdscde 333	0.273546	0.165935	1.65	0.10	-0.051681	0.598773
_Iloanbal1_3	-0.210312	0.075762	-2.78	0.01	-0.358802	-0.061822
_Iloanbal1_2	-0.170823	0.076164	-2.24	0.03	-0.320102	-0.021545
_Ispsetld_5	0.335652	0.091573	3.67	0.00	0.156174	0.515131
_Ispsetld_6	0.221784	0.102694	2.16	0.03	0.020507	0.423062
_Ispsetld_4	0.134714	0.084666	1.59	0.11	-0.031228	0.300656
_Itimadd1_4	-0.228007	0.166804	-1.37	0.17	-0.554937	0.098923
_Itimadd1_2	-0.192363	0.090699	-2.12	0.03	-0.370130	-0.014597
_Iage_2	-0.328309	0.104115	-3.15	0.00	-0.532371	-0.124247
_Iloanbal6_2	0.165756	0.183640	0.90	0.37	-0.194172	0.525685
_Itimebank_6	0.028010	0.072789	0.38	0.70	-0.114654	0.170675
_Itimebank_2	0.087259	0.054747	1.59	0.11	-0.020043	0.194561
_Itosettl3_2	-0.024020	0.065056	-0.37	0.71	-0.151527	0.103488
_Issrc4to6_5	-0.498108	0.294184	-1.69	0.09	-1.074698	0.078483
_Isocworst_4	0.042037	0.263097	0.16	0.87	-0.473624	0.557698
_Isocworst_3	0.471756	0.244356	1.93	0.05	-0.007174	0.950685

_Ibrand_2	0.122322	0.064609	1.89	0.06	-0.004308	0.248953
_Iinc_surp_4	0.174286	0.074212	2.35	0.02	0.028833	0.319738
_Iinc_surp_5	0.121224	0.058079	2.09	0.04	0.007390	0.235057
_Imor_rent_7	0.313040	0.132692	2.36	0.02	0.052968	0.573111
_Ialcifdet_2	0.303238	0.238565	1.27	0.20	-0.164341	0.770817
_Isocbal_3	-0.497193	0.264105	-1.88	0.06	-1.014829	0.020444
_Imor_rent_4	0.068703	0.076450	0.90	0.37	-0.081136	0.218542
_Inetincm_7	-0.335513	0.172204	-1.95	0.05	-0.673026	0.002000
_Inetincm_4	-0.041142	0.073861	-0.56	0.58	-0.185907	0.103623
_Ino_store_1	-0.111476	0.227058	-0.49	0.62	-0.556502	0.333550
_Ispsetld_3	0.086048	0.083835	1.03	0.31	-0.078265	0.250361
_Itosettl4_3	0.639930	0.182950	3.50	0.00	0.281355	0.998505
_Iloanbal4_3	0.245851	0.117124	2.10	0.04	0.016292	0.475411
_Itimadd1_8	-0.142080	0.082196	-1.73	0.08	-0.303181	0.019020
_Iccjgt500_3	0.080876	0.061888	1.31	0.19	-0.040423	0.202174
_Issrc4to6_4	0.159480	0.172311	0.93	0.36	-0.178244	0.497203
_Itosettl5_3	-0.550424	0.306429	-1.80	0.07	-1.151015	0.050167
_Ispl6m12_5	-0.106735	0.072303	-1.48	0.14	-0.248446	0.034977
_Isocnoact_2	0.589262	0.427374	1.38	0.17	-0.248376	1.426899
Igdscde3 2	-0.601198	0.385620	-1.56	0.12	-1.356999	0.154602
_Ispl6m12_3	-0.333338	0.306520	-1.09	0.28	-0.934105	0.267430
_Isocsett_2	0.212194	0.224009	0.95	0.34	-0.226856	0.651243

Table D.30: Model 4 paying back early on Non-Internet segment

_t	Coef.	Std. Err.	z	$P > \|z\|$	[95% Conf.	Interval]
raw_loanapr1	0.031840	0.005501	5.79	0.00	0.021059	0.042621
_Icpi_2	0.376970	0.037046	10.18	0.00	0.304361	0.449578
_Ispsetld_9	0.557654	0.060507	9.22	0.00	0.439062	0.676246
_Ispl6m4_4	-0.150461	0.040841	-3.68	0.00	-0.230508	-0.070415
_Itosettl4_2	0.016420	0.077619	0.21	0.83	-0.135712	0.168551
_Iinc_surp_3	-0.266164	0.122057	-2.18	0.03	-0.505391	-0.026937
_Iage_6	-0.493233	0.095156	-5.18	0.00	-0.679735	-0.306731
_Iage_7	-0.622869	0.113400	-5.49	0.00	-0.845129	-0.400608
_Isocsett_3	-0.085353	0.070771	-1.21	0.23	-0.224063	0.053356
_Iage_5	-0.455380	0.093406	-4.88	0.00	-0.638453	-0.272308
_Iage_4	-0.364367	0.087781	-4.15	0.00	-0.536413	-0.192320
_Imortbal_2	0.126971	0.048432	2.62	0.01	0.032046	0.221896
_Itosettl1_3	0.082902	0.047011	1.76	0.08	-0.009239	0.175042
_Iloanbal5_2	0.127925	0.140797	0.91	0.36	-0.148031	0.403881
_Ispsetld_7	0.417191	0.077418	5.39	0.00	0.265455	0.568927
_Ispsetld_8	0.422096	0.082103	5.14	0.00	0.261178	0.583015

L	-0.020827	0.005116	-4.07	0.00	-0.030854	-0.010800
_Issrc4to6_3	0.201625	0.067963	2.97	0.00	0.068419	0.334831
_Imor_rent_5	0.213825	0.060533	3.53	0.00	0.095182	0.332467
_Igdscde 999	0.360476	0.068045	5.30	0.00	0.227110	0.493843
Igdscde2 1	0.241203	0.052791	4.57	0.00	0.137736	0.344671
Igdscde2 4	0.210699	0.066219	3.18	0.00	0.080912	0.340486
_Isnbalall_4	0.147495	0.050697	2.91	0.00	0.048132	0.246858
_Iage_3	-0.198157	0.089474	-2.21	0.03	-0.373522	-0.022792
_Inoopen6_4	-0.078698	0.040063	-1.96	0.05	-0.157219	-0.000177
_Imor_rent_6	0.218466	0.064480	3.39	0.00	0.092087	0.344845
_Inetincm_5	-0.198733	0.063937	-3.11	0.00	-0.324047	-0.073420
_Inetincm_6	-0.290827	0.081691	-3.56	0.00	-0.450938	-0.130715
_Igdscde 333	0.432409	0.135280	3.20	0.00	0.167264	0.697553
_Iloanbal1_3	-0.097750	0.057352	-1.70	0.09	-0.210157	0.014657
_Iloanbal1_2	-0.132176	0.054430	-2.43	0.02	-0.238856	-0.025496
_Ispsetld_5	0.216261	0.066712	3.24	0.00	0.085509	0.347014
_Ispsetld_6	0.300590	0.071185	4.22	0.00	0.161069	0.440111
_Ispsetld_4	0.241130	0.061698	3.91	0.00	0.120204	0.362055
_Itimadd1_4	-0.322037	0.098642	-3.26	0.00	-0.515372	-0.128702
_Itimadd1_2	-0.156688	0.060495	-2.59	0.01	-0.275255	-0.038120
_Iage_2	-0.114436	0.088936	-1.29	0.20	-0.288747	0.059875
_Iloanbal6_2	0.280068	0.139796	2.00	0.05	0.006074	0.554062
_Itimebank_6	0.206089	0.056063	3.68	0.00	0.096207	0.315970
_Itimebank_2	0.084634	0.041324	2.05	0.04	0.003640	0.165627
_Itosettl3_2	-0.124450	0.049308	-2.52	0.01	-0.221092	-0.027809
_Issrc4to6_5	-0.377999	0.209567	-1.80	0.07	-0.788743	0.032745
_Isocworst_4	0.323424	0.100705	3.21	0.00	0.126046	0.520801
_Isocworst_3	0.292311	0.093063	3.14	0.00	0.109912	0.474710
_Ibrand_2	0.094813	0.048359	1.96	0.05	0.000031	0.189595
_Iinc_surp_4	0.086053	0.049535	1.74	0.08	-0.011034	0.183140
_Iinc_surp_5	0.068835	0.040512	1.70	0.09	-0.010567	0.148237
_Imor_rent_7	0.336915	0.131665	2.56	0.01	0.078857	0.594974
_Ialcifdet_2	0.406191	0.193923	2.09	0.04	0.026109	0.786272
_Isocbal_3	-0.230412	0.095224	-2.42	0.02	-0.417048	-0.043776
_Imor_rent_4	0.133249	0.055336	2.41	0.02	0.024793	0.241704
_Inetincm_7	-0.304668	0.175873	-1.73	0.08	-0.649374	0.040037
_Inetincm_4	-0.144123	0.058285	-2.47	0.01	-0.258359	-0.029886
_Ino_store_1	-0.164431	0.072813	-2.26	0.02	-0.307141	-0.021721
_Ispsetld_3	0.122144	0.060808	2.01	0.05	0.002963	0.241325
_Itosettl4_3	0.329474	0.115464	2.85	0.00	0.103170	0.555779
_Iloanbal4_3	0.214878	0.085226	2.52	0.01	0.047838	0.381917
_Itimadd1_8	-0.073447	0.055073	-1.33	0.18	-0.181387	0.034494

_Iccjgt500_3	0.068213	0.042809	1.59	0.11	-0.015691	0.152117
_Issrc4to6_4	0.229106	0.116486	1.97	0.05	0.000797	0.457415
_Itosettl5_3	-0.186668	0.146613	-1.27	0.20	-0.474024	0.100689
_Ispl6m12_5	-0.085915	0.061071	-1.41	0.16	-0.205610	0.033781
_Isocnoact_2	0.228590	0.125667	1.82	0.07	-0.017712	0.474892
Igdscde3 2	-0.352478	0.248197	-1.42	0.16	-0.838935	0.133980
_Ispl6m12_3	-0.312152	0.182088	-1.71	0.09	-0.669037	0.044733
_Isocsett_2	0.188781	0.094387	2.00	0.05	0.003786	0.373776

The actual symbolic function used to calculated the expected profit in Matlab is printed below. L is for the loan amount and x is for the interest rate charged.

```
(-1.*L+.98787691278353016554353 8*(2250134495133037/2251799813685248) ^exp(1014624516889591/9007199254740992*x+
7827322445356437/59029581035870565171 2*L-1297761051646769/18014398509481984) *((4500844009866497/4503599627370496)
^exp(5595364410765473/144115188075855872*x-5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496)
 -1.*(8991702818783173/9007199254740992) ^exp(5595364410765473/144115188075855872*x-
5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496)) *(.31622776601683793319988 9*L*(100.+x)
^(1/4)-.9999999999999999999999994e-4*L*(100.+x) ^2/(.9999999999999999999999994e-4*(100.+x)^2-1.)*(.316227766016837933199889
*(100.+x)^(1/4)-1.)) *(-1.+1.3625841381159225709959 8*(100.+x)^(1/12))+.93701913094311458552430 7*(8937376977342113/
9007199254740992) ^exp(1014624516889591/9007199254740992*x+7827322445356437/590295810358705651712*L-1297761051646769/
18014398509481984) *((8481104869158237/9007199254740992) ^exp(5595364410765473/144115188075855872*x-5176894256845849/
295147905179352825856*L+5425581046685411/4503599627370496) -1.*(4212378365670251/4503599627370496) ^exp(5595364410765473
/144115188075855872*x-5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496))
 *(.2154434690031883721759 29e-2*L*(100.+x)^(4/3)-.9999999999999999999999994e-4*L*(100.+x) ^2/
(.9999999999999999999999994e-4*(100.+x)^2-1.)*(.215443469003188372175929e-2*(100.+x)^(4/3)-1.))
 *(-1.+1.362584138115922570995 98*(100.+x)^(1/12))+.91072741116713927898615 6*(8907375347704459/9007199254740992)
 ^exp(1014624516889591/9007199254740992*x+7827322445356437/590295810358705651712*L-1297761051646769/18014398509481984)
 *((8085879414211337/9007199254740992) ^exp(5595364410765473/144115188075855872*x-5176894256845849/295147905179352825856
*L+5425581046685411/4503599627370496) -1.*(4017172091621691/4503599627370496) ^exp(5595364410765473/144115188075855872*x
-5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496)) *(.146779926762206954092052e-3*L*(100.+x)
^(23/12)-.9999999999999999999999994e-4*L*(100.+x) ^2/(.9999999999999999999999994e-4*(100.+x)^2-1.)
*(.146779926762206954092052e-3*(100.+x)^(23/12)-1.)) *(-1.+1.362584138115922570995 98*(100.+x)^(1/12))
+.97987661464877733278930 8*(8992420872707761/9007199254740992) ^exp(1014624516889591/9007199254740992*x
+7827322445356437/590295810358705651712*L-1297761051646769/18014398509481984) *((8971003374175853/9007199254740992)
 ^exp(5595364410765473/144115188075855872*x-5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496)
 -1.*(8945078853280857/9007199254740992) ^exp(5595364410765473/144115188075855872*x
-5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496))
 *(.146779926762206954092052*L*(100.+x)^(5/12)-.9999999999999999999999994e-4*L*(100.+x)
 ^2/(.9999999999999999999999994e-4*(100.+x)^2-1.)*(.146779926762206954092052*(100.+x)^(5/12)-1.))
 *(-1.+1.362584138115922570995 98*(100.+x)^(1/12))+.24596715799421411299996 6*(8971003374175853/9007199254740992)
 ^exp(5595364410765473/144115188075855872*x-5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496)
 *((2250134495133037/2251799813685248) ^exp(1014624516889591/9007199254740992*x
+7827322445356437/590295810358705651712*L-1297761051646769/18014398509481984) -1.*(4498200802209197/4503599627370496)
 ^exp(1014624516889591/9007199254740992*x+7827322445356437/590295810358705651712*L-1297761051646769/18014398509481984))
 *(.316227766016837933199889*L*(100.+x)^(1/4)-.9999999999999999999999994e-4*L*(100.+x)
 ^2/(.9999999999999999999999994e-4*(100.+x)^2-1.)*(.316227766016837933199889*(100.+x)^(1/4)-1.))
 +.98386863197685645199986 4*(4498200802209197/4503599627370496) ^exp(1014624516889591/9007199254740992*x
+7827322445356437/590295810358705651712*L-1297761051646769/18014398509481984) *((8991702818783173/9007199254740992)
 ^exp(5595364410765473/144115188075855872*x-5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496)
 -1.*(8971003374175853/9007199254740992) ^exp(5595364410765473/144115188075855872*x
-5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496))
 *(.215443469003188372175929*L*(100.+x)^(1/3)-.9999999999999999999999994e-4*L*(100.+x)
 ^2/(.9999999999999999999999994e-4*(100.+x)^2-1.)*(.215443469003188372175929*(100.+x)^(1/3)-1.))
 *(-1.+1.362584138115922570995 98*(100.+x)^(1/12))+.92943070336027704538154 7
*(8932128122048405/9007199254740992) ^exp(1014624516889591/9007199254740992*x+7827322445356437
/590295810358705651712*L-1297761051646769/18014398509481984) *((8364791122158079/9007199254740992)
 ^exp(5595364410765473/144115188075855872*x-5176894256845849/295147905179352825856*L+5425581046685411
/4503599627370496) -1.*(4149995179171859/4503599627370496) ^exp(5595364410765473/144115188075855872*x
-5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496))
 *(.1000000000000000000000000e-2*L*(100.+x)^(3/2)-.9999999999999999999999994e-4*L*(100.+x)
 ^2/(.9999999999999999999999994e-4*(100.+x)^2-1.)*(.1000000000000000000000000e-2*(100.+x)^(3/2)-1.))
 *(-1.+1.362584138115922570995 98*(100.+x)^(1/12))+.99190152332621026779764 9*(2251126683166943/2251799813685248)
 ^exp(1014624516889591/9007199254740992*x+7827322445356437/590295810358705651712*L-1297761051646769/18014398509481984)
 *((9005957161963763/9007199254740992) ^exp(5595364410765473/144115188075855872*x
-5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496) -1.*(4500844009866497/4503599627370496)
 ^exp(5595364410765473/144115188075855872*x-5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496))
 *(.464158883361277889241008*L*(100.+x)^(1/6)-.9999999999999999999999994e-4*L*(100.+x)
 ^2/(.9999999999999999999999994e-4*(100.+x)^2-1.)*(.464158883361277889241008*(100.+x)^(1/6)-1.))
 *(-1.+1.362584138115922570995 98*(100.+x)^(1/12))+.94851809857858287786767 1*(559183992747409/562949953421312)
 ^exp(1014624516889591/9007199254740992*x+7827322445356437/590295810358705651712*L-1297761051646769/18014398509481984)
 *((8653328193444415/9007199254740992) ^exp(5595364410765473/144115188075855872*x
-5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496) -1.*(4300974033223813/4503599627370496)
 ^exp(5595364410765473/144115188075855872*x-5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496))
 *(.681292069057961285497988e-2*L*(100.+x)^(13/12)-.9999999999999999999999994e-4*L*(100.+x)
 ^2/(.9999999999999999999999994e-4*(100.+x)^2-1.)*(.681292069057961285497988e-2*(100.+x)^(13/12)-1.))
 *(-1.+1.362584138115922570995 98*(100.+x)^(1/12))+.23809559032804800621436 5*(8653328193444415/9007199254740992)
```

```
^exp(5595364410765473/144115188075855872*x-5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496)
*((8953690005984367/9007199254740992) ^exp(1014624516889591/9007199254740992*x+7827322445356437/
590295810358705651712*L-1297761051646769/18014398509481984) -1.*(559399050888615/562949953421312)
^exp(1014624516889591/9007199254740992*x+7827322445356437/590295810358705651712*L-1297761051646769
/18014398509481984)) *(.1467799267622069540920052e-1*L*(100.+x)^(11/12)-.9999999999999999999999994e-4*L*(100.+x)
^2/(.9999999999999999999999994e-4*(100.+x)^2-1.)*(.1467799267622069540920052e-1*(100.+x)^(11/12)-1.))
+.2352091367206442086717260*(8481104869158237/9007199254740992) ^exp(5595364410765473/144115188075855872
*x-5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496) *((8943931065879825/9007199254740992)
^exp(1014624516889591/9007199254740992*x+7827322445356437/590295810358705651712*L-1297761051646769/18014398509481984)
-1.*(1117463989200615/1125899906842624) ^exp(1014624516889591/9007199254740992*x+7827322445356437
/590295810358705651712*L-1297761051646769/18014398509481984)) *(.4641588833612778892410080e-2*L*(100.+x)^(7/6)
-.9999999999999999999999994e-4*L*(100.+x) ^2/(.9999999999999999999999994e-4*(100.+x)^2-1.)*(.4641588833612778892410080e
-2*(100.+x)^(7/6)-1.)) +.2410174660901514733546730*(2197633039383027/2251799813685248)
^exp(5595364410765473/144115188075855872*x-5176894256845849/295147905179352825856*L+5425581046685411
/4503599627370496) *((2244442328010009/2251799813685248) ^exp(1014624516889591/9007199254740992*x
+7827322445356437/590295810358705651712*L-1297761051646769/18014398509481984) -1.*(4484484909400055/4503599627370496)
^exp(1014624516889591/9007199254740992*x+7827322445356437/590295810358705651712*L-1297761051646769/18014398509481984))
*(.4641588833612778892410080e-1*L*(100.+x)^(2/3)-.9999999999999999999999994e-4*L*(100.+x)
^2/(.9999999999999999999999994e-4*(100.+x)^2-1.)*(.4641588833612778892410080e-1*(100.+x)^(2/3)-1.))
+.2479753808315525669494120*(4500844009866497/4503599627370496) ^exp(5595364410765473/144115188075855872*x
-5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496) *((4503497800982921/4503599627370496)
^exp(1014624516889591/9007199254740992*x+7827322445356437/590295810358705651712*L-1297761051646769/18014398509481984)
-1.*(2251126683166943/2251799813685248) ^exp(1014624516889591/9007199254740992*x+7827322445356437/
590295810358705651712*L-1297761051646769/18014398509481984)) *(.6812920690579612854979880*L*(100.+x)
^(1/12)-.9999999999999999999999994e-4*L*(100.+x) ^2/(.9999999999999999999999994e-4*(100.+x)^2-1.)
*(.6812920690579612854979880*(100.+x)^(1/12)-1.)) +.2449691536621943331973270*(8945078853280857/9007199254740992
^exp(5595364410765473/144115188075855872*x-5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496)
*((4498200802209197/4503599627370496) ^exp(1014624516889591/9007199254740992*x+7827322445356437/
590295810358705651712*L-1297761051646769/18014398509481984) -1.*(8992420872707761/9007199254740992)
^exp(1014624516889591/9007199254740992*x+7827322445356437/590295810358705651712*L-1297761051646769/18014398509481984))
*(.2154434690031883721759290*L*(100.+x)^(1/3)-.9999999999999999999999994e-4*L*(100.+x)
^2/(.9999999999999999999999994e-4*(100.+x)^2-1.)*(.2154434690031883721759290*(100.+x)^(1/3)-1.))
+.2390655917630444738783730*(8701426817608717/9007199254740992)
^exp(5595364410765473/144115188075855872*x-5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496)
*((8961090140748077/9007199254740992) ^exp(1014624516889591/9007199254740992*x
+7827322445356437/590295810358705651712*L-1297761051646769/18014398509481984)
-1.*(8953690005984367/9007199254740992) ^exp(1014624516889591/9007199254740992*x+7827322445356437/
590295810358705651712*L-1297761051646769/18014398509481984)) *(.2154434690031883721759290e-1*L*(100.+x)^(5/6)
-.9999999999999999999999994e-4*L*(100.+x) ^2/(.9999999999999999999999994e-4*(100.+x)^2-1.)
*(.2154434690031883721759290e-1*(100.+x)^(5/6)-1.)) +.2314148915737976097988080*(8239055304305507/9007199254740992)
^exp(5595364410765473/144115188075855872*x-5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496)
*((8932128122048405/9007199254740992) ^exp(1014624516889591/9007199254740992*x+7827322445356437/
590295810358705651712*L-1297761051646769/18014398509481984) -1.*(2231562280773709/2251799813685248)
^exp(1014624516889591/9007199254740992*x+7827322445356437/590295810358705651712*L-1297761051646769/18014398509481984))
*(.1000000000000000000000000e-2*L*(100.+x)^(3/2)-.9999999999999999999999994e-4*L*(100.+x)
^2/(.9999999999999999999999994e-4*(100.+x)^2-1.)*(.1000000000000000000000000e-2*(100.+x)^(3/2)-1.))
+.9408365468825768346869020*(1117463989200615/1125899906842624) ^exp(1014624516889591/9007199254740992*x
+7827322445356437/590295810358705651712*L-1297761051646769/18014398509481984) *((8540854395630539/9007199254740992)
^exp(5595364410765473/144115188075855872*x-5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496)
-1.*(8481104869158237/9007199254740992) ^exp(5595364410765473/144115188075855872*x
-5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496)) *(.3162277660168379331998890e-2*L*(100.+x)
^(5/4)-.9999999999999999999999994e-4*L*(100.+x) ^2/(.9999999999999999999999994e-4*(100.+x)^2-1.)
*(.3162277660168379331998890e-2*(100.+x)^(5/4)-1.)) *(-1.+1.362584138115922570995980*(100.+x)^(1/12))
+.9256595662951904391952330*(2231562280773709/2251799813685248) ^exp(1014624516889591/9007199254740992*x
+7827322445356437/590295810358705651712*L-1297761051646769/18014398509481984) *((4149995179171859/4503599627370496)
^exp(5595364410765473/144115188075855872*x-5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496)
-1.*(8239055304305507/9007199254740992) ^exp(5595364410765473/144115188075855872*x-5176894256845849/
295147905179352825856*L+5425581046685411/4503599627370496)) *(.6812920690579612854979880e-3*L*(100.+x)^(19/12)
-.9999999999999999999999994e-4*L*(100.+x) ^2/(.9999999999999999999999994e-4*(100.+x)^2-1.)
*(.6812920690579612854979880e-3*(100.+x)^(19/12)-1.)) *(-1.+1.362584138115922570995980*(100.+x)^(1/12))
+.9219037304971450698450190*(1115049012713431/1125899906842624) ^exp(1014624516889591/9007199254740992*x
+7827322445356437/590295810358705651712*L-1297761051646769/18014398509481984) *((8239055304305507/9007199254740992)
^exp(5595364410765473/144115188075855872*x-5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496)
-1.*(8190338516056353/9007199254740992) ^exp(5595364410765473/144115188075855872*x
-5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496)) *(.4641588833612778892410080e
-3*L*(100.+x)^(5/3)-.9999999999999999999999994e-4*L*(100.+x) ^2/(.9999999999999999999999994e-4*(100.+x)^2-1.)
*(.4641588833612778892410080e-3*(100.+x)^(5/3)-1.)) *(-1.+1.362584138115922570995980*(100.+x)^(1/12))
+.9959425301322412487898080*(4503497800982921/4503599627370496) ^exp(1014624516889591/9007199254740992*x
+7827322445356437/590295810358705651712*L-1297761051646769/18014398509481984)
*(1-1.*(9005957161963763/9007199254740992) ^exp(5595364410765473/144115188075855872*x
-5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496))
*(.6812920690579612854979880*L*(100.+x)^(1/12)-.9999999999999999999999994e-4*L*(100.+x)
^2/(.9999999999999999999999994e-4*(100.+x)^2-1.)*(.6812920690579612854979880*(100.+x)^(1/12)-1.))
*(-1.+1.362584138115922570995980*(100.+x)^(1/12))+.9446695149745764530539750*(8943931065879825/9007199254740992)
^exp(1014624516889591/9007199254740992*x+7827322445356437/590295810358705651712*L-1297761051646769/18014398509481984)
*((4300974033223813/4503599627370496) ^exp(5595364410765473/144115188075855872*x
-5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496) -1.*(8540854395630539/9007199254740992)
^exp(5595364410765473/144115188075855872*x-5176894256845849/295147905179352825856*L
+5425581046685411/4503599627370496)) *(.4641588833612778892410080e-2*L*(100.+x)^(7/6)-.9999999999999999999999994e-4
*L*(100.+x) ^2/(.9999999999999999999999994e-4*(100.+x)^2-1.)*(.4641588833612778892410080e-2*(100.+x)^(7/6)-1.))
*(-1.+1.362584138115922570995980*(100.+x)^(1/12))+.9070321621385866765356990*(2223127196297631/2251799813685248)
^exp(1014624516889591/9007199254740992*x+7827322445356437/590295810358705651712*L-1297761051646769/18014398509481984)
*((4017172091621691/4503599627370496) ^exp(5595364410765473/144115188075855872*x
-5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496) -1.*(3996675083817621/4503599627370496)
^exp(5595364410765473/144115188075855872*x-5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496))
*(.1000000000000000000000000e-3*L*(100.+x)^2-.9999999999999999999999994e-4*L*(100.+x) ^2/(.9999999999999999999999994e-4*
```

```
(100.+x)^2-1.)*(.1000000000000000000000000e-3*(100.+x)^2-1.)) *(-1.+1.362584138115922570995 98*(100.+x)^(1/12))
+.964069864360605893418690*(4484484909400055/4503599627370496) ^exp(1014624516889591/9007199254740992*x
+7827322445356437/590295810358705651712*L-1297761051646769/18014398509481984) *((8837970463919045/9007199254740992)
 ^exp(5595364410765473/144115188075855872*x-5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496)
 -1.*(2197633039383027/2251799813685248) ^exp(5595364410765473/144115188075855872*x-5176894256845849
/295147905179352825856*L+5425581046685411/4503599627370496)) *(.316227766016837933199889e-1*L*(100.+x)^(3/4)
-.99999999999999999999999994e-4*L*(100.+x) ^2/(.99999999999999999999999994e-4*(100.+x)^2-1.)
*(.316227766016837933199889e-1*(100.+x)^(3/4)-1.)) *(-1.+1.36258413811592257099598*(100.+x)^(1/12))
+.236167378743644113263494*(8540854395630539/9007199254740992) ^exp(5595364410765473/144115188075855872*x
-5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496) *((559183992747409/562949953421312)
 ^exp(1014624516889591/9007199254740992*x+7827322445356437/590295810358705651712*L-1297761051646769/18014398509481984)
-1.*(8943931065879825/9007199254740992) ^exp(1014624516889591/9007199254740992*x+7827322445356437/590295810358705651712*L
-1297761051646769/18014398509481984)) *(.681292069057961285497988e-2*L*(100.+x)^(13/12)-.99999999999999999999999994e
-4*L*(100.+x) ^2/(.99999999999999999999999994e-4*(100.+x)^2-1.)*(.681292069057961285497988e-2*(100.+x)^(13/12)-1.))
+.914437714640234204740341*(4457727132502003/4503599627370496) ^exp(1014624516889591/9007199254740992*x+
7827322445356437/590295810358705651712*L-1297761051646769/18014398509481984) *((2033127443706281/2251799813685248)
 ^exp(5595364410765473/144115188075855872*x-5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496)
 -1.*(8085879414211337/9007199254740992) ^exp(5595364410765473/144115188075855872*x
-5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496)) *(.215443469003188372175929e-3*
L*(100.+x)^(11/6)-.99999999999999999999999994e-4*L*(100.+x) ^2/(.99999999999999999999999994e-4*(100.+x)^2-1.)*
(.215443469003188372175929e-3*(100.+x)^(11/6)-1.)) *(-1.+1.36258413811592257099598*(100.+x)^(1/12)
)+.230475932624286267461255*(8190338516056353/9007199254740992) ^exp(5595364410765473/144115188075855872*x
-5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496) *((2231562280773709/2251799813685248)
 ^exp(1014624516889591/9007199254740992*x+7827322445356437/590295810358705651712*L-1297761051646769/18014398509481984)
 -1.*(1115049012713431/1125899906842624) ^exp(1014624516889591/9007199254740992*x+7827322445356437/
590295810358705651712*L-1297761051646769/18014398509481984)) *(.681292069057961285497988e-3*L*(100.+x)^(19/12)
-.99999999999999999999999994e-4*L*(100.+x) ^2/(.99999999999999999999999994e-4*(100.+x)^2-1.)*(.681292069057961285497988e-3
*(100.+x)^(19/12)-1.)) +.229540783472419629805850*(2033127443706281/2251799813685248)
^exp(5595364410765473/144115188075855872*x-5176894256845849/295147905179352825856*L+5425581046685411
/4503599627370496) *((1115049012713431/1125899906842624) ^exp(1014624516889591/9007199254740992*x+
7827322445356437/590295810358705651712*L-1297761051646769/18014398509481984) -1.*(1114792330052669/
1125899906842624) ^exp(1014624516889591/9007199254740992*x+
7827322445356437/590295810358705651712*L-1297761051646769/18014398509481984)) *(.464158883361277889241008e-3
*L*(100.+x)^(5/3)-.99999999999999999999999994e-4*L*(100.+x) ^2/(.99999999999999999999999994e-4*(100.+x)^2-1.)
*(.464158883361277889241008e-3*(100.+x)^(5/3)-1.)) +.226758040534646669133925*(3996675083817621/4503599627370496)
^exp(5595364410765473/144115188075855872*x-5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496)
 *((8907375347704459/9007199254740992) ^exp(1014624516889591/9007199254740992*x+7827322445356437/
590295810358705651712*L-1297761051646769/18014398509481984) -1.*(2223127196297631/2251799813685248)
 ^exp(1014624516889591/9007199254740992*x+7827322445356437/590295810358705651712*L-1297761051646769/18014398509481984))
 *(.146779926762206954092052e-3*L*(100.+x)^(23/12)-.99999999999999999999999994e-4*L*(100.+x)
 ^2/(.99999999999999999999999994e-4*(100.+x)^2-1.)*(.146779926762206954092052e-3*(100.+x)^(23/12)-1.))
 +.237129524644645719466918*(4300974033223813/4503599627370496) ^exp(5595364410765473/144115188075855872*x-
5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496) *((559399050888615/562949953421312)
 ^exp(1014624516889591/9007199254740992*x+7827322445356437/590295810358705651712*L-1297761051646769/18014398509481984)
 -1.*(559183992747409/562949953421312) ^exp(1014624516889591/9007199254740992*x+7827322445356437/
590295810358705651712*L-1297761051646769/18014398509481984)) *(L*(1.+.1000000000000000000000000e-1*x)
-.99999999999999999999999994e-6*L*(100.+x)^2/(.99999999999999999999999994e-4*(100.+x) ^2-1.)*x)+
.234254782735778646381077*(4212378365670251/4503599627370496) ^exp(5595364410765473/144115188075855872*x
-5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496) *((1117463989200615/1125899906842624)
^exp(1014624516889591/9007199254740992*x+7827322445356437/590295810358705651712*L-1297761051646769/18014398509481984)
 -1.*(8937376977342113/9007199254740992) ^exp(1014624516889591/9007199254740992*x+7827322445356437
/590295810358705651712*L-1297761051646769/18014398509481984)) *(.316227766016837933199889e-2*L*(100.+x)
^(5/4)-.99999999999999999999999994e-4*L*(100.+x) ^2/(.99999999999999999999999994e-4*(100.+x)^2-1.)
*(.316227766016837933199889e-2*(100.+x)^(5/4)-1.)) +.246969228195882541385885*(8991702818783173/9007199254740992)
 ^exp(5595364410765473/144115188075855872*x-5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496)
 *((2251126683166943/2251799813685248) ^exp(1014624516889591/9007199254740992*x+7827322445356437
/590295810358705651712*L-1297761051646769/18014398509481984) -1.*(2250134495133037/2251799813685248)
 ^exp(1014624516889591/9007199254740992*x+7827322445356437/590295810358705651712*L-1297761051646769/18014398509481984))
 *(.464158883361277889241008*L*(100.+x)^(1/6)-.99999999999999999999999994e-4*L*(100.+x)
^2/(.99999999999999999999999994e-4*(100.+x)^2-1.)*(.464158883361277889241008*(100.+x)^(1/6)-1.))
+.227681852791784819746539*(4017172091621691/4503599627370496) ^exp(5595364410765473/144115188075855872*x-
5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496) *((4457727132502003/4503599627370496)
^exp(1014624516889591/9007199254740992*x+7827322445356437/590295810358705651712*L-1297761051646769/18014398509481984)
 -1.*(8907375347704459/9007199254740992) ^exp(1014624516889591/9007199254740992*x+7827322445356437/
590295810358705651712*L-1297761051646769/18014398509481984)) *(.215443469003188372175929e-3*L*(100.+x)
^(11/6)-.99999999999999999999999994e-4*L*(100.+x) ^2/(.99999999999999999999999994e-4*(100.+x)^2-1.)
*(.215443469003188372175929e-3*(100.+x)^(11/6)-1.)) +.240039544983887117151997*(4371919373261769/4503599627370496)
 ^exp(5595364410765473/144115188075855872*x-5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496)
 *((4484484909400055/4503599627370496) ^exp(1014624516889591/9007199254740992*x+7827322445356437/
590295810358705651712*L-1297761051646769/18014398509481984) -1.*(8961090140748077/9007199254740992)
^exp(1014624516889591/9007199254740992*x+7827322445356437/590295810358705651712*L-1297761051646769/18014398509481984))
 *(.316227766016837933199889e-1*L*(100.+x)^(3/4)-.99999999999999999999999994e-4*L*(100.+x)
^2/(.99999999999999999999999994e-4*(100.+x)^2-1.)*(.316227766016837933199889e-1*(100.+x)^(3/4)-1.))
 +.241999371247002750544874*(8837970463919045/9007199254740992) ^exp(5595364410765473/144115188075855872*x
-5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496) *((8983141115603571/9007199254740992)
 ^exp(1014624516889591/9007199254740992*x+7827322445356437/590295810358705651712*L-1297761051646769/18014398509481984)
 -1.*(2244442328010009/2251799813685248) ^exp(1014624516889591/9007199254740992*x+7827322445356437/
590295810358705651712*L-1297761051646769/18014398509481984)) *(.681292069057961285497988e-1*L*(100.+x)
^(7/12)-.99999999999999999999999994e-4*L*(100.+x) ^2/(.99999999999999999999999994e-4*(100.+x)^2-1.)
*(.681292069057961285497988e-1*(100.+x)^(7/12)-1.)) +.233304301013449851600709*(8364791122158079/9007199254740992)
 ^exp(5595364410765473/144115188075855872*x-5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496)
*((8937376977342113/9007199254740992) ^exp(1014624516889591/9007199254740992*x+7827322445356437/
590295810358705651712*L-1297761051646769/18014398509481984) -1.*(2233726327948655/2251799813685248)
^exp(1014624516889591/9007199254740992*x+7827322445356437/590295810358705651712*L-1297761051646769/18014398509481984))
 *(.215443469003188372175929e-2*L*(100.+x)^(4/3)-.99999999999999999999999994e-4*L*(100.+x)
^2/(.99999999999999999999999994e-4*(100.+x)^2-1.)*(.215443469003188372175929e-2*(100.+x)^(4/3)-1.))
```

```
+.2323576758420692613453B7*(4149995179171859/4503599627370496) ^exp(5595364410765473/144115188075855872*
x-5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496) *((2233726327948655/2251799813685248)
^exp(1014624516889591/9007199254740992*x+7827322445356437/590295810358705651712*L-1297761051646769/18014398509481984)
 -1.*(8932128122048405/9007199254740992) ^exp(1014624516889591/9007199254740992*x+7827322445356437
/590295810358705651712*L-1297761051646769/18014398509481984)) *(.14677992676220695409205 2e-2*L*(100.+x)^
(17/12)-.99999999999999999999999994e-4*L*(100.+x) ^2/(.99999999999999999999999994e-4*(100.+x)^2-1.)
*(.14677992676220695409205 2e-2*(100.+x)^(17/12)-1.)) +.95626236705217789551349l*(8953690005984367/9007199254740992)
 ^exp(1014624516889591/9007199254740992*x+7827322445356437/590295810358705651712*L-1297761051646769/18014398509481984)
 *((4371919373261769/4503599627370496) ^exp(5595364410765473/144115188075855872*x-5176894256845849/
295147905179352825856*L+5425581046685411/4503599627370496) -1.*(8701426817608717/9007199254740992)
^exp(5595364410765473/144115188075855872*x-5176894256845849/295147905179352825856*L+5425581046685411/
4503599627370496)) *(.14677992676220695409205 2e-1*L*(100.+x)^(11/12)-.99999999999999999999999994e-4*L*(100.+x)
 ^2/(.99999999999999999999999994e-4*(100.+x)^2-1.)*(.14677992676220695409205 2e-1*(100.+x)^(11/12)-1.))
 *(-1.+1.3625841381159225709959 8*(100.+x)^(1/12))+.24397519870267961635148 0*(4457337796314217/4503599627370496)
^exp(5595364410765473/144115188075855872*x-5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496)
 *((8992420872707761/9007199254740992) ^exp(1014624516889591/9007199254740992*x+7827322445356437/590295810358705651712*L
-1297761051646769/18014398509481984) -1.*(8988458785899585/9007199254740992) ^exp(1014624516889591/9007199254740992*x+
7827322445356437/590295810358705651712*L-1297761051646769/18014398509481984)) *(.14677992676220695409205 2*L*(100.+x)
^(5/12)-.99999999999999999999999994e-4*L*(100.+x) ^2/(.99999999999999999999999994e-4*(100.+x)^2-1.)*
(.14677992676220695409205 2*(100.+x)^(5/12)-1.)) +.96015817993554868607988*(8961090140748077/9007199254740992)
^exp(1014624516889591/9007199254740992*x+7827322445356437/590295810358705651712*L-1297761051646769/18014398509481984)
*((2197633039383027/2251799813685248) ^exp(5595364410765473/144115188075855872*x-5176894256845849/295147905179352825856*L
+5425581046685411/4503599627370496) -1.*(4371919373261769/4503599627370496) ^exp(5595364410765473/144115188075855872*x
-5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496)) *(.21544346900318837217592 9e-1*L*(100.+x)
^(5/6)-.99999999999999999999999994e-4*L*(100.+x) ^2/(.99999999999999999999999994e-4*(100.+x)^2-1.)*
(.215443469003188372175929e-1*(100.+x)^(5/6)-1.))
 *(-1.+1.36258413811592257099598*(100.+x)^(1/12))+.228609428660058551185085*(8085879414211337/9007199254740992)
 ^exp(5595364410765473/144115188075855872*x-5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496)
 *((1114792330052669/1125899906842624) ^exp(1014624516889591/9007199254740992*x+7827322445356437/
590295810358705651712*L-1297761051646769/18014398509481984) -1.*(4457727132502003/4503599627370496)
^exp(1014624516889591/9007199254740992*x+7827322445356437/590295810358705651712*L-1297761051646769/18014398509481984))
 *(.316227766016837933199889e-3*L*(100.+x)^(7/4)-.999999999999999999999994e-4*L*(100.+x)
^2/(.999999999999999999999994e-4*(100.+x)^2-1.)*(.316227766016837933199889e-3*(100.+x)^(7/4)-1.))
 +.248985632533060312197452*(9005957161963763/9007199254740992) ^exp(5595364410765473/144115188075855872
*x-5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496) *(1-1.*(4503497800982921
/4503599627370496) ^exp(1014624516889591/9007199254740992*x+7827322445356437/590295810358705651712*L
-1297761051646769/18014398509481984)) *L+.936384219791599825654103e-4*(8085879414211337/9007199254740992)
^exp(5595364410765473/144115188075855872*x-5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496)
*(4457727132502003/4503599627370496) ^exp(1014624516889591/9007199254740992*x+7827322445356437/590295810358705651712*L
-1297761051646769/18014398509481984) *L*(.681292069057961285497988*(100.+x)^(1/12)-1.)*(100.+x)^2/
(.999999999999999999999994e-4*(100.+x)^2-1.) +.979212663861430165005809e-4*(8701426817608717/9007199254740992)
 ^exp(5595364410765473/144115188075855872*x-5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496)
 *(8953690005984367/9007199254740992) ^exp(1014624516889591/9007199254740992*x+7827322445356437/590295810358705651712
*L-1297761051646769/18014398509481984) *L*(.681292069057961285497988*(100.+x)^(1/12)-1.)*(100.+x)^2/
(.999999999999999999999994e-4*(100.+x)^2-1.) +.967341583333966287927264e-4*(8540854395630539/9007199254740992)
^exp(5595364410765473/144115188075855872*x-5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496)
*(8943931065879825/9007199254740992) ^exp(1014624516889591/9007199254740992*x+7827322445356437/590295810358705651712*L
-1297761051646769/18014398509481984) *L*(.681292069057961285497988*(100.+x)^(1/12)-1.)*(100.+x)^2/
(.999999999999999999999994e-4*(100.+x)^2-1.) +.932584869035150621681818e-4*(4017172091621691/4503599627370496)
^exp(5595364410765473/144115188075855872*x-5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496)
*(8907375347704459/9007199254740992) ^exp(1014624516889591/9007199254740992*x+7827322445356437/590295810358705651712*L
-1297761051646769/18014398509481984) *L*(.681292069057961285497988*(100.+x)^(1/12)-1.)*(100.+x)^2/
(.999999999999999999999994e-4*(100.+x)^2-1.) +.100748147914430100684785e-3*(8971003374175853/9007199254740992)
^exp(5595364410765473/144115188075855872*x-5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496)
*(4498200802209197/4503599627370496) ^exp(1014624516889591/9007199254740992*x+7827322445356437/590295810358705651712*L
-1297761051646769/18014398509481984) *L*(.681292069057961285497988*(100.+x)^(1/12)-1.)*(100.+x)^2
/(.999999999999999999999994e-4*(100.+x)^2-1.) +.951737040249115694470698e-4*(4149995179171859/4503599627370496)
 ^exp(5595364410765473/144115188075855872*x-5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496)
*(8932128122048405/9007199254740992) ^exp(1014624516889591/9007199254740992*x+7827322445356437/590295810358705651712*L
-1297761051646769/18014398509481984) *L*(.681292069057961285497988*(100.+x)^(1/12)-1.)*(100.+x)^2/
(.999999999999999999999994e-4*(100.+x)^2-1.) +.952382361312192024857459*(559399050888615/562949953421312)
^exp(1014624516889591/9007199254740992*x+7827322445356437/590295810358705651712*L-1297761051646769/18014398509481984)
*((8701426817608717/9007199254740992) ^exp(5595364410765473/144115188075855872*x-
5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496)
-1.*(8653328193444415/9007199254740992) ^exp(5595364410765473/144115188075855872*x-
5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496)) *(L*(1.+.100000000000000000000000e-1*x)
-.999999999999999999999994e-6*L*(100.+x)^2/(.999999999999999999999994e-4*(100.+x)
^2-1.)*x)*(-1.+1.36258413811592257099598*(100.+x)^(1/12)) +.963416624007758678719382e-
4*(8481104869158237/9007199254740992) ^exp(5595364410765473/144115188075855872*x-
5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496) *
(1117463989200615/1125899906842624) ^exp(1014624516889591/9007199254740992*x+7827322445356437/
590295810358705651712*L-1297761051646769/18014398509481984) *L*(.681292069057961285497988*(100.+x)^
(1/12)-1.)*(100.+x)^2/(.999999999999999999999994e-4*(100.+x)^2-1.) +.999322413886175708575654e-
4*(4457337796314217/4503599627370496) ^exp(5595364410765473/144115188075855872*x-5176894256845849
/295147905179352825856*L+5425581046685411/4503599627370496) *(8988458785899585/9007199254740992)
^exp(1014624516889591/9007199254740992*x+7827322445356437/590295810358705651712*L-1297761051646769/18014398509481984)
 *L*(.681292069057961285497988*(100.+x)^(1/12)-1.)*(100.+x)^2/(.999999999999999999999994e-4*(100.+x)^2-1.)
+.101570715988603931422478e-3*(4500844009866497/4503599627370496) ^exp(5595364410765473/144115188075855872*x-
5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496) *(2251126683166943/2251799813685248)
^exp(1014624516889591/9007199254740992*x+7827322445356437/590295810358705651712*L-1297761051646769/18014398509481984)
*L*(.681292069057961285497988*(100.+x)^(1/12)-1.)*(100.+x)^2/(.999999999999999999999994e-4*(100.+x)^2-1.
) +.967997484988011002179496*(2244442328010009/2251799813685248) ^exp(1014624516889591/9007199254740992*x+
7827322445356437/590295810358705651712*L-1297761051646769/18014398509481984) *((8878938178505361/9007199254740992)
 ^exp(5595364410765473/144115188075855872*x-5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496)
 -1.*(8837970463919045/9007199254740992) ^exp(5595364410765473/144115188075855872*x-
5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496))
```

```
*(.4641588833612778892410008e-1*L*(100.+x)^(2/3)-.9999999999999999999999994e-4*L*(100.+x)
^2/(.9999999999999999999999994e-4*(100.+x)^2-1.)*(.4641588833612778892410008e-1*(100.+x)^(2/3)-1.))
*(-1.+1.3625841381159225709598*(100.+x)^(1/12))+.1011585958690334889516 57e-3*(8991702818783173/9007199254740992)
^exp(5595364410765473/144115188075855872*x-5176894256845849/2951479051793528258 56*L+5425581046685411/4503599627370496)
*(2250134495133037/2251799813685248) ^exp(1014624516889591/9007199254740992*x+7827322445356437/5902958103587056517 12*L
-1297761051646769/18014398509481984) *L*(.68129206905796128549 7988*(100.+x)^(1/12)-1.)*(100.+x)^2/
(.9999999999999999999999994e-4*(100.+x)^2-1.) +.92880093402991275677 2550e-4*(3996675083817621/4503599627370496)
^exp(5595364410765473/144115188075855872*x-5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496)
*(2223127196297631/2251799813685248) ^exp(1014624516889591/9007199254740992*x+7827322445356437/590295810358705651712*L
-1297761051646769/18014398509481984) *L*(.681292069057961285497988*(100.+x)^(1/12)-1.)*(100.+x)^2/
(.9999999999999999999999994e-4*(100.+x)^2-1.) +.9332172040537994064028 37*(2233726327948655/2251799813685248)
^exp(1014624516889591/9007199254740992*x+7827322445356437/590295810358705651712*L-1297761051646769/18014398509481984)
*((4212378365670251/4503599627370496) ^exp(5595364410765473/144115188075855872*x-
5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496)
-1.*(8364791122158079/9007199254740992) ^exp(5595364410765473/144115188075855872*x-
5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496))
*(.1467799267622069540920 52e-2*L*(100.+x)^(17/12)-.9999999999999999999999994e-4*L*(100.+x)
^2/(.9999999999999999999999994e-4*(100.+x)^2-1.)*(.146779926762206954092052e-2*(100.+x)^(17/12)-1.))
*(-1.+1.3625841381159225709598*(100.+x)^(1/12))+.9759007948107184654059 18*(8988458785899585/9007199254740992)
^exp(1014624516889591/9007199254740992*x+7827322445356437/590295810358705651712*L-1297761051646769/18014398509481984)
*((8945078853280857/9007199254740992) ^exp(5595364410765473/144115188075855872*x-
5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496)
-1.*(4457337796314217/4503599627370496) ^exp(5595364410765473/144115188075855872*x-
5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496))
*(.1000000000000000000000000*L*(100.+x)^(1/2)-.9999999999999999999999994e-4*L*(100.+x)
^2/(.9999999999999999999999994e-4*(100.+x)^2-1.)*(.1000000000000000000000000*(100.+x)^(1/2)-1.))
*(-1.+1.3625841381159225709598*(100.+x)^(1/12))+.9181631338896785192234 00*(1114792330052669/1125899906842624)
^exp(1014624516889591/9007199254740992*x+7827322445356437/590295810358705651712*L-1297761051646769/18014398509481984)
*((8190338516056353/9007199254740992) ^exp(5595364410765473/144115188075855872*x-5176894256845849/
295147905179352825856*L+5425581046685411/4503599627370496) -1.*(2033127443706281/2251799813685248)
^exp(5595364410765473/144115188075855872*x-5176894256845849/295147905179352825856*L+5425581046685411/
4503599627370496)) *(.3162277660168379331998 89e-3*L*(100.+x)^(7/4)-.9999999999999999999999994e-4*L*(100.+x)
^2/(.9999999999999999999999994e-4*(100.+x)^2-1.)*(.316227766016837933199889e-3*(100.+x)^(7/4)-1.))
*(-1.+1.3625841381159225709598*(100.+x)^(1/12))+.2429852766854630398387 06*(8878938178505361/9007199254740992)
^exp(5595364410765473/144115188075855872*x-5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496)
*((8988458785899585/9007199254740992) ^exp(1014624516889591/9007199254740992*x+7827322445356437/
590295810358705651712*L-1297761051646769/18014398509481984) -1.*(8983141115603571/9007199254740992)
^exp(1014624516889591/9007199254740992*x+7827322445356437/590295810358705651712*L-1297761051646769/18014398509481984))
*(.1000000000000000000000000*L*(100.+x)^(1/2)-.9999999999999999999999994e-4*L*(100.+x)
^2/(.9999999999999999999999994e-4*(100.+x)^2-1.)*(.1000000000000000000000000*(100.+x)^(1/2)-1.))
+.9556144169510905921564 99e-4*(8364791122158079/9007199254740992) ^exp(5595364410765473/144115188075855872*x-
5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496) *(2233726327948655/2251799813685248)
^exp(1014624516889591/9007199254740992*x+7827322445356437/590295810358705651712*L-1297761051646769/18014398509481984)
*L*(.681292069057961285497988*(100.+x)^(1/12)-1.)*(100.+x)^2/(.9999999999999999999999994e-4*(100.+x)^2-1.)
+.995267693303656611179333e-4*(8878938178505361/9007199254740992) ^exp(5595364410765473/144115188075855872*x-
5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496) *(8983141115603571/9007199254740992)
^exp(1014624516889591/9007199254740992*x+7827322445356437/590295810358705651712*L-1297761051646769/18014398509481984)
*L*(.681292069057961285497988*(100.+x)^(1/12)-1.)*(100.+x)^2/(.9999999999999999999999994e-4*(100.+x)^2-1.)
+.940199049103030803684756e-4*(2033127443706281/2251799813685248) ^exp(5595364410765473/144115188075855872*x-
5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496) *(1114792330052669/1125899906842624)
^exp(1014624516889591/9007199254740992*x+7827322445356437/590295810358705651712*L-1297761051646769/18014398509481984)
*L*(.681292069057961285497988*(100.+x)^(1/12)-1.)*(100.+x)^2/(.9999999999999999999999994e-4*(100.+x)^2-1.)
+.971282532944468866936489e-4*(4300974033223813/4503599627370496) ^exp(5595364410765473/144115188075855872*x-
5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496) *(559183992747409/562949953421312)
^exp(1014624516889591/9007199254740992*x+7827322445356437/590295810358705651712*L-1297761051646769/18014398509481984)
*L*(.681292069057961285497988*(100.+x)^(1/12)-1.)*(100.+x)^2/(.9999999999999999999999994e-4*(100.+x)^2-1.)
+.947875395886275009735913e-4*(8239055304305507/9007199254740992) ^exp(5595364410765473/144115188075855872*x-
5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496) *(2231562280773709/2251799813685248)
^exp(1014624516889591/9007199254740992*x+7827322445356437/590295810358705651712*L-1297761051646769/18014398509481984)
*L*(.681292069057961285497988*(100.+x)^(1/12)-1.)*(100.+x)^2/(.9999999999999999999999994e-4*(100.+x)^2-1.)
+.101984515085541503876075e-3*(9005957161963763/9007199254740992) ^exp(5595364410765473/144115188075855872*x-
5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496) *(4503497800982921/4503599627370496)
^exp(1014624516889591/9007199254740992*x+7827322445356437/590295810358705651712*L-1297761051646769/18014398509481984)
*L*(.681292069057961285497988*(100.+x)^(1/12)-1.)*(100.+x)^2/(.9999999999999999999999994e-4*(100.+x)^2-1.)
+.975239537983684633454032e-4*(8653328193444415/9007199254740992) ^exp(5595364410765473/144115188075855872*x-
5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496) *(559399050888615/562949953421312)
^exp(1014624516889591/9007199254740992*x+7827322445356437/590295810358705651712*L-1297761051646769/18014398509481984)
*L*(.681292069057961285497988*(100.+x)^(1/12)-1.)*(100.+x)^2/(.9999999999999999999999994e-4*(100.+x)^2-1.)
+.987207541105260434860733e-4*(2197633039383027/2251799813685248) ^exp(5595364410765473/144115188075855872*x-
5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496) *(4484484909400055/4503599627370496)
^exp(1014624516889591/9007199254740992*x+7827322445356437/590295810358705651712*L-1297761051646769/18014398509481984)
*L*(.681292069057961285497988*(100.+x)^(1/12)-1.)*(100.+x)^2/(.9999999999999999999999994e-4*(100.+x)^2-1.)
+.959507590085749335576884e-4*(4212378365670251/4503599627370496) ^exp(5595364410765473/144115188075855872*x-
5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496) *(8937376977342113/9007199254740992)
^exp(1014624516889591/9007199254740992*x+7827322445356437/590295810358705651712*L-1297761051646769/18014398509481984)
*L*(.681292069057961285497988*(100.+x)^(1/12)-1.)*(100.+x)^2/(.9999999999999999999999994e-4*(100.+x)^2-1.)
+.983201976254001631854574e-4*(4371919373261769/4503599627370496) ^exp(5595364410765473/144115188075855872*x-
5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496) *(8961090140748077/9007199254740992)
^exp(1014624516889591/9007199254740992*x+7827322445356437/590295810358705651712*L-1297761051646769/18014398509481984)
*L*(.681292069057961285497988*(100.+x)^(1/12)-1.)*(100.+x)^2/(.9999999999999999999999994e-4*(100.+x)^2-1.)
+.100339365340034798877624e-3*(8945078853280857/9007199254740992) ^exp(5595364410765473/144115188075855872*x-
5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496) *(8992420872707761/9007199254740992)
^exp(1014624516889591/9007199254740992*x+7827322445356437/590295810358705651712*L-1297761051646769/18014398509481984)
*L*(.681292069057961285497988*(100.+x)^(1/12)-1.)*(100.+x)^2/(.9999999999999999999999994e-4*(100.+x)^2-1.)
+.944029420029076551521293e-4*(8190338516056353/9007199254740992) ^exp(5595364410765473/144115188075855872*x-
5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496) *(1115049012713431/1125899906842624)
```

```
^exp(1014624516889591/9007199254740992*x+7827322445356437/590295810358705651712*L-1297761051646769/18014398509481984)
*L*(.681292069057961285497988*(100.+x)^(1/12)-1.)*(100.+x)^2/(.9999999999999999999999994e-4*(100.+x)^2-1.)
+.971941106741852159354823*(8983141115603571/9007199254740992) ^exp(1014624516889591/9007199254740992*x+
7827322445356437/590295810358705651712*L-1297761051646769/18014398509481984) *((4457337796314217/4503599627370496)
^exp(5595364410765473/144115188075855872*x-5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496)
-1.*(8878938178505361/9007199254740992) ^exp(5595364410765473/144115188075855872*x-5176894256845849/
295147905179352825856*L+5425581046685411/4503599627370496)) *(.681292069057961285497988e-1*L*(100.+x)
^(7/12)-.9999999999999999999999994e-4*L*(100.+x) ^2/(.9999999999999999999999994e-4*(100.+x)^2-1.)
*(.681292069057961285497988e-1*(100.+x)^(7/12)-1.)) *(-1.+1.36258413811592257099598*(100.+x)^(1/12))
+.991229424627723266231798e-4*(8837970463919045/9007199254740992) ^exp(5595364410765473/144115188075855872*x-
5176894256845849/295147905179352825856*L+5425581046685411/4503599627370496) *(2244442328010009/2251799813685248)
^exp(1014624516889591/9007199254740992*x+7827322445356437/590295810358705651712*L-1297761051646769/18014398509481984)
*L*(.681292069057961285497988*(100.+x)^(1/12)-1.)*(100.+x)^2/(.9999999999999999999999994e-4*(100.+x)^2-1.))
*exp(2.71908869999999991406980-.177276789999999989655421*x-.817499999999999948712900e-4*L) /
(1.+exp(2.71908869999999991406980-.177276789999999989655421*x-.817499999999999948712900e-4*L))
```