UNIVERSITY OF SOUTHAMPTON

FACULTY OF BUSINESS AND LAW

Southampton Management School

SELECTED MODELLING PROBLEMS IN CREDIT SCORING

by

Katarzyna Helena Bijak

Thesis for the degree of Doctor of Philosophy

August 2013

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF BUSINESS AND LAW

Southampton Management School

Thesis for the degree of Doctor of Philosophy

SELECTED MODELLING PROBLEMS IN CREDIT SCORING

by Katarzyna Helena Bijak

This research addresses three selected modelling problems that occur in credit scoring. The focus is on segmentation, modelling Loss Given Default (LGD) for unsecured loans and affordability assessment.

It is usually expected that segmentation, i.e. dividing the population into a number of groups and building separate scorecards for them, will improve the model performance. The most common statistical methods for segmentation are the two-step approaches, where logistic regression follows Classification and Regression Trees (CART) or Chi-square Automatic Interaction Detection (CHAID) trees. In this research, these approaches and a simultaneous method, in which both segmentation and scorecards are optimised at the same time: Logistic Trees with Unbiased Selection (LOTUS), are applied to the data provided by two UK banks and a European credit bureau. The model performance measures are compared to assess an improvement due to the segmentation.

For unsecured retail loans, LGD is often found difficult to model. In the frequentist (classical) two-step approach, the first model (logistic regression) is used to separate positive values from zeroes and the second model (e.g. linear regression) is applied to estimate these values. Instead, one can build a Bayesian hierarchical model, which is a more coherent approach. In this research, Bayesian methods and the frequentist approach are applied to the data on personal loans provided by a UK bank. The Bayesian model generates an individual predictive distribution of LGD for each loan, whose potential applications include approximating the downturn LGD and stress testing LGD under Basel II.

An applicant's affordability (ability to repay) is often checked using a simple, static approach. In this research, a theoretical framework for dynamic affordability assessment is proposed. Both income and consumption are allowed to vary over time and their changes are described with random effects models for panel data. On their basis a simulation is run for a given applicant. The ability to repay is checked over the life of the loan and for all possible instalment amounts. As a result, a probability of default is assigned to each amount, which can help find the maximum affordable instalment. This is illustrated with an example based on artificial data.

Contents

List of tables

List of figures

DECLARATION OF AUTHORSHIP

I, Katarzyna Helena Bijak

declare that the thesis entitled

SELECTED MODELLING PROBLEMS IN CREDIT SCORING

and the work presented in the thesis are both my own, and have been generated by me as the result of my own original research. I confirm that:

- this work was done wholly or mainly while in candidature for a research degree at this University;
- where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- where I have consulted the published work of others, this is always clearly attributed;
- where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
- parts of this work have been published as:
    - Bijak, K. and Thomas, L.C. (2012) Does segmentation always improve model performance in credit scoring?, *Expert Systems with Applications*, 39(3), pp. 2433-2442.
    - Bijak, K. and Thomas, L.C. (2013) Modelling LGD for unsecured retail loans using Bayesian methods, accepted for publication in the *Journal of the Operational Research Society*.

Signed: …….

Date: …….

Acknowledgements

I would like to take this opportunity to thank those who have helped and supported me during the course of my research and throughout the writing of this thesis.

First of all, I am truly indebted and thankful to my supervisor Professor Lyn Thomas whose support, guidance and advice have been invaluable to me. I have been extremely lucky to have a supervisor who is not only one of the most respected experts in credit scoring but also a great, friendly and warm person. It has been both a privilege and a pleasure to work with him. He has been a constant source of inspiration to me and I have really enjoyed our meetings. I would also like to express my gratitude to my second supervisor Dr Christophe Mues for his helpful comments and suggestions and, in particular, for his involvement in the project on the dynamic affordability assessment. I would like to thank my examiners Professor David Hand and Professor Bart Baesens for their feedback and the very interesting and stimulating discussion during my viva.

On the basis of this research, a paper has been published in *Expert Systems with Applications* and another one has been accepted for publication in the *Journal of the Operational Research Society*. I am grateful to anonymous journal referees for their comments and suggestions that have also helped improve this thesis. Moreover, I am obliged to the credit bureau which has provided me with a large and unique dataset generated exclusively for this research.

I would like to thank all members of the Centre for Operational Research, Management Sciences and Information Systems (CORMSIS) at the University of Southampton for creating a friendly and inspiring environment, in which it has been a pleasure to work, both as a PhD student and – for a period of time – as a CORMSIS research facilitator.

I would also like to thank my managers and friends from the Polish credit bureau (BIK S.A.), Piotr Woźniak and Mariola Kapla. The experience and skills that I have gained over the years of working with them are invaluable and have helped me confidently conduct my research.

Finally, I would like to express my gratitude to my family. I am indebted to my parents and parents in law for their constant encouragement. I am also thankful to my son Jerzy, who was born during the course of my research and has taught me how to manage my time more effectively. My ultimate thanks go to my husband Jakub. Without his love, patience and incredible support (including childminding our son in the last few weeks) this thesis would not have been possible.

Abbreviations

AID	Automatic Interaction Detection
AIRB	Advanced Internal Ratings-Based
AOC	Area Over the Regression Error Characteristic Curve
APR	Annual Percentage Rate
ART	Adaptive Random Trees
AUC	Area Under the ROC Curve
AUROC	Area Under the ROC Curve
BC	Bound and Collapse
BCBS	Basel Committee on Banking Supervision
CAIS	Credit Account Information Sharing
CART	Classification and Regression Trees
CATO	Current Account Turnover
CCF	Credit Conversion Factor
CCJ	County Court Judgement
CDF	Cumulative Distribution Function
CF	Conversion Factor
CHAID	Chi-square Automatic Interaction Detection
CLV	Customer Lifetime Value
CRA	Credit Reference Agency
CRD	Capital Requirements Directive
CRM	Customer Relationship Management
EAD	Exposure at Default
EBA	European Banking Authority
EFS	Expenditure and Food Survey
EM	Expectation Maximisation
FCA	Financial Conduct Authority

FD	First Difference
FE	Fixed Effects
FiNPiN	Financial Pinpoint Identified Neighbourhood
FIRB	Foundation Internal Ratings-Based
FSA	Financial Services Authority
GA	Genetic Algorithm
GDP	Gross Domestic Product
GLS	Generalised Least Squares
IRB	Internal Ratings-Based
KS	Kolmogorov-Smirnov
LAV	Least Absolute Value
LDP	Low Default Portfolio
LEQ	Loan Equivalent
LGD	Loss Given Default
LMT	Logistic Model Trees
LOTUS	Logistic Trees with Unbiased Selection
LSSVM	Least Squares Support Vector Machine
LTV	Loan to Value
MAE	Mean Absolute Error
MARS	Multivariate Adaptive Regression Splines
MC	Monte Carlo
MCMC	Markov Chain Monte Carlo
ML	Maximum Likelihood
MNAR	Missing Not at Random
MSE	Mean Square Error
NN	Neural Network
OFT	Office of Fair Trading
OLS	Ordinary Least Squares
ONS	Office for National Statistics

PD	Probability of Default
PIH	Permanent Income Hypothesis
PIT	Point-in-Time
PRA	Prudential Regulation Authority
RBP	Risk-Based Pricing
RE	Random Effects
RMSE	Root Mean Square Error
ROC	Receiver Operating Characteristic
RR	Recovery Rate
SOM	Self-Organising Map
SVM	Support Vector Machine
TTC	Through-the-Cycle
TTD	Through-the-Door
VaR	Value at Risk
VBA	Visual Basic for Applications
WoE	Weight of Evidence

Chapter 1
Introduction

1.1 Research aim

The last global financial crisis has focused attention on risk management tools, including credit scoring models used by lenders and credit bureaus. Although credit scoring models have generally proven to work better than rating models used by rating agencies, there are still numerous unsolved problems in this area. The aim of this research is to address three selected modelling problems in credit scoring which are actually faced in banking practice. The specific objectives are set in section 1.3 of this Introduction after defining the key concepts in section 1.2. The thesis structure is outlined in section 1.4.

1.2 Credit scoring

Credit scoring is "the set of decision models and their underlying techniques that aid lenders in the granting of consumer credit" (Thomas *et al.*, 2002, p. 1). It can be classified as business analytics, a subset of business intelligence (Davenport and Harris, 2007). Since the 1950s, credit scoring has been used to identify applicants who are likely to repay and thus can be granted a loan (Mays, 2004). Nowadays most lenders use scoring to assess the credit risk of their actual or potential customers (individuals as well as small and medium enterprises). Credit scoring models are also developed by credit bureaus (also known as credit reference agencies, CRAs) to help banks estimate that risk on the basis of data coming from the banking sector as a whole.

For about ten years, the need for quantitative models in banking has been largely related to the New Basel Capital Accord (Basel II) and in particular to the Internal Ratings-Based (IRB) approach. Basel II has replaced the Basel Capital Accord (Basel I) and its recommendations have been implemented through legislation in many countries. In the EU member states, they have been transposed into law via the Capital Requirements

Directive (CRD), under which name the Council Directives 2006/48/EC and 2006/49/EC are collectively known.

The Basel II document contains the revised capital adequacy framework agreed by the Basel Committee on Banking Supervision (BCBS, 2006). Basel II consists of three pillars: minimum capital requirements, supervisory review process and market discipline. The first pillar focuses on credit, operational and market risk. As far as credit risk is concerned, there are two possible approaches: the Standardised Approach and the IRB approach. Contrary to the former, the latter allows lenders to assess credit risk parameters on their own. It can be implemented as either Foundation (FIRB) or Advanced (AIRB). Under the FIRB approach, which is not available for retail exposures, banks use internal estimates of Probability of Default (PD), whereas other risk parameters are provided by the regulator. Under the AIRB approach for retail exposures, lenders are allowed to use their own estimates of PD, Loss Given Default (LGD) and Exposure at Default (EAD). The product of PD, LGD and EAD is the expected loss amount that should be covered by profits from lending. Regulatory capital is computed as the difference between the Value at Risk (VaR) at a 99.9% confidence level (such amount that the probability of having a higher loss is 0.1%) and the expected loss. Estimates of PD and LGD are employed to calculate capital requirements, i.e. capital needed to cover the unexpected loss for one unit of money at risk. Capital requirements are applied along with EAD estimates to determine risk-weighted assets. More on PD, LGD, EAD and Basel II can be found e.g. in the books by Thomas (2009a) and van Gestel and Baesens (2009).

Basel II has recently been reformed by Basel III that has set even higher requirements, e.g. a higher Tier 1 capital ratio, i.e. the ratio of the bank's core capital to risk-weighted assets (BCBS, 2011). The Third Basel Accord has also introduced a number of novelties, such as the capital conservation buffer and the countercyclical buffer. Nevertheless, Basel III still requires the same risk parameters to be estimated.

The banks which have received approval from the regulator to use the IRB approach are expected to stress test the estimated risk parameters. According to the European Banking Authority (EBA), "a deep (probabilistic) understanding of how macro-

economic variables and institution specific effects would impact the institution at any given point in time is important in stress testing modelling. Ideally, this transformation should be based on quantitative modelling where data is relatively rich" (EBA, 2010, point 52).

Currently there is much more to credit scoring than just determining which applicants are likely to repay their loans. This section provides a review of selected credit scoring areas. To learn about other related issues, it is recommended to refer to the literature, e.g. the books by Thomas *et al.* (2002) and Anderson (2007).

1.2.1 PD modelling

The original objective of credit scoring was to classify applicants into Goods (creditworthy) and Bads (uncreditworthy, defaulters). With the introduction of Basel II, the focus has started to shift from the classification to the accurate assessment of credit risk. The Council Directive 2006/48/EC has adopted the Basel definition of default, according to which the obligor defaults when at least one of the following events occurs: "(a) the credit institution considers that the obligor is unlikely to pay its credit obligations to the credit institution, the parent undertaking or any of its subsidiaries in full, without recourse by the credit institution to actions such as realising security (if held); (b) the obligor is past due more than 90 days on any material credit obligation to the credit institution, the parent undertaking or any of its subsidiaries" (Annex VII, Part 4, point 44). In the UK, the regulator has changed 90 to 180 days overdue. PD is defined as the probability of default over a one year period (Council Directive 2006/48/EC, Article 4(25)). In case of retail exposures, the above definition of default may be applied either at the customer or at the account (facility) level and thus, PD models can be developed at both levels.

There are different types of scoring. Application scoring is used to accept and reject applicants, especially new customers. Behavioural scoring is employed to assess the credit risk of existing customers and can also be implemented in the credit decision making process. Application scoring is mainly based on data from loan applications, whereas behavioural scoring is mostly based on data on customers' behaviour stored in bank databases. Moreover, there is credit bureau scoring. Credit bureaus are institutions

that collect and analyse data on loans granted by banks operating in a given country (Anderson, 2007; van Gestel and Baesens, 2009). Such data enable tracking the credit history of a customer in the banking sector. Credit bureau scoring is based on data on customers' credit histories. Both application and behavioural scoring can be enriched with credit bureau data or scores. As a rule, this increases the model performance (van Gestel and Baesens, 2009).

In application scoring, one of the key challenges is sample bias, since models are built using data on the accepted applicants and then used to assess the through-the-door (TTD) population, including those who would have previously been rejected. The set of techniques proposed to deal with this problem is referred to as 'reject inference'. They include, among other methods, extrapolation from the estimated model to the rejected applicants and augmentation (e.g. weighting the sample to take into account the probability of acceptance). A number of reject inference techniques were presented by Hand and Henley (1993). Some methods which are used in the industry were also described by Siddiqi (2005). Nevertheless, Hand and Henley (1993) concluded that it is impossible to construct a reliable reject inference technique, at least without the use of additional information. Additional information on the rejected applicants sometimes can be obtained from a credit bureau.

Many statistical and data mining methods have been proposed to develop credit scoring models. Among other things, they include: discriminant analysis, logistic and other forms of regression, classification trees, *k*-nearest neighbour algorithm, linear programming, neural networks (NNs), genetic algorithms (GAs) and support vector machines (SVMs) (e.g. Baesens *et al.*, 2003; Thomas *et al.*, 2002). A single scoring model (scorecard) can be developed for the entire customer population. However, the population is often segmented into a number of groups, for which separate scorecards are built, since this is expected to improve the model performance (more on segmentation can be found in Chapter 2). As far as the model performance is concerned, discriminatory power (ability to separate Goods and Bads), calibration (accuracy of PD estimates) and stability are usually validated. Before Basel II, classification was the major task and thus, separation ability was considered most important. There is a wide selection of discriminatory power measures, including the Gini coefficient and the

Kolmogorov-Smirnov (KS) statistic (Thomas, 2009a) as well as the H measure (Hand, 2009). Nowadays many PD models perform very well but there is still room for improvement in areas such as low default portfolios (LDPs). Pluto and Tasche (2006) proposed a methodology to obtain the most prudent PD estimates based on the assumption that the ordinal borrower ranking is correct. Bayesian methods also seem suitable for LDPs (e.g. Kiefer, 2009).

Such models as logistic regression enable predicting whether the customer will default within a given time horizon (usually 12 months). Nevertheless, sometimes it may be worth forecasting not only if but also when he or she will default. For this purpose, survival analysis is used, e.g. the Cox proportional hazards model (Banasik *et al.*, 1999). As a result, one can obtain the distribution of time until default. Survival analysis has the advantage of allowing for time-variant regressors in general and for macroeconomic variables in particular. Introducing macroeconomic variables enables producing through-the-cycle (TTC) estimates as opposed to point-in-time (PIT) estimates provided by traditional credit scoring models (Thomas, 2010). It also enables taking into account correlations in defaults, which makes survival analysis especially useful in portfolio PD modelling.

1.2.2 Portfolio PD modelling

For a long time, lenders have been using well-performing models to assess PD at the individual level. However, they often lack similarly effective tools to model PD at the portfolio level. Such tools are needed, since banks are expected to stress test their consumer loan portfolios under Basel II. Portfolio PD models can also be applied to assess the risk of asset-backed securities in the securitization process. Traditional credit scoring models cannot be directly used to estimate portfolio PD, since they assume independence of the default events. In fact, this assumption is never satisfied. It is estimated that coefficients of correlation in defaults typically vary from 0.5% to 3% (BCBS, 2005b). In particular, numerous default events occur simultaneously in the economic downturn. Nevertheless, the traditional models can still be applied, if default correlations are modelled separately, e.g. using copulas (Li, 2000).

Many tools which are proposed for assessing PD of consumer loan portfolios are motivated by models of corporate credit risk, both structural and reduced-form. In a structural model, the firm's asset value follows a stochastic process, and if it falls below the threshold, the firm is assumed to default. Reduced-form models allow for estimating PD based on the firm's characteristics, such as its bond price or rating, industry and region, as well as macroeconomic variables. In structural models for retail portfolios, the firm's asset value can be replaced with the consumer's score provided by a credit bureau (de Andrade and Thomas, 2007). Similarly, the consumer's behavioural score can be used instead of the firm's bond rating in a reduced-form model (Malik and Thomas, 2010). The latter work is an example of employing survival analysis in portfolio PD modelling. Another possible approach is using frailty models that are an extension of the Cox proportional hazards model (Thomas, 2009a). Frailty models assume that there are unobserved factors which affect different firms or consumers and thus may connect their defaults. Thomas (2009b) described a few more approaches to modelling PD of retail portfolios by using analogies to corporate credit risk models (e.g. based on Markov chains). Taking into account the state of the economy, it is possible to assume conditional independence of the default events given the macroeconomic environment (e.g. Rosch and Scheule, 2003). One can also use macroeconomic variables to model the exogenous function, i.e. one of the products of the default rate decomposition in the dual-time dynamics approach (Breeden *et al.*, 2008).

1.2.3 LGD modelling

LGD is the loss borne by the bank when a customer defaults on a loan. The Council Directive 2006/48/EC defines LGD as "the ratio of the loss on an exposure due to the default of a counterparty to the amount outstanding at default" (Article 4(27)), where 'loss' means "economic loss, including material discount effects, and material direct and indirect costs associated with collecting on the instrument" (Article 4(26)). According to the EBA guidelines, "the data used to calculate the realised LGD of an exposure should include all relevant information" (EBA, n.d., section 3.3.2.2). Among the relevant information, the guidelines mention: outstanding amount of the exposure at default (including principal as well as interests and fees), recoveries (e.g. proceeds from the sale of collateral or the loan) and work-out costs (including the costs of both in-house and outsourced collection).

LGD modelling has a much shorter history than PD modelling and is mostly associated with Basel II. LGD for corporate loans has been assessed for a longer time than for retail loans, first with a fixed value based on historical data, and then using more complicated models (Thomas, 2009a). Various approaches to modelling corporate LGD were presented e.g. by Altman *et al.* (2005). Since the sale of collateral can have a large impact on LGD, there are separate models for secured and unsecured loans. In particular, mortgage LGD can be modelled either directly or as a combination of repossession and haircut models, where a 'haircut' is the ratio of the sale price to the estimated value of a property. In a two-stage approach, the first model (e.g. logistic regression) separates repossessed properties from the rest, whereas the second model (e.g. linear regression) allows for the haircut estimation; it is often assumed that LGD is equal to zero in case of the properties that are not repossessed. Among other techniques, survival analysis and quantile regression were also suggested. The most important regressor seems to be Loan to Value (LTV), i.e. the ratio of the outstanding debt to the value of a property. Examples of mortgage LGD models include those built by Somers and Whittaker (2007), Qi and Yang (2009), Leow *et al.* (2009 and 2010), Zhang *et al.* (2010) and Tong *et al.* (2011). LGD models for unsecured retail loans can be classified as either one-stage or multi-stage approaches, where various regression models and data mining techniques were proposed (details can be found in Chapter 3).

Under the AIRB approach, "credit institutions shall use LGD estimates that are appropriate for an economic downturn if those are more conservative than the long-run average" (Council Directive 2006/48/EC, Annex VII, Part 4, point 74). This is referred to as the 'downturn LGD'. The estimation of the downturn LGD can be challenging, since there is no Basel formula for it but only a principles-based approach was suggested (BCBS, 2005a).

1.2.4 EAD modelling

EAD is the exposure of a facility at the time of default. EAD is straightforward to estimate for on-balance sheet positions, since it can be determined on the basis of the current outstanding amount (and thus, it can be easily obtained e.g. for instalment loans or mortgages). It is more difficult to assess for off-balance sheet positions, e.g. in case of credit cards. For such products as credit cards, EAD can be calculated as a sum of the

current exposure and a product of the currently undrawn part of the allocated limit and a (credit) conversion factor, (C)CF. The Council Directive 2006/48/EC defines a conversion factor as "the ratio of the currently undrawn amount of a commitment that will be drawn and outstanding at default to the currently undrawn amount of the commitment, the extent of the commitment shall be determined by the advised limit, unless the unadvised limit is higher" (Article 4(28)). In some, especially American, literature (e.g. Moral, 2006; Jiménez *et al.*, 2009; Qi, 2009), CCF is called 'loan equivalent' (LEQ).

EAD modelling attracts less attention than LGD or PD and is strongly related to Basel II. Although EAD can be modelled directly (Taplin *et al.*, 2007), it is CCF/LEQ that is usually modelled (e.g. Valvonis, 2008; Qi, 2009; Brown, 2011). Those models include, among other things, linear regression and logit. Thomas (2009a) also suggested modelling CCF using probit, hazards models and Markov chains. The best regressors seem to be based on credit limit usage (e.g. change in utilisation rate within the last 12 months).

As with LGD, the banks which are permitted to use the AIRB approach need to estimate the downturn conversion factors, since "credit institutions shall use conversion factor estimates that are appropriate for an economic downturn if those are more conservative than the long-run average" (Council Directive 2006/48/EC, Annex VII, Part 4, point 88).

1.2.5 Affordability modelling

Credit scoring focuses mainly on creditworthiness, i.e. an applicant's propensity to repay a loan, derived from the fact that similar applicants repaid their loans in the past. However, even high creditworthiness does not necessarily mean ability to repay. Therefore, affordability should also be assessed. A loan can be considered affordable "if its level and terms allow the consumer to meet current and future payment obligations in full, without recourse to further debt relief or rescheduling, avoiding accumulation of arrears while allowing an acceptable level of consumption" (Financial Services Authority, 2010, paragraph 2.16). Thus, affordability assessment can be defined as "a 'borrower-focussed test' which involves a creditor assessing a borrower's ability to

undertake a specific credit commitment, or specific additional credit commitment, in a sustainable manner, without the borrower incurring (further) financial difficulties and/or experiencing adverse consequences" (Office of Fair Trading, 2011, paragraph 4.1). This is inextricably linked to the concepts of consumer overindebtedness and responsible lending. Irresponsible lending practices, such as granting credit without reasonable affordability assessment, may lead to the customer being overindebted. Therefore, regulations put affordability assessment at the centre of responsible lending. Nevertheless, there is little literature on statistical models and methods dedicated to this purpose. The solutions which are used in banking practice are described in Chapter 4.

1.2.6 Risk-based pricing

Risk-based pricing (RBP) means adjusting loan terms to the credit risk that is specific to the customer. Lenders most often adjust interest rates but other loan features can also be varied. Among such features are loan amounts, credit limits, initial discounts and some extra offers, e.g. insurance policies or loyalty programmes (Thomas, 2009a). They are determined at the time of application but some adjustments can be made over the life of the loan (to reflect changes in the customer's behavioural score, in particular).

In RBP, one of the major problems is adverse selection. Scorecards are usually scaled so that the higher the score, the lower the credit risk and the better the customer. Thus, applicants with lower scores are offered worse loan terms than applicants with higher scores. Those who nevertheless accept such offers generate even higher risk than the lender has predicted. However, if the lender increases interest rates for applicants with lower scores, it will attract even worse customers. As a result, worse customers drive out better customers due to asymmetric information between them and the lender. Huang and Thomas (2009) analysed the impact of adverse selection on the profitability of a lender that uses RBP for credit cards.

Thomas (2009a) proposed models for RBP that allow determining the optimal interest rate for any level of credit risk under various assumptions. Those assumptions include adverse selection (or the lack of it) as well as different shapes of the response rate function (i.e. function that gives probability of the applicant accepting the offer of a loan with a given interest rate). Konstantinos *et al.* (2003) applied Bayesian methods in RBP

for credit cards. In that approach, application scoring is combined with the data on the use of the credit card in the initial period in order to update the Annual Percentage Rate (APR). Since its ultimate goal is maximising profit, RBP is closely related to modelling profitability (Edelman, 2003).

1.2.7 Profit scoring

Predicting profit on a customer is a challenge, since there are so many factors that may have an impact on the final profit. Among other things, profit may be affected by: initial terms and conditions, changes in interest rates, limit increases, usage, such events as default, attrition or prepayment, choice and effectiveness of marketing or collection strategies, and possible sale of the portfolio. It may even be difficult to calculate the actual profit on a given customer. Nevertheless, there are attempts to develop models that could support profitability analysis, especially by predicting one or more of the above-mentioned factors. For example, Whittaker *et al.* (2005) proposed quantile regression to predict credit card balance. Some approaches, such as survival analysis used by Thomas *et al.* (2005), focus on propensity to purchase financial products, which rather resembles propensity scoring. Other approaches include Markov chains (Thomas *et al.*, 2001) and segmentation based on a number of scores that measure risk, attrition etc. (Thomas *et al.*, 2002). Some of the above-mentioned models can only be applied to make short-term predictions. Ideally, though, credit decisions should be based on Customer Lifetime Value (CLV), i.e. the predicted net profit from the whole relationship with a given customer. In consumer finance, CLV was modelled with quantile regression (Benoit and van den Poel, 2009) and other methods including linear regression, probit and tobit II (Donkers *et al.*, 2007). Most of those models were built using insurance data and thus did not take into account probability of default. Thomas (2010) suggested that competing risks analysis could be applied to allow for various possible events that may affect the lifetime value of a bank customer. Crowder *et al.* (2005) noted that the lender's actions change the expected CLV and proposed the model which can be used to choose optimal actions.

1.2.8 Propensity, attrition, collection and fraud scoring

Many credit scoring techniques can be adapted for other bank activities such as marketing, collection and fraud detection. Some of them, including propensity and

attrition/churn scoring tools, can be implemented in Customer Relationship Management (CRM) systems. Propensity scoring is used to select customers for marketing campaigns (especially direct-mail ones). Propensity scoring models allow for the prediction of which customers will be interested in new loans in general or specific products in particular (e.g. mortgages or credit cards). Willingness to apply for new loans (credit propensity) can be assessed in a similar way as credit risk (Bijak, 2011). Andreeva *et al.* (2005) used survival analysis to model propensity to purchase with a card. Attrition/churn scoring is employed to identify customers who are most likely to move to another lender or to close or stop using their accounts. The identified customers can be targeted with anti-churn campaigns to prevent unwanted events from occurring. Burez and van den Poel (2008) developed churn models using survival analysis and random forests. Since there may be a few types of unwanted events (including prepayment and default), competing risks analysis can be applied (Thomas, 2009a).

Collection/recovery scoring is used to support the choice of appropriate actions against customers who are past due on their obligations. It can also help value portfolios for sale. Similarly to application and behavioural scoring, collection/recovery scoring can be classified as either entry (applied before the first contact) or sequential. Sequential models can repeatedly predict migration to a worse level of delinquency based on data on the lender's actions and the customer's responses to them (Anderson, 2007). Fraud scoring is employed to detect and prevent possible frauds. There are two main types of such models and techniques: application and transaction fraud scoring (Anderson, 2007). The former is similar to application scoring but its objective is to identify potential fraudsters, i.e. applicants who deliberately would not repay their loans, whereas the latter can be e.g. part of processing credit card transactions. Both supervised and unsupervised methods can be used to produce 'suspicion scores' (Bolton and Hand, 2002). In order to recognise unusual transaction patterns, NNs are often applied (Anderson, 2007).

1.3 Modelling problems

There are some common modelling problems that occur in social sciences and economics. They are also familiar to modellers in credit scoring. Three selected problems (heterogeneity, uncertainty and dynamics) are briefly discussed below. It is

described how the next chapters of this thesis address them, each in one of the following areas of credit scoring: PD, LGD and affordability modelling.

1.3.1 Heterogeneity

Heterogeneity of populations is an inherent problem in social and economic modelling, since all people as well as all economic agents (consumers, customers, companies etc.) are unique. There is both observed and unobserved heterogeneity. It is never feasible to collect data on all characteristics that differentiate the population members. One can control only for some variables in the model, assuming that the error terms capture the effect of the remaining characteristics (unobserved heterogeneity). Unobserved heterogeneity can be taken into account by using latent or instrumental variables, fixed or random effects, frailty terms etc. Improper treatment of unobserved heterogeneity may lead to false conclusions about relationships being drawn from the model (e.g. Heckman, 1981).

On the one hand, heterogeneity justifies and enables modelling. If populations were perfectly homogeneous with respect to the dependent variable, no models would be needed. If populations were perfectly homogeneous with respect to the characteristics, it would not be possible to build any models. On the other hand, heterogeneity may potentially hamper modelling when there are a number of subpopulations with unique relationships between the characteristics and the dependent variable. In credit scoring, this problem has been recognised for some time, especially in PD modelling (e.g. Makuch, 2001). In response, various segmentation methods were proposed (e.g. Siddiqi, 2005). It is often believed that segmentation improves the model performance.

Chapter 2 challenges this common belief. In that chapter, three segmentation methods are used: two popular two-step approaches and a new, simultaneous method. In the two-step approaches, logistic regression follows Classification and Regression Trees (CART) or Chi-square Automatic Interaction Detection (CHAID) trees. In the simultaneous method, called Logistic Trees with Unbiased Selection (LOTUS), both segmentation and scorecards are optimised at the same time. A single-scorecard logistic regression model serves as a reference. The above-mentioned segmentation methods are applied to the data provided by two of the major UK banks and one of the European

credit bureaus. Once the models are developed, their performance measures are compared to find out whether there is any improvement due to the methods used. The segmentation contribution is also assessed. Furthermore, it is analysed in what situations segmentation can improve the model performance and when the simultaneous approach can perform better than the two-step approaches.

1.3.2 Uncertainty

There are several sources of uncertainty in a model, including the stochastic nature of the model, measurement error and inability to capture all influences on the dependent variable. There is also model uncertainty. Uncertainty can be classified as either aleatory or epistemic (Wagenmakers *et al.*, 2008). Aleatory uncertainty arises from the fact that if it were possible to repeat the experiment many times, results would vary. By definition, it cannot be reduced. Epistemic uncertainty is related to inaccurate measurement, omitted variables etc. and thus can potentially be reduced. Frequentist (classical) statistics better deals with aleatory than with epistemic uncertainty such as uncertainty about parameters (Wagenmakers *et al.*, 2008).

The need for taking uncertainty into account in financial risk management has recently been emphasised and Bayesian methods are often recommended in this context (e.g. Böcker, 2010). Most credit scoring models are developed using frequentist statistics, though. As far as LGD for unsecured retail loans is concerned, a two-step approach can be employed, in which the two models are estimated independently (e.g. Matuszyk *et al.*, 2010). The use of the second model is conditional on the outcome of the first one but uncertainty is not propagated from one model to another. As a result, a part of uncertainty about the LGD estimates is ignored, which makes that approach problematic from the methodological point of view.

Chapter 3 suggests using Bayesian methods to model LGD, since Bayesian statistics offers a more coherent description of uncertainty than the frequentist framework. In the frequentist two-step approach, the first model (logistic regression) separates positive values from zeroes, whereas the second model (e.g. linear regression) allows for the estimation of the positive values. In the Bayesian framework, they are replaced with a single, hierarchical model, as Bayesian statistics enables an integrated estimation of

hierarchical models. For each loan, an individual predictive distribution of LGD is produced, rather than just a point estimate as in the two-step approach. The predictive distributions provide more information and offer more possibilities than the point estimates. They can be used, among other applications, in the LGD stress testing process and to approximate the downturn LGD. Both Bayesian methods and the frequentist approach are applied to the data on personal loans granted by a large UK bank.

1.3.3 Dynamics

Most social and economic phenomena are dynamic in nature. The macroeconomic environment is never static. Many characteristics of economic agents are time-variant. The relationships between the characteristics and the dependent variable may also vary over time. Dynamics is reflected in numerous economic theories. Examples include the Life-Cycle Theory (Modigliani and Brumberg, 1954) as well as various theories of business cycle developed by different schools of economic thought (Snowdon and Vane, 2005). Dynamics can be taken into account by using time series, panel data (time-series cross-sections), time-variant regressors such as macroeconomic variables etc.

In credit scoring, though, most approaches are static and introducing dynamics into models has been recognised as one of the current challenges (Crook and Bellotti, 2008; Thomas, 2011). One of the areas where there is a need for incorporating dynamics is affordability assessment. In the UK, both the Office of Fair Trading (OFT) and the Financial Services Authority (FSA) recommend a long term perspective and taking into consideration variability of the applicant's income and expenditure over time when assessing affordability. Nevertheless, a static approach is often used in practice.

Chapter 4 proposes a theoretical framework for dynamic affordability assessment. Affordability is defined as a function that assigns to each possible instalment amount a probability of the applicant defaulting over the loan repayment period. Affordability assessment consists in the estimation of this function. Both income and consumption are allowed to vary over time. Their changes are modelled with random effects models for panel data which are derived from the economic literature, including the Euler equation approach. Once the models are estimated, they are applied in a simulation that is run for

a given applicant. Each iteration generates a pair of the predicted income and consumption time series. On this basis, the applicant's ability to repay is assessed over the life of the loan and for all possible instalment amounts. As a result, each amount is assigned with probabilities of default and failure to pay. This allows for the identification of the maximum affordable instalment. The proposed approach is illustrated with an example based on artificial data.

1.4 Thesis structure

The thesis is structured as follows. The next three chapters focus on the problems discussed in section 1.3. Chapter 2 deals with segmentation in the context of modelling PD. In Chapter 3, Bayesian methods are used to model LGD for unsecured retail loans. In Chapter 4, the theoretical framework for dynamic affordability assessment is presented. Chapter 5 includes a summary, conclusions and recommendations. In the beginning of each chapter, there is a short description of its contents and structure.

Chapter 2

Does segmentation always improve model performance in credit scoring?[1]

2.1 Introduction

A scoring model describes the relationship between a customer's characteristics (independent variables) and his or her creditworthiness status (a dependent variable). A customer's status can be either 'good' or 'bad' (and sometimes also 'indeterminate' or 'other'). The most common form of scoring models is referred to as a 'scorecard'. According to Mays (2004), the scorecard is "a formula for assigning points to applicant characteristics in order to derive a numeric value that reflects how likely a borrower is, relative to other individuals, to experience a given event or perform a given action" (p. 63). Different points are assigned to different attributes of a characteristic (values of a variable). Scorecards are used to calculate scores and/or probabilities of default. They are sometimes scaled to obtain a required relationship between scores and PD. A scoring model can consist of one or more scorecards. In the latter case, it can be referred to as a 'suite of scorecards'. In order to develop such a multi-scorecard model, segmentation is applied.

It is commonly expected that segmentation will improve the model performance. Segmentation is often carried out using the two-step approaches, where logistic regression follows Classification and Regression Trees (CART) or Chi-square Automatic Interaction Detection (CHAID) trees. In this research, these approaches are applied as well as Logistic Trees with Unbiased Selection (LOTUS). The latter is a simultaneous method, in which both segmentation and scorecards are optimised at the same time. A single-scorecard logistic regression model is used as a reference. All these

[1] This chapter is based on the following paper: Bijak, K. and Thomas, L.C. (2012) Does segmentation always improve model performance in credit scoring?, *Expert Systems with Applications*, 39(3), pp. 2433-2442.

methods are applied to the data provided by two of the major UK banks and one of the European credit bureaus. Once the models are developed, the obtained results are analysed to examine whether there is an improvement in the model performance (in terms of the discriminatory power) due to the segmentation methods used. Moreover, the segmentation contribution is assessed. Finally, it is discussed in which situations segmentation improves the model performance and when the simultaneous approach outperforms the two-step approaches.

This chapter is structured as follows. In section 2.2, the theoretical background of segmentation is presented as well as segmentation methods and other researchers' findings on its impact on the model performance in credit scoring. In section 2.3, the basics of logistic regression, CART, CHAID and LOTUS are introduced. In section 2.4, the three datasets are described. Section 2.5 is on the empirical results. Section 2.6 is a discussion on when segmentation can improve the model performance, and section 2.7 includes the research findings and conclusions.

2.2 Background

2.2.1 Segmentation

For a long time, segmentation has been a key element of marketing (Wedel and Kamakura, 2000). According to the original definition by Smith (1956), market segmentation is a strategy of "viewing a heterogeneous market […] as a number of smaller homogeneous markets in response to differing product preferences among important market segments". In credit scoring, segmentation can be defined as "the process of identifying homogeneous populations with respect to their predictive relationships" (Makuch, 2001, p. 140). The identified populations are treated separately in the process of a scoring model development, usually because of possible unique relationships between a customer's characteristics and the dependent variable.

Nowadays segmentation is widely used in the industry. There are various segmentation drivers, i.e. factors that can drive the division of a scoring model into two or more scorecards. Thomas *et al*. (2001) classify them into strategic, operational and variable interactions. Segmentation for strategic reasons is aimed at varying strategies for

different groups of customers, whereas operational reasons are related to differences in the scope of available characteristics etc. An interaction occurs when the relationship between a characteristic and the dependent variable varies amongst groups with different attributes of another characteristic.

Similarly but not identically, Anderson (2007) classifies segmentation drivers into: marketing, customer, data, process and model fit factors. The first four factors reflect, respectively, the special treatment of some market segments, or customer groups, data issues (such as data availability) and business process requirements (e.g. different definitions of a dependent variable). The model fit relates to interactions within the data and using segmentation to improve the model performance. In this research, the focus is on segmentation which is driven by the model fit factors.

There are two key concepts related to segmentation: a segmentation basis and a segmentation method. A segmentation basis is a set of variables that allow for the assignment of potential customers to homogeneous groups. Segmentation bases can be classified as either general or product-specific, and either observable or unobservable (Wedel and Kamakura, 2000). General bases are related to the customer but independent of products, whereas product-specific bases depend on both the customer and the product (e.g. a loan). Contrary to unobservable bases such as intentions, observable bases can be directly measured. As far as scorecard segmentation is concerned in this research, there is an unobservable product-specific basis. Once the segmentation is implemented, customers are grouped on the basis of their unobservable behavioural intentions to repay the loans or the relationship between their intentions and characteristics. On the date of grouping, it is not known whether they are going to repay or not.

According to Wedel and Kamakura (2000), there are six criteria for effective segmentation: identifiability, substantiality, accessibility, stability, responsiveness and actionability. Identifiability means that customers can be easily assigned to segments. Substantiality guarantees sufficient size of segments from the profitability point of view. Accessibility ensures that segments can be reached using available tools. Stability is defined as time invariability. Responsiveness means that segments differ from each

other in their response/behaviour, and actionability refers to the possibility of taking effective actions towards them. Unobservable product-specific bases which contain behavioural intentions are characterised by good identifiability and substantiality, moderate stability and very good responsiveness (Wedel and Kamakura, 2000). They are also characterised by poor accessibility and actionability but these criteria seem to be less important when segmentation is driven by the model fit factors. The above-mentioned features make the unobservable product-specific bases promising as far as scorecard segmentation is concerned.

2.2.2 Segmentation methods

Segmentation methods can be classified as either associative (descriptive) or regressive (predictive) approaches (Aurifeille, 2000; Wedel and Kamakura, 2000). Since the ultimate goal is to assess the credit risk, the latter are applied in this research. There are two types of regressive approaches: two-step (a priori) and simultaneous (post hoc) methods (Aurifeille, 2000; Wedel and Kamakura, 2000). In the two-step approaches, segmentation is followed by the development of a regression model in each segment. In the simultaneous methods, both segmentation and regression models are optimised at the same time.

The two-step approaches are not designed to yield optimal results in terms of the prediction accuracy but rather to aid the understanding of overall strategy. On the other hand, the simultaneous methods give priority to a low, tactical level rather than to a high, strategic level of decision: the optimisation objective is to obtain the most accurate prediction, and not necessarily a meaningful and easily understandable segmentation (Desmet, 2001).

There is not much literature on segmentation methods in credit scoring. According to Siddiqi (2005), segmentation methods can be classified as either experience-based (heuristic) or statistical. As far as the experience-based methods are concerned, one approach is to define segments that are homogeneous with respect to some customers' characteristics. This allows for the development of segment-specific variables. For example, creating a segment of customers who have credit cards enables the construction of such characteristics as credit card utilisation rate. Another approach is to

define segments that are homogeneous with respect to the length of customers' credit history (cohorts) or data availability (thin/thick credit files). For instance, creating a segment of established customers allows building behavioural variables based on the data from the last 12 months, the last 24 months etc.

Furthermore, when there is a group (e.g. mortgage loan owners or consumer finance borrowers) that is expected to behave differently from other customers, or for whom the previous scoring model turns out to be inefficient, it may be worth creating a separate segment for such a group. Moreover, customers can be grouped into segments in order to make it easier for the bank to treat them in different ways, e.g. by setting different cut-offs, i.e. score thresholds used in the decision making (Thomas, 2009a).

Finally, segmentation can be based on variables (e.g. age) that are believed to have strong interactions with other characteristics (Thomas, 2009a). This is a heuristic approach but it has been developed into statistical methods based on interactions. An alternative to segmentation based on the selected variable is to include all its interactions with other characteristics in a single-scorecard model (Banasik *et al*., 1996). However, such a model has a large number of parameters and is less understandable than a multi-scorecard one. Therefore, Thomas *et al*. (2001) suggest including only single interactions in a model. They recommend segmenting the population instead, if there is a variable that has strong interactions with a number of other characteristics.

The experience-based segmentation methods can help achieve various goals such as improving the model performance for a certain group of customers or supporting the decision making process. The experience-based segmentation may also allow for better risk assessment for the entire population of customers. There is no guarantee, though, that segmentation which intuitively seems reasonable will increase the model performance (Makuch, 2001).

As far as statistical methods are concerned, segmentation can be carried out using statistical tools as well as data mining and machine learning techniques. One approach is to do the cluster analysis (Siddiqi, 2005). The cluster analysis can be conducted using hierarchical clustering, the *k*-means algorithm or Self-Organising Maps (SOMs).

Regardless of the algorithm applied, clustering is based on customers' characteristics. Therefore, customers with different demographic or behavioural profiles are classified into different segments. The resulting groups are homogeneous with respect to the characteristics but, since the customer's status is not used in segmentation, they do not need to differ in risk profiles.

Another approach is to use tree-structured classification methods such as CART or CHAID (VantageScore, 2006). In this approach, grouping is based on the customer's status and thus, segments differ in risk profiles. Both the cluster analysis and classification trees can constitute the first step in the two-step regressive approaches.

However, the classification trees often yield sub-optimal results (VantageScore, 2006). In 2006 VantageScore introduced a new, multi-level segmentation approach: combining experience-based segmentation (at higher levels) and segmentation based on a dedicated score (at lower levels). The score is calculated using an additional scoring model that has to be built first. The split points on the score are determined using CART. Using the score enables dividing customers in such a way that in each segment, customers are similar to one another as far as their risk profile is concerned. There is an assumption that different risk profiles are associated with different relationships between the dependent variable and a customer's characteristics. The VantageScore approach makes it easier for the bank to treat subprime and prime customers in different ways, but it seems that this approach is not necessarily optimal in terms of the model performance.

There have also been some attempts to develop methods that would allow for the optimal segmentation, i.e. a segmentation that would maximise the model performance. They can be classified as the simultaneous methods. Hand *et al*. (2005) suggested a method for the optimal division into two segments. In both segments, the same set of variables is used to develop a scorecard. The optimal division into the two groups is found using exhaustive search (each possible split point is examined on each variable or the linear combination of variables). For each possible pair of segments, two logistic regression models are built. The fit of the two-scorecard model is assessed using its overall likelihood, i.e. a product of likelihoods of the scorecards, and the division is chosen that gives the highest overall likelihood. However, the adopted assumptions

(only two segments, the same variables) result in the limited usefulness of the suggested method. In banking practice, customers are usually divided into at least several segments, in which different sets of variables are used.

Another approach to the optimal segmentation is FICO's Adaptive Random Trees (ART) technology (Ralph, 2006). In this approach, trees are not built level by level as in most tree-structured classification methods. In the beginning, the trees are randomly created using some predefined split points on the possible splitting variables. Then a GA is applied to find the best tree, i.e. the tree which gives the highest divergence in the system of scorecards in its leaves, where the scorecards are naïve Bayes models. In all models, there is the same set of characteristics as in the parent scorecard that is built on the entire sample.

The ART technology has fewer drawbacks than other methods. It should allow for the maximisation of the model performance (measured with divergence). The number of segments is not predetermined. The use of the GAs avoids the exhaustive search that is both expensive and time-consuming. However, there is still a serious disadvantage, since – as in Hand *et al*. (2005) – the same set of variables is used in all scorecards. This disadvantage is shared by many simultaneous methods, including those which probably have not been used in credit scoring yet, e.g. clusterwise logistic regression (Qian *et al*., 2008).

2.2.3 Impact of segmentation

It is commonly asserted by scorecard developers that a suite of scorecards allows for better risk assessment than a single scorecard used for all customers. According to Makuch (2001), who measured model performance using the KS statistic, segmentation usually increases performance by 5-10 per cent in comparison with a single-scorecard system. It is also believed that segmentation itself can significantly contribute to performance of a scoring model.

The impact of segmentation on the model performance measures can be assessed using simulated results of random scorecards applied to the identified segments (Thomas, 2009a). The segmentation contribution to the model performance can also be assessed

using difference between a performance measure of the model and the weighted average amongst the scorecards. The average is calculated using weights equal to percentages of customers classified to the segments.

Banasik *et al*. (1996) analysed the impact of some experience-based divisions on the discrimination of a model. They set a few cut-offs and measured the discrimination in terms of errors that occur on a holdout sample. As a result, they found that "it is not the case that creating scorecards on separate subpopulations is necessarily going to give better discrimination than keeping to one scorecard on the full population". For a suite of scorecards, it is difficult to choose cut-offs that are independent, good and robust at the same time. However, if cut-offs are chosen in the same way for all models, multi-scorecard models reject fewer applicants than single-scorecard ones (Banasik *et al*., 1996). This may also be considered an advantage of segmentation.

2.3 Methodology

2.3.1 Logistic regression

Logistic regression is the most commonly used method for developing scoring models. Since there is a binary dependent variable (either good or bad), binomial logistic regression is applied. In binomial logistic regression, the dependent variable *y* is equal to the cumulative distribution function (CDF) *F* of a logistic distribution:

$$y = F(\boldsymbol{\beta x}) = \frac{1}{1+e^{-\boldsymbol{\beta x}}}$$

where $\boldsymbol{x}$ is a vector of independent variables (covariates) and $\boldsymbol{\beta}$ is a vector of model parameters (Greene, 2000, p. 815). The parameters are usually estimated using the Maximum Likelihood (ML) method. The estimated value of the dependent variable lies between zero and one. Thus, it can be interpreted as a probability of the dependent variable being equal to one. In credit scoring, this is a probability of the customer being bad (probability of default).

In scorecards, covariates are often used in the form of Weights of Evidence (WoE). If a discrete or discretised variable X takes K values, then the WoE for its nth value ($n \leq K$) is computed using the following formula (Anderson, 2007, p. 192):

$$WoE_n = ln\left(\frac{P(n \mid G)}{P(n \mid B)}\right) = ln\left[\left(G_n \Big/ \sum_{k=1}^{K} G_k\right) \Big/ \left(B_n \Big/ \sum_{k=1}^{K} B_k\right)\right]$$

where G_n (B_n) is the number of Goods (Bads) for whom X takes the nth value. Weights of Evidence allow for the assessment and comparison of the relative credit risk associated with different attributes of a characteristic. The advantage of using Weights of Evidence is that scorecards are more parsimonious and thus more robust than when coding characteristics as sets of dummies.

The ratio of Goods to Bads is referred to as the 'odds' in credit scoring. The population odds are the ratio of the proportion of Goods p_G to the proportion of Bads p_B in the population. It is often assumed that there is a linear relationship between the score and the log odds (Mays, 2004). Using Bayes' theorem, it can be shown that the log odds s_n amongst customers, for whom X takes the nth value, are equal to a sum of the population log odds s_{pop} and the WoE for the nth value of X (Thomas, 2009a, p. 33):

$$s_n = ln\left(\frac{P(G \mid n)}{P(B \mid n)}\right) = ln\left(\frac{P(n \mid G)\,p_G}{P(n \mid B)\,p_B}\right) = ln\left(\frac{p_G}{p_B}\right) + ln\left(\frac{P(n \mid G)}{P(n \mid B)}\right) = s_{pop} + WoE_n$$

Sometimes there is no theory that would support the choice of covariates. In such case, the best set of covariates can be identified using the stepwise selection of variables (Hosmer and Lemeshow, 2000). The stepwise selection is a procedure of alternate inclusion and exclusion of variables from a model based on the statistical significance of their coefficients that is measured with a p-value. In logistic regression, the likelihood ratio test or the Wald test are used to assess the significance of the coefficients. In both tests, the chi-square test statistics are computed. In a forward selection step, such a variable is included that, once added to the model, it has the most significant coefficient. In a backward elimination step, the variable which has the least significant coefficient is excluded from the model. The stepwise selection is especially

useful in case of a large number of possible covariates. Therefore, it is popular in behavioural scoring.

The goodness-of-fit of a logistic regression model can be measured e.g. using the deviance. In logistic regression, the deviance plays the same role as the residual sum of squares in linear regression. It is calculated according to the following formula:

$$D = -2\sum_{i=1}^{n}\left[y_i ln\left(\frac{\hat{p}_i}{y_i}\right) + (1-y_i) ln\left(\frac{1-\hat{p}_i}{1-y_i}\right)\right]$$

where y_i is the dependent variable value and $\hat{p}_i$ is the estimated probability of $y_i = 1$ for the *i*th observation, $i = 1, \ldots, n$ (Hosmer and Lemeshow, 2000, p. 13).

2.3.2 CART

Classification and Regression Trees (CART) are a popular nonparametric statistical method (Breiman *et al*., 1998). In this research, the focus is on classification trees, i.e. those with nominal dependent variables, as opposed to regression trees, where dependent variables are continuous. In CART, predictors can be both continuous and categorical, while splits are binary. All possible splits on all variables are examined and assessed. In order to measure quality of a split, the impurity function values are calculated for both child nodes. The impurity is often assumed to take the form of the entropy:

$$I(N) = (-p) log(p) - (1-p) log(1-p)$$

or the Gini index:

$$I(N) = 2p(1-p)$$

where p is a fraction of observations with a positive response (value of the binary dependent variable) in the node N (Izenman, 2008, p. 288). Once all splits are assessed,

such a split of the node N into N_1 and N_2 is selected that results in the largest decrease in impurity (Breiman *et al*., 1998, p. 32):

$$IG(N_1, N_2) = I(N) - \frac{|N_1|}{|N|} I(N_1) - \frac{|N_2|}{|N|} I(N_2)$$

The tree is grown using the recursive partitioning, i.e. each child node is split in the same way (Berk, 2008). The growing process continues until no more nodes can be split. In order to avoid excessively large structures and overfitting, the tree is then pruned back. The pruning process consists in minimising the cost-complexity measure that is defined as follows:

$$R_\alpha(T) = R(T) + \alpha|T|$$

where $R(T)$ is an estimate of the misclassification cost of the tree T and α is the complexity parameter, while $|T|$ denotes the number of leaves (Breiman *et al*., 1998, p. 66). For each value of α, such a tree can be identified that minimises the cost-complexity measure (if there are two or more such trees, the smallest one is selected). As α increases, new minimising trees appear only for some values of α. As a result, there are a certain number of the minimising trees. It can be demonstrated that they constitute a sequence of nested subtrees (Breiman *et al*., 1998). This sequence is identified on a training sample. Then the cost-complexity of each subtree is assessed either on a test sample or using cross-validation. On this basis, the final tree is selected amongst the subtrees. In this research, test samples have been used in pruning the CART trees. Splits have been selected using the Gini index as the impurity function. CART has served as the first step in the two-step approach and the trees have been created in SAS Enterprise Miner.

The CART method is often compared to the C4.5 algorithm, another popular method for building classification trees (Hand *et al*., 2001; Larose, 2005). Nevertheless, there are some important differences between them, e.g. the latter allows splitting into three or more child nodes (multi-way splits). Moreover, in the C4.5 algorithm, the split selection is always based on the information gain, i.e. reduction in entropy.

2.3.3 CHAID

Chi-square Automatic Interaction Detection (CHAID) is also a tree-structured classification method (Kass, 1980). It belongs to the family of methods known as Automatic Interaction Detection (AID). As its name suggests, AID allows for the detection of interactions between variables. Thus, the segmentation is based on the interactions. The standard AID can be described as "a stepwise application of a one-way analysis of variance model", since the sequential partitioning of the dataset is driven by maximising the between group sum of squares of the (continuous) dependent variable (Abid Ali *et al*., 1975). AID requires that the dataset is partitioned on the basis of predictors which are categorical, i.e. either discrete or discretised (if originally continuous).

Contrary to the standard AID, where splits can only be binary, CHAID allows for multi-way splits (Kass, 1980). In CHAID, the dependent variable has to be nominal, and the split selection is based on the chi-square tests of independence between the grouped predictors and the dependent variable. The original categories of each predictor can be grouped into a number of classes using a stepwise procedure that includes both merging and splitting steps (Hawkins and Kass, 1982). In a merging step, all categories or classes which can be merged are compared to one another using the above-mentioned tests. The least significantly different ones are then grouped into a new class. In a splitting step, all possible binary divisions of this class are analysed and such a division is selected that leads to the most significantly different classes. Only the classes which consist of three or more categories can be divided. The procedure continues until no more merging is possible. Among the grouped predictors, the one is used to split the node that produces the most significant split. In order to account for multiple testing, the Bonferroni correction is used (Hawkins and Kass, 1982). The Bonferroni correction adjusts the test significance level for numerous tests that are performed at the same time on the basis of the Bonferroni inequality (e.g. Hand *et al*., 2001).

Once a node is split, the grouping and testing process is repeated for each child node. Growing the tree stops when there are no more nodes that can be split. No pruning is carried out. Nevertheless, in this research, manual pruning has been performed to ensure that in each leaf there are enough Bads to build a logistic regression model. Similarly to

CART, CHAID has been used as the first step in the two-step approach and the trees have been produced in SAS Enterprise Miner. CART and CHAID have been chosen for this purpose, since they were mentioned by VantageScore (2006) as typical segmentation methods used in credit scoring, and CHAID was also suggested by Anderson (2007) as an approach to identify the best scorecard split.

Classification trees, including CART and CHAID, can be employed not only for segmenting customers but also for developing scoring models (Makuch, 2001; Thomas *et al*., 2002; Yobas *et al*., 2004). They can be applied instead of e.g. logistic regression. In such an application, each customer can be assigned a probability of default equal to the bad rate in the leaf that he or she falls into.

2.3.4 LOTUS

There is selection bias in CART (but not CHAID) and in all other methods where exhaustive search is used for variable selection (Chan and Loh, 2004). If all possible splits based on all variables are considered, then variables with more unique values are more likely to be selected to split the node. Chan and Loh (2004) proposed a new method, in which the selection bias problem is overcome: the Logistic Tree with Unbiased Selection (LOTUS) algorithm. The algorithm allows for the development of classification trees with logistic regression models in their leaves (Chan and Loh, 2004; Loh, 2006). Since the trees are built together with the models, this is a simultaneous method.

Before the algorithm starts, continuous variables need to be classified into f-, s- and n-variables. F-variables are potential regressors and s-variables are allowed to split the nodes, whereas n-variables can be used in both roles. Categorical (both ordinal and nominal) variables can be used only for splitting. The pseudocode for the algorithm is presented in Figure 2.1. The algorithm starts with a regression model developed using the entire training sample (at the root). Once a node is split, new models are built in the child nodes. As far as the models are concerned, one can use either multiple regression (with or without stepwise selection) or the simple regression which gives the lowest deviance. As a result, a different set of regressors may be used in each model. In order

to avoid the bias, the split selection is divided into two separate steps: variable selection and split point selection (Chan and Loh, 2004). Only binary splits are allowed.

```
LOTUS {
    model_development(root)
    repeat until impossible to split a node or to develop a model {
        for each node {
            split_variable_selection(node)
            split_point_selection(node, split_variable)
            node_split(node, split_variable, split_point)
            model_development(node_1)
            model_development(node_2)
        }
    }
    CART_pruning(tree)
}

split_variable_selection(node) {
    discretise n- and s-variables
    for each n-variable that is used as a regressor {
        calculate the trend-adjusted chi-square statistic
    }
    for each other n-, s- or categorical variable {
        calculate the ordinary chi-square statistic
    }
    select split_variable with the lowest p-value
    return split_variable
}

split_point_selection(node, split_variable) {
    if split_variable is continuous or ordinal {
        determine selected quantiles
    }
    if split_variable is nominal {
        determine potential split points
    }
    select split_point with the lowest total deviance
    return split_point
}

node_split(node, split_variable, split_point) {
    split node into node_1 and node_2
}
```

Figure 2.1. Pseudocode for the LOTUS algorithm

In the first step, a simple discretisation method is applied to s- and n-variables. For the discretised and categorical variables, the chi-square statistics are computed as in tests of independence. The statistic used depends on whether the analysed variable serves as a regressor in the parent node, i.e. the node to be split. For all categorical and s-variables as well as for some n-variables, the ordinary chi-square statistic is calculated, while for those n-variables which are used as regressors, the trend-adjusted chi-square statistic is computed. The variable with the lowest *p*-value is selected to split the node.

With regard to the above-mentioned test statistics, the ordinary chi-square statistic can be decomposed into the Cochran-Armitage trend test statistic and the trend-adjusted chi-square statistic (Chan and Loh, 2004). The former is used to test for a linear trend in proportions (Cochran, 1954; Armitage, 1955). The latter can be applied to test for independence after adjusting for a linear trend. It is implemented in LOTUS to distinguish nonlinear effects from linear ones. If the ordinary chi-square statistic were used instead, the variables which – like regressors – have strong linear effects would be more likely to be selected to split the node. This would lead to another selection bias (Chan and Loh, 2004).

In the second step, only some values of the selected variable are taken into account and thus, exhaustive search is avoided. If the variable is either continuous or ordinal, five sample quantiles are considered as potential split points. If the variable is nominal, its values are arranged in order of the proportion of cases with a positive response in the node. It is assumed that potential split points surround such a value that, if used for splitting, minimises the sum of the variances of the response variable in the resulting subsets. In consequence, five potential split points are considered. Eventually, amongst the potential split points, the one is selected which minimises the total deviance, i.e. the sum of deviances of regression models built in the child nodes.

The algorithm stops when there are too few observations to split a node or to develop a model. The CART pruning method is then used to prune the tree. The pruning process can be performed either on a test sample or using cross-validation. The cost-complexity measure is based on the total deviance (summed over all leaves). Finally, the subtree with the lowest total deviance is selected (Chan and Loh, 2004).

In this research, LOTUS has been chosen, since it is one of the very few methods which enable a simultaneous optimisation of segmentation and logistic regression models and allow for different sets of variables in the models. It has also been important that the LOTUS software, in which the algorithm is implemented (Chan, 2005), can process large datasets. In the software, such options had been chosen that logistic regression models have been built using the stepwise selection and the pruning process has been based on test samples.

2.3.5 Discriminatory power measures

In credit scoring, it is important not only how well the model fits the data but also how effectively it separates Goods and Bads. As mentioned in section 1.2.1, this separation ability is referred to as the 'discriminatory power'. There are a number of discriminatory power measures, e.g. the Gini coefficient and the KS statistic.

Both the Gini coefficient and the KS statistic can be calculated using the CDFs of scores, computed separately for Goods and Bads (Thomas, 2009a). The discriminatory power measures can be derived from the Receiver Operating Characteristic (ROC) curve, drawn by plotting the above-mentioned CDFs against each other. The KS statistic is equal to the maximum difference between these CDFs. It can also be obtained as the maximum vertical distance between the ROC curve and the diagonal. The Gini coefficient is equal to the double area under the ROC curve (AUROC, also known as AUC) less one. Similarly to the KS statistic, it takes values between zero and one with higher values meaning stronger discriminatory power. It can be demonstrated that the Gini coefficient is the probability that a randomly selected Good will have a higher score than a randomly selected Bad (Thomas *et al.*, 2002).

Among other discriminatory power measures, there are divergence, information value and the Mahalanobis distance. Divergence is given by the following formula (Thomas, 2009a, p. 105):

$$D = \int \big(f(s|G) - f(s|B)\big) ln\left(\frac{f(s|G)}{f(s|B)}\right) ds$$

where s is a score and $f(s|G)$ and $f(s|B)$ are conditional probability density functions of scores for Goods and Bads, respectively. Information value is a discrete analogue to divergence, calculated for a number of score bands. If the score distributions are normal with identical variances: $N(\mu_G, \sigma^2)$ and $N(\mu_B, \sigma^2)$, divergence boils down to the square of the Mahalanobis distance (Thomas, 2009a, p. 108):

$$D = \frac{(\mu_G - \mu_B)^2}{\sigma^2} = D_M^2$$

Moreover, there are such discriminatory power measures as the Somers D-concordance statistic and the Mann-Whitney *U*-statistic. The relationship between these statistics, the Gini coefficient and AUROC is as follows (Thomas, 2009a, p. 113 and p. 120):

$$GINI = 2AUROC - 1 = D_S = 2\frac{U}{n_G n_B} - 1$$

where n_G and n_B are numbers of Goods and Bads, respectively.

AUROC and the related measures (including the Gini coefficient) are criticised for being incoherent in terms of misclassification costs (Hand, 2009). It can be demonstrated that AUROC is equivalent to the weighted average of the misclassification losses where a weight distribution depends on the score distributions and thus on the model used. Hand (2009) proposed an alternative to AUROC which is free from this disadvantage, since it is based on an objective weight distribution (the H measure uses a beta distribution). While the H measure is gaining popularity, the Gini coefficient and the KS statistic still remain the most commonly used measures of the discriminatory power.

2.4 Data

In this research, three real-world datasets are used. The data describe individual customers. There are two datasets containing application data and one dataset with behavioural (credit bureau) data. The datasets are referred to as A_1, A_2 and B, respectively.

In order to get unbiased results, each dataset has been randomly divided into training, validation and test samples. In all these samples, the bad rate is the same as in the original dataset. The datasets A_2 and B have been divided into the samples that contain ca 50, 30 and 20 per cent of customers, respectively. The samples which have been created as a result of the division of A_1 include ca 50, 25 and 25 per cent of customers (there would be an insufficient number of Bads in a smaller test sample).

The training samples have been used to develop models. The validation samples have served as holdout ones, i.e. they have not been used in the model development. Once a model had been built, its stability has been evaluated through the comparison of its discriminatory power on the training and validation samples. The smaller the difference, the more stable the model. The test samples have only been used to prune the trees.

2.4.1 Dataset A_1

The dataset A_1 has been provided by one of the major UK banks. There are data on 7835 applicants, of whom 6440 are Goods and 1395 are Bads. Originally, there had also been data on some rejected applications but they have been excluded from the dataset. The applications were made between April and September 1994. Customers applied for personal loans for different purposes. Loan amounts ranged from £500 to £50000, while repayment periods varied from 6 months to 5 years.

The characteristics are listed in Appendix A. They describe both a customer and a loan that he or she applied for. There are also some credit bureau variables in the dataset.

2.4.2 Dataset A_2

The dataset A_2 has been provided by another major UK bank. There are data on 39858 customers, including 38135 Goods and 1723 Bads. Originally, there had also been some Indeterminates who have been eliminated from the dataset. The loans were opened between May 1994 and August 1996. Loan amounts ranged from £300 to £15000, while loan terms (durations) varied from 6 months to 10 years.

In the original dataset, there have been 111946 customers. There have not only been application but also credit bureau data (see Appendix A). However, the additional data have been provided only for a part of the dataset. There are reasons to assume that the bank had such data for other customers, too. In order to account for this, the bad rate should be the same amongst customers with and without the credit bureau data (4.32%). All Goods and Bads, for whom there is the additional data, are included in the dataset. As far as customers without the credit bureau data are concerned, all Bads are included as well as such a number of randomly sampled Goods that the bad rate is equal to 4.32%. The resulting numbers of Goods and Bads are mentioned above.

2.4.3 Dataset *B*

The dataset *B* has been sampled from the database of a European credit bureau. This large and unique dataset has been generated and provided exclusively for this research. There are data on 186574 customers, of whom 179544 are Goods and 7030 are Bads. In the original dataset, there had also been data on some Indeterminates but they have been excluded. The customers had different credit products with different banks. Their characteristics and statuses have been determined by the credit bureau on the basis of data from the whole banking sector.

There are 324 characteristics based on the customer's credit history. They cannot be listed, though, since this is proprietary information. Some examples include: worst payment status within the last 12 months, number of credit inquiries within the last 12 months, number of loans granted within the last 12 months, number of open accounts, number of different products, number of past loans, time since last credit inquiry, time since last opening of an account, time since last delinquency over 30 days, total debt, total outstanding amount, total credit limit, credit card utilisation rate, sum of monthly instalments, number of banks the customer had accounts with etc. The other characteristics are related to various types of products as well as different time periods and payment statuses (describing delinquencies). The characteristics are as of the 1st of July 2008 (observation point) and the customer's status is as of the 1st of July 2009 (outcome point). Thus, the outcome period length is exactly equal to twelve months.

2.5 Results

In this research, suites of scorecards have been developed based on the datasets described in section 2.4. Both the two-step and simultaneous approaches have been adopted. In the two-step approaches, segmentation has been performed using CART and CHAID, and scorecards have been built for the identified segments. In the simultaneous approach, the LOTUS algorithm has been used to develop both segmentation and scorecards. For reference purposes, a single-scorecard model has been estimated based on each dataset. All the scorecards have been built using logistic regression with stepwise selection. No interaction variables have been allowed in the scorecards. The model performance is measured in terms of the discriminatory power.

The variable grouping process has been performed in the Interactive Grouping node in SAS Enterprise Miner. Categories of discrete variables have been grouped into classes, while continuous variables have been discretised (binned) first. For each variable, such a division has been selected that maximises reduction in entropy on the entire training sample. No more than five classes have been allowed. The groupings have sometimes been modified manually to put them in line with the banking experience.

2.5.1 Trees

In all the adopted approaches, only grouped variables and those original ones which are categorical have been allowed to split the nodes. If necessary, the CART and CHAID trees have been pruned back manually until there have been at least a minimum number of Bads in each leaf. This number has been assumed to be equal to 100 for the datasets A_1 and A_2 and 500 for the dataset *B*. The same minimum numbers of Bads have been set as an option in the LOTUS algorithm. Therefore, the final trees are rather compact. The CART, CHAID and LOTUS trees are presented in Figures 2.2-2.10.

In each leaf, the numbers represent: the number of Bads and the bad rate in the leaf as well as the number of all customers and their share in the training sample. In the CHAID tree for the dataset *B*, there is one leaf with only 16 Bads (marked with an asterisk in Figure 2.9). It has not been possible to prune the tree more because this leaf is a child node of the root. However, with such a number of Bads, it has not been possible to build a scorecard, either. Therefore, in this leaf all customers have been assigned the same probability of default that is equal to the bad rate (0.3%). As a result, there is no separating ability and both the Gini coefficient and the KS statistic are equal to zero in this leaf.

For each dataset, there is at least one variable that has been selected to split nodes in most trees based on this dataset. Time with Bank has been used in all trees for the dataset A_1. For the dataset A_2, all nodes have been split using either Loan Amount or Loan Purpose. For the dataset *B*, Var2 has been used in both the CART and the LOTUS trees. The variables Var1, Var2 and Var3 are based on the payment statuses of a customer's loans.

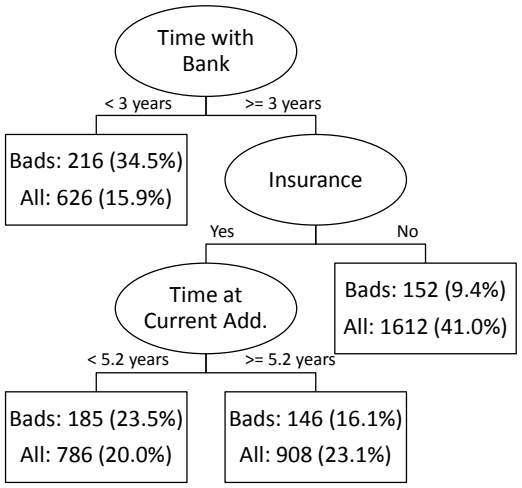

Figure 2.2. CART tree for the dataset A_1

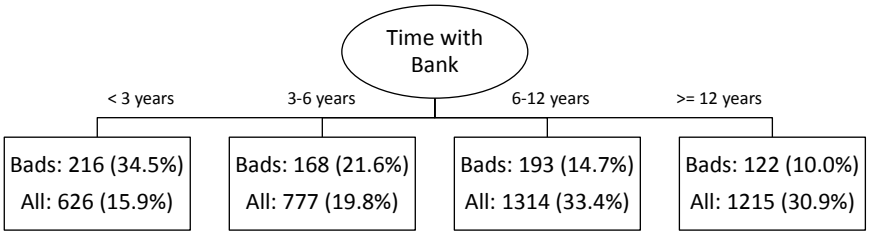

Figure 2.3. CHAID tree for the dataset A_1

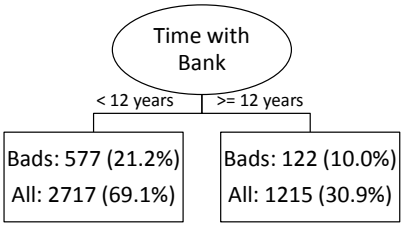

Figure 2.4. LOTUS tree for the dataset A_1

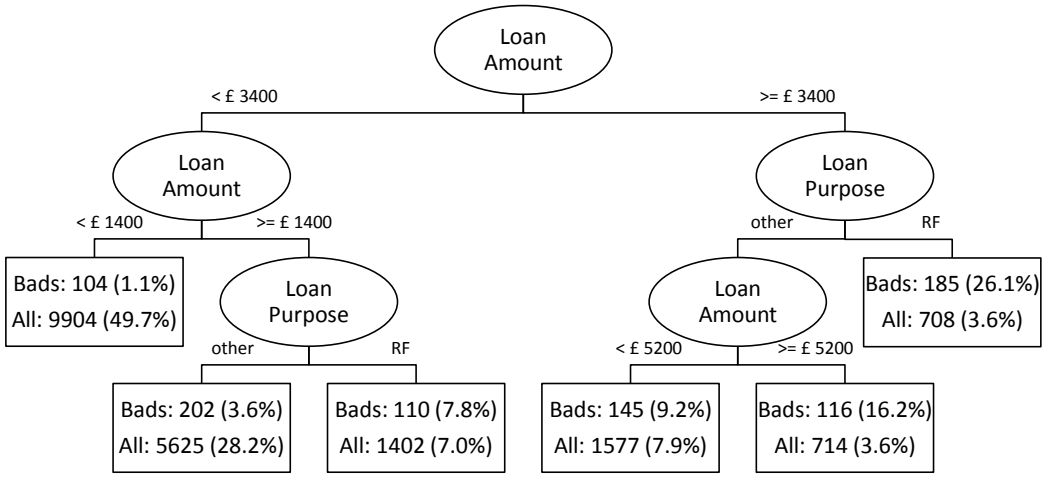

Figure 2.5. CART tree for the dataset A_2

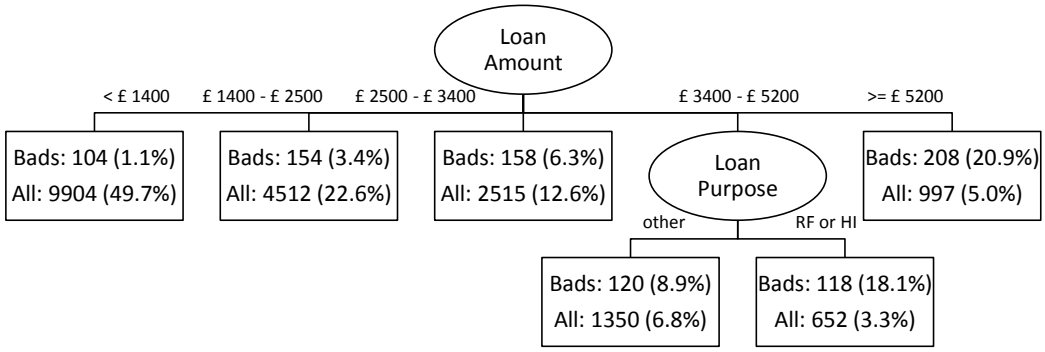

Figure 2.6. CHAID tree for the dataset A_2

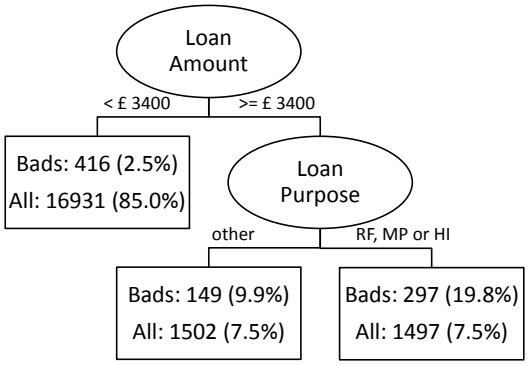

Figure 2.7. LOTUS tree for the dataset A_2

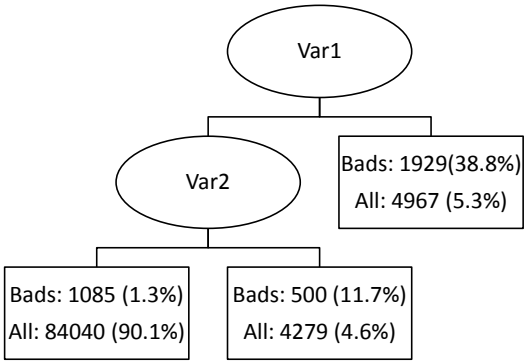

Figure 2.8. CART tree for the dataset *B*

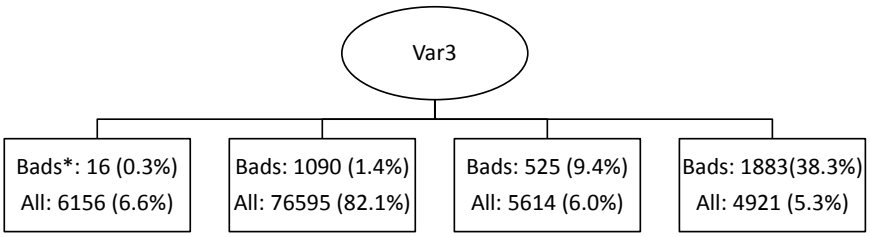

Figure 2.9. CHAID tree for the dataset *B*

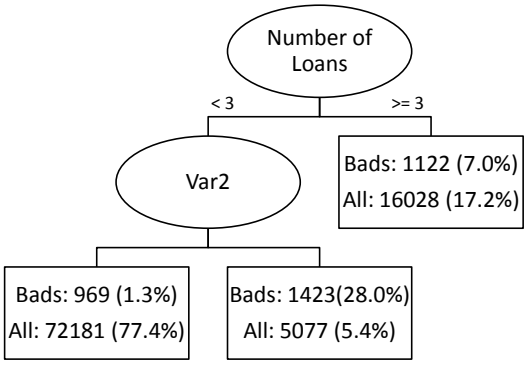

Figure 2.10. LOTUS tree for the dataset *B*

2.5.2 Scorecards

In all the developed scorecards, characteristics have been used in the form of WoE (based on the training sample). It has been assumed that no scorecard could consist of more than 10 characteristics, since in a credit scoring application, there are usually between 6 and 15 best variables (Anderson, 2007). In Appendix A, the characteristics which have been used in the reference logistic regression models based on the datasets A_1 and A_2 are marked with a bold font. In the reference scorecard based on the dataset *B*, there are, amongst other variables, Var1, number of credit inquiries within the last 9 months and age of the oldest loan. Some variables have been used both in the reference models and in the trees: Time with Bank and Insurance (A_1), Loan Amount and Loan Purpose (A_2) as well as Var1 (*B*).

In each suite, the scorecards are consistent in terms of scale, i.e. there is the same relationship between the score and PD. This enables the calculation of discriminatory power measures for the entire model. The Gini coefficients and the KS statistics are presented in Tables 2.1 and 2.2, respectively. There are values obtained on the training, test and validation samples. In addition, there are the estimates of means and standard deviations derived from the 100-fold bootstrap (in the brackets).

Only for the dataset A_1 do the multi-scorecard models perform slightly better than the reference logistic regression model: both the Gini coefficients and the KS statistics are higher by 2-3 percentage points on a training sample. For the other datasets, the differences in the Gini coefficient do not exceed one percentage point, which makes them negligible.

All the models for the dataset *B* are perfectly stable: the Gini coefficients and the KS statistics are very similar on the training and validation samples. The perfect stability is probably due to the size of the training sample and the power of the credit bureau variables. The models for A_2 are still stable, while those for A_1 cannot be considered stable: the Gini coefficients are lower by more than 10 percentage points on the validation sample as compared to the training sample (this is likely to be caused by overfitting, especially in case of using CART, CHAID and LOTUS, since there may be too many parameters for such a number of observations). For both A_1 and A_2, logistic

regression models are the most stable, probably due the smallest number of parameters and the simplest structure.

	Training sample	***Test sample***	***Validation sample***
	Dataset A_1		
CART	0.527 (0.527 ± 0.019)	0.374 (0.375 ± 0.034)	0.359 (0.364 ± 0.027)
CHAID	0.531 (0.529 ± 0.019)	0.392 (0.385 ± 0.031)	0.351 (0.353 ± 0.035)
LOTUS	0.520 (0.520 ± 0.020)	0.425 (0.427 ± 0.034)	0.386 (0.380 ± 0.027)
Logistic regression	0.499 (0.502 ± 0.020)	0.404 (0.404 ± 0.028)	0.397 (0.402 ± 0.032)
	Dataset A_2		
CART	0.663 (0.663 ± 0.014)	0.623 (0.625 ± 0.021)	0.618 (0.620 ± 0.021)
CHAID	0.664 (0.665 ± 0.012)	0.621 (0.621 ± 0.024)	0.622 (0.620 ± 0.019)
LOTUS	0.664 (0.664 ± 0.014)	0.634 (0.641 ± 0.026)	0.634 (0.633 ± 0.018)
Logistic regression	0.657 (0.658 ± 0.013)	0.640 (0.641 ± 0.024)	0.635 (0.637 ± 0.017)
	Dataset B		
CART	0.807 (0.807 ± 0.005)	0.813 (0.812 ± 0.010)	0.808 (0.808 ± 0.009)
CHAID	0.807 (0.807 ± 0.006)	0.814 (0.812 ± 0.011)	0.805 (0.806 ± 0.009)
LOTUS	0.805 (0.805 ± 0.006)	0.817 (0.818 ± 0.011)	0.803 (0.802 ± 0.008)
Logistic regression	0.801 (0.802 ± 0.006)	0.818 (0.819 ± 0.010)	0.807 (0.807 ± 0.009)

Table 2.1. Gini coefficient values for training, test and validation samples

	Training sample	***Test sample***	***Validation sample***
	Dataset A_1		
CART	0.389 (0.393 ± 0.018)	0.296 (0.303 ± 0.028)	0.267 (0.282 ± 0.022)
CHAID	0.386 (0.389 ± 0.017)	0.320 (0.319 ± 0.027)	0.283 (0.292 ± 0.028)
LOTUS	0.379 (0.385 ± 0.019)	0.344 (0.347 ± 0.027)	0.298 (0.305 ± 0.024)
Logistic regression	0.362 (0.370 ± 0.018)	0.317 (0.322 ± 0.024)	0.316 (0.319 ± 0.029)
	Dataset A_2		
CART	0.516 (0.513 ± 0.014)	0.479 (0.486 ± 0.021)	0.477 (0.484 ± 0.019)
CHAID	0.520 (0.523 ± 0.013)	0.469 (0.478 ± 0.023)	0.489 (0.490 ± 0.018)
LOTUS	0.502 (0.506 ± 0.014)	0.491 (0.499 ± 0.025)	0.487 (0.487 ± 0.019)
Logistic regression	0.497 (0.497 ± 0.014)	0.505 (0.508 ± 0.025)	0.485 (0.487 ± 0.018)
	Dataset B		
CART	0.705 (0.705 ± 0.006)	0.704 (0.702 ± 0.010)	0.701 (0.701 ± 0.009)
CHAID	0.705 (0.706 ± 0.006)	0.712 (0.711 ± 0.010)	0.696 (0.697 ± 0.009)
LOTUS	0.702 (0.702 ± 0.007)	0.710 (0.710 ± 0.011)	0.700 (0.699 ± 0.009)
Logistic regression	0.692 (0.693 ± 0.007)	0.708 (0.709 ± 0.011)	0.698 (0.698 ± 0.008)

Table 2.2. KS statistic values for training, test and validation samples

The Gini coefficients and the KS statistics which have been obtained on the validation samples for the datasets A_2 and B are similar for single- and multi-scorecard models. Nevertheless, on the validation sample for the dataset A_1, the discriminatory power

measures are higher by 3-5 percentage points for the logistic regression than for the CART- and CHAID-based models.

In order to test whether there are statistically significant differences in performance between the models for the datasets A_2 and B, each of the obtained values has been compared to the 95% bootstrap confidence intervals for the other models. The comparisons have been made for each sample separately. On their basis the following conclusions can be drawn. As far as the Gini coefficient is concerned, for each sample all obtained values fall within all confidence intervals. Thus, there is not enough evidence to reject the hypotheses that all models perform equally well. With regard to the KS statistic, there are some differences in performance between the models: the reference scorecards perform slightly worse than CHAID on the A_2 and B training samples, and the reference scorecard also performs slightly worse than CART on the B training sample. Although the above-mentioned differences are statistically significant – albeit only marginally – they are sufficiently small to be devoid of any practical significance. Furthermore, there is not enough evidence to reject the hypotheses that all models perform equally well on the test samples, and the same is true for the validation samples. This means that the slight superiority of CHAID and CART over the reference scorecards which has been observed on the training samples has not been confirmed on the test and validation samples for the datasets A_2 and B.

2.5.3 Segmentation contribution

For each approach, the segmentation contribution to the model performance has been assessed using the difference between the Gini coefficient or the KS statistic of the model and the weighted average amongst the scorecards. For comparison purposes, the discriminatory power measures have also been calculated for the CART, CHAID and LOTUS trees. In order to compute these measures, each customer has been assigned a probability of default equal to the bad rate in his or her segment. The results for the training samples are presented in Tables 2.3 and 2.4. There are the Gini coefficients and the KS statistics of the entire models (‘Model’) and the scorecard averages calculated using weights equal to the percentages of customers classified to the segments (‘Scorecards’). There are also the differences between the former and the latter (‘Difference’) as well as the discriminatory power measures of the trees (‘Tree’).

Moreover, for each of the above-mentioned measures there are the estimates of its mean and standard deviation derived from the 100-fold bootstrap (in the brackets).

For the dataset A_1, the trees are much weaker than the scorecards, the segmentation contribution does not exceed 9 percentage points and the scorecards are comparable to the logistic regression. As a result, the multi-scorecard models slightly outperform the single-scorecard one. For the datasets A_2 and *B*, both the Gini coefficients and the KS statistics of the trees are high, often higher than those of the scorecards. The segmentation contribution is up to even 20 percentage points (for example with regard to the KS statistic, it is equal to 0.209 in case of CHAID for *B*). However, the scorecards which have been built for the identified segments are much weaker than the logistic regression models developed on the entire training samples. Therefore, there is no difference in performance between the single- and multi-scorecard models.

	Model ***(1)***	***Scorecards*** ***(2)***	***Difference*** ***(1) – (2)***	***Tree***
	Dataset A_1			
CART	0.527 (0.527 ± 0.019)	0.442 (0.442 ± 0.001)	0.086 (0.086 ± 0.018)	0.328 (0.327 ± 0.019)
CHAID	0.531 (0.529 ± 0.019)	0.453 (0.453 ± <0.000)	0.077 (0.076 ± 0.019)	0.295 (0.295 ± 0.024)
LOTUS	0.520 (0.520 ± 0.020)	0.485 (0.485 ± <0.000)	0.036 (0.036 ± 0.020)	0.164 (0.163 ± 0.015)
	Dataset A_2			
CART	0.663 (0.663 ± 0.014)	0.502 (0.502 ± 0.001)	0.161 (0.161 ± 0.013)	0.567 (0.562 ± 0.017)
CHAID	0.664 (0.665 ± 0.012)	0.499 (0.499 ± 0.001)	0.165 (0.166 ± 0.012)	0.563 (0.565 ± 0.015)
LOTUS	0.664 (0.664 ± 0.014)	0.554 (0.554 ± <0.000)	0.110 (0.110 ± 0.014)	0.397 (0.398 ± 0.017)
	Dataset B			
CART	0.807 (0.807 ± 0.005)	0.671 (0.671 ± <0.000)	0.136 (0.136 ± 0.005)	0.634 (0.633 ± 0.008)

	Model ***(1)***	***Scorecards*** ***(2)***	***Difference*** ***(1) – (2)***	***Tree***
CHAID	0.807 (0.807 ± 0.006)	0.635 (0.635 ± 0.001)	0.172 (0.172 ± 0.006)	0.619 (0.619 ± 0.008)
LOTUS	0.805 (0.805 ± 0.006)	0.608 (0.608 ± <0.000)	0.197 (0.197 ± 0.006)	0.572 (0.571 ± 0.008)

Table 2.3. Gini coefficient values of models, scorecards and trees

	Model ***(1)***	***Scorecards*** ***(2)***	***Difference*** ***(1) – (2)***	***Tree***
	Dataset A_1			
CART	0.389 (0.393 ± 0.018)	0.353 (0.353 ± 0.001)	0.036 (0.039 ± 0.017)	0.261 (0.260 ± 0.018)
CHAID	0.386 (0.389 ± 0.017)	0.355 (0.355 ± <0.000)	0.031 (0.034 ± 0.017)	0.234 (0.235 ± 0.023)
LOTUS	0.379 (0.385 ± 0.019)	0.370 (0.370 ± <0.000)	0.009 (0.014 ± 0.019)	0.164 (0.163 ± 0.015)
	Dataset A_2			
CART	0.516 (0.513 ± 0.014)	0.395 (0.395 ± 0.001)	0.121 (0.119 ± 0.013)	0.443 (0.443 ± 0.016)
CHAID	0.520 (0.523 ± 0.013)	0.389 (0.389 ± 0.001)	0.130 (0.134 ± 0.012)	0.443 (0.443 ± 0.015)
LOTUS	0.502 (0.506 ± 0.014)	0.433 (0.433 ± <0.000)	0.070 (0.073 ± 0.013)	0.384 (0.384 ± 0.017)
	Dataset B			
CART	0.705 (0.705 ± 0.006)	0.514 (0.514 ± <0.000)	0.190 (0.190 ± 0.006)	0.615 (0.615 ± 0.008)
CHAID	0.705 (0.706 ± 0.006)	0.496 (0.496 ± <0.000)	0.209 (0.210 ± 0.006)	0.595 (0.594 ± 0.008)
LOTUS	0.702 (0.702 ± 0.007)	0.546 (0.546 ± <0.000)	0.156 (0.156 ± 0.006)	0.517 (0.517 ± 0.008)

Table 2.4. KS statistic values of models, scorecards and trees

In order to test whether the segmentation contributions are statistically significant, the corresponding 95% bootstrap confidence intervals have been examined. On the basis of these examinations the following conclusions can be drawn. With regard to both the Gini coefficient and the KS statistic, the contribution has turned out to be insignificant in case of LOTUS for A_1. As far as the KS statistic is concerned, the contribution has also been found insignificant in case of CHAID for A_1. For the other models and/or datasets the confidence intervals do not include zero and thus, the hypotheses that the segmentation contributions are null must be rejected. This means that for the datasets A_2 and B in particular, all segmentation contributions are statistically significant.

The results which have been obtained on the test and validation samples confirm that the segmentation contribution is lowest for the dataset A_1. It is not stable as the models for A_1 are not stable, either. On the other hand, the results show that if a dataset is large, all scorecards are stable and the segmentation contribution is stable as well.

2.6 Discussion

It can be surprising that there is no improvement in the model performance due to segmentation and the multi-scorecard models do not perform considerably better than the single-scorecard ones, especially on the credit bureau dataset. As far as the credit bureau is concerned, the population is highly heterogeneous because there are customers of different banks, using different products etc. It could be expected that segmentation would bring an improvement in risk assessment for this population.

There are a number of possible reasons for the lack of superiority of the multi-scorecard models. The distributions may be such that most of the separation between Goods and Bads is actually achieved by the single-scorecard models, so that there is little extra separation to be obtained by including the interactions implicit in the segmentation. Some characteristics may be effective on the entire sample but may lose their discriminatory power and/or independence from other variables in the identified segments. The sample sizes may be too small to allow for the identification of the optimal segments (very unlikely in case of the credit bureau dataset, though). Moreover, it cannot be excluded that applying other segmentation methods would lead to better multi-scorecard models. Other possible reasons include the adopted assumptions, in

particular the maximum number of characteristics in a scorecard: if fewer variables were allowed, segmentation might play a more important role.

It is worth seeing, in what situations segmentation improves the model performance and the simultaneous approach performs better than the two-step approaches. In order to show an example of such a situation, an artificial dataset has been constructed.

It is assumed that there is a random variable X and two simple logistic regression models based on this variable. In the first model, the parameter coefficient is equal to β, while in the second model it is equal to $-\beta$. It means that the relationship between X and the binary dependent variable Y is positive in the former and negative in the latter model. Values of Y are randomly generated using these two models. As a result, there are two groups of customers: G_1 and G_2. Their sizes do not have to be equal but should not differ much. In G_1, the bad rate is higher than in G_2. Subsequently, G_1 is split into G_{11} and G_{12} so that G_{12} is similar to G_2 in terms of the bad rate. Ultimately, there are three groups of customers: G_{11} (the first model, high bad rate), G_{12} (the first model, low bad rate) and G_2 (the second model, low bad rate).

In order to distinguish them from one another, a new variable Z is created. For different groups, Z takes random values from different, non-overlapping intervals, e.g. $[a, b)$ for G_{11}, $[b, c)$ for G_{12} and $[c, d)$ for G_2. It is determined for each customer separately. The artificial dataset contains three variables (X, Y and Z). There are training, validation and test samples having at least a few thousand customers each.

CART and CHAID should produce similar segmentations, where the sample is split on Z almost equal to b so that G_{11} is mostly in one node, while G_{12} and G_2 are mostly in another node. For illustration purposes, the ‘perfect’ split is presented in Figure 2.11. The high-bad-rate group would be separated from the low-bad-rate ones, since this is how the classification trees work. It may be difficult, though, to build a good scorecard in the node which contains both G_{12} and G_2 customers, as their data have been generated using the completely different models.

The LOTUS algorithm should split the sample on Z approximately equal to c so that G_{11} and G_{12} are mainly in one node and G_2 is mainly in another node. The 'perfect' split is demonstrated in Figure 2.12. The groups whose data come from the different models would be separated from each other, since LOTUS is aimed at identifying such divisions. This should allow for the development of good scorecards in both nodes. One can expect the scorecards to reflect the models which have been used to generate the data for G_1 and G_2 customers, respectively.

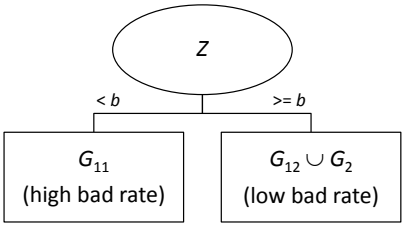

Figure 2.11. Generic CART/CHAID tree for an artificial dataset

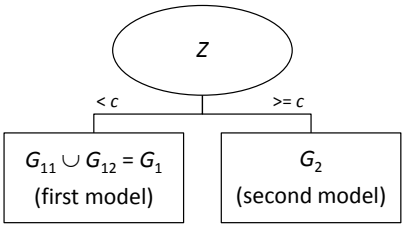

Figure 2.12. Generic LOTUS tree for an artificial dataset

The two-step approaches based on CART and CHAID as well as the LOTUS algorithm and logistic regression have been applied to an artificial dataset that has been constructed as described above. Depending on the group (G_1 or G_2), values of the dependent variable have been generated using one of the following models:

$$Y = \frac{1}{1+e^{-(-2+0.025X)}} \text{ and } Y = \frac{1}{1+e^{-(0.005-0.025X)}}$$

The error term (a random component) has been taken into account in the calculations. It has been assumed that the customer is bad if the result is greater than 0.5 and that he or she is good otherwise. Sizes and bad rates of the customer groups in the training sample are presented in Table 2.5. As far as the variable *Z* is concerned, its threshold values *b* and *c* have been assumed to equal –100 and –10, respectively.

Group	***Size***	***Bad rate***
G_{11}	750	52.3%
G_{12}	1250	14.4%
G_1	2000	28.6%
G_2	3000	14.5%
All	5000	20.2%

Table 2.5. Sizes and bad rates of the customer groups in the training sample (artificial dataset)

Both CART and CHAID have produced the same segmentation that separates the high-bad-rate and low-bad-rate groups of customers (see Figure 2.13). As expected, it has not been possible to build an effective scorecard in the second (mixed) node. Therefore, the entire model performs only slightly better than the single-scorecard one. In turn, the LOTUS algorithm has separated the groups whose data come from the different models (see Figure 2.14), which indeed has enabled the development of effective scorecards in both nodes. As a result, the simultaneous approach outperforms the two-step approaches on the artificial dataset, whereas the single-scorecard model performance is relatively poor, since both *X* and *Z* are weak variables on the entire sample (see Tables 2.6 and 2.7).

This is an example of a situation, when segmentation improves the model performance and the simultaneous approach outperforms the two-step approaches. However, it seems rather unusual in banking practice that the same characteristic affects the score positively in one group and negatively in another. Provided that there is such a characteristic in a real-world application, will it make a difference in a ten-or-more-characteristic scorecard?

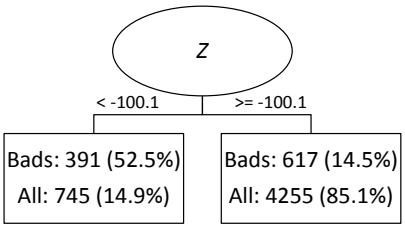

Figure 2.13. CART/CHAID tree for the artificial dataset

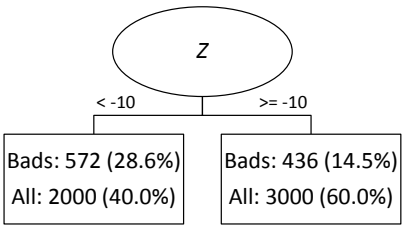

Figure 2.14. LOTUS tree for the artificial dataset

	Training sample	***Test sample***	***Validation sample***
	Artificial Dataset		
CART/CHAID	0.528	0.519	0.517
LOTUS	0.636	0.635	0.633
Logistic regression	0.482	0.479	0.469

Table 2.6. Gini coefficient values for training, test and validation samples (artificial dataset)

	Training sample	***Test sample***	***Validation sample***
	Artificial Dataset		
CART/CHAID	0.392	0.388	0.380
LOTUS	0.486	0.497	0.499
Logistic regression	0.335	0.344	0.330

Table 2.7. KS statistic values for training, test and validation samples (artificial dataset)

2.7 Conclusions

In this research, the LOTUS algorithm, in which both segmentation and scorecards are optimised at the same time, is compared to the two-step approaches, where logistic regression follows CART or CHAID trees. A logistic regression model serves as a reference for the multi-scorecard models. The above-mentioned methods have been applied to the data provided by two UK banks and a European credit bureau. The model performance measures have been calculated to assess an improvement due to the segmentation.

For none of the analysed real-world datasets do the multi-scorecard models perform considerably better than the logistic regression. Thus, the first and most important finding is that segmentation does not always improve model performance in credit scoring. The performance improvement is not necessary to occur even if it is the only goal of segmentation as in this research. This is in line with the findings of Banasik *et al*. (1996) that have been confirmed here for the statistical methods of segmentation.

Secondly, there is no difference in performance between the two-step and simultaneous approaches. Classification trees (CART and CHAID) followed by logistic regression in their leaves yield similar results to the LOTUS algorithm, in which both segmentation and scorecards are optimised at the same time. The LOTUS algorithm had seemed promising as a method for the optimal segmentation. However, it outperforms neither the two-step approaches nor the logistic regression.

Thirdly, for a large sample including strong characteristics, all the models have the same separating ability and are equally stable. In this case, the two-step and simultaneous approaches as well as the logistic regression perform very similarly. For smaller samples and/or weaker characteristics, the logistic regression models are the most stable, since they have fewer parameters and a simpler structure than the multi-scorecard models.

Fourthly, the segmentation contribution can be up to 20 percentage points. The discriminatory power measures of the trees which are used for segmentation can be even higher than those of the scorecards developed in their leaves. This means that

segmentation itself can be a very powerful tool. However, it seems that such a strong segmentation does not leave much space for the scorecards to further discriminate customers. Thus, the scorecards on average are weaker than the single-scorecard model.

Fifthly, it is possible to show an example of a situation when segmentation improves the model performance and the simultaneous approach outperforms the two-step approaches on an artificial dataset. Nevertheless, such a situation as in the example seems rather unusual in banking practice.

Building more than one scorecard requires more time and resources to be allocated to development, implementation, maintenance, monitoring and validation of the model. These additional costs should be compensated for by the improvement in performance, if it is the goal of segmentation. As this research shows, such an improvement is not necessary to occur. If it does not occur, it makes sense to use a single-scorecard model. Generally, it is advised to minimise the number of segments (Anderson, 2007).

In banking practice it is common not to compare the developed multi-scorecard model with a single-scorecard one. Building the latter is usually considered a waste of time, since there is a strong belief that segmentation allows for better risk assessment. Moreover, it seems that there is a pressure in the industry to choose multi-scorecard models, e.g. each new version may be expected to consist of more scorecards than the previous one. However, maintaining a number of scorecards which perform like a single one is a great waste of resources. In light of this research, it is strongly recommended to develop a single-scorecard model for comparison purposes.

Understandably, in practice segmentation is rarely chosen solely on the basis of the model performance. There are various criteria: identifiability, substantiality, accessibility, stability, responsiveness and actionability (see section 2.2.1) as well as operational reasons: maintainability, impact on backtesting and stress testing, properties of possible cut-offs etc. They all should be taken into account when deciding on segmentation.

As far as the model performance is concerned, usually the discriminatory power is an important but not the only criterion for the model choice. For example, if a multi-scorecard model is similar to a single-scorecard one in terms of the discriminatory power but produces more accurate PD estimates, then it makes sense to choose the former and not the latter. Therefore, further analysis could investigate the impact of segmentation on the model calibration.

Further analysis could also include comparing results obtained using LOTUS and other simultaneous approaches, e.g. Logistic Model Trees (LMT). Similarly to LOTUS, LMT are classification trees with logistic regression models in the leaves but they employ additive logistic regression models estimated using a boosting algorithm called LogitBoost (Landwehr *et al.*, 2005).

In the future, once the cross-border data exchange amongst the European credit bureaus emerges from its infancy stage, it will be interesting to find out how a performance-driven segmentation divides the customer population of the European banks. In particular, one may wonder whether there are any groups of countries with different relationships between a customer's characteristics and the dependent variable. At the moment such an analysis is not feasible, since there are large discrepancies in the scope of data collected by different credit bureaus (Association of Consumer Credit Information Suppliers and European Credit Research Institute, 2011) but the need for standardisation has already been recognised (European Commission's Expert Group on Credit Histories, 2009).

Finally, it should not be forgotten that segmentation is sometimes driven by other factors than the model fit. For example, it may result from marketing strategies or data availability, and the model performance improvement may not be its goal.

Chapter 3

Modelling LGD for unsecured retail loans using Bayesian methods[2]

3.1 Introduction

Loss Given Default (LGD) is defined in section 1.2.3. This chapter is on modelling LGD for unsecured retail loans. Because of the LGD distribution shape, it is often difficult to fit a model to the data. Therefore, multi-stage models were proposed, such as the two-step approach presented by Matuszyk *et al.* (2010). In this frequentist approach, two separate models are estimated independently, which can be considered problematic from the methodological point of view. The first model (logistic regression) separates positive values from zeroes, whereas the second model (e.g. linear regression) allows for the estimation of the positive values. The result is a point estimate of LGD for each loan. In order to apply this approach, one has either to set a cut-off for the first model or to calculate a product of the estimated value and probability that this value is greater than zero. One can also draw a number from a Bernoulli distribution with the estimated probability, whether to assign the value or zero, which is equivalent to using a random cut-off.

Alternatively, LGD can be modelled using Bayesian methods. The Bayesian framework offers a more coherent approach, since there is a single, hierarchical model instead of two separate ones. The result is an individual predictive distribution of LGD for each loan, rather than just a single number. Having a distribution, one can use its characteristics such as quantiles. The predictive distributions can be used, for example, in the LGD stress testing process or to approximate the downturn LGD. In this research, Bayesian methods as well as the frequentist approach are applied to the data on personal

[2] This chapter is based on the following paper: Bijak, K. and Thomas, L.C. (2013) Modelling LGD for unsecured retail loans using Bayesian methods, accepted for publication in the *Journal of the Operational Research Society*.

loans that were provided by a large UK bank. The data are such that the empirical distribution of LGD has a high peak at zero, which justifies the use of multi-stage approaches. With regard to Bayesian methods, they are argued to be an appropriate choice here, because they allow for an integrated estimation of hierarchical models.

This chapter is structured as follows. Section 3.2 is on the research background that covers various techniques of LGD modelling as well as a short introduction to Bayesian statistics and a review of Bayesian methods in credit risk modelling. In section 3.3, the frequentist and Bayesian approaches to LGD modelling are presented. In section 3.4, the data are described. In section 3.5, the empirical results are demonstrated. Section 3.6 is a discussion on the possible uses of the results, whereas section 3.7 includes the research findings and conclusions.

3.2 Background

3.2.1 LGD modelling for unsecured retail loans

LGD usually takes values from the interval [0,1]. It can exceed one, if a bank hardly manages to recover any of the loan and adds in its collection costs. LGD can also be negative, if the principal, interests, fees and penalties which have been paid sum up to more than the outstanding amount plus work-out costs. Some models cannot cope with values outside the interval [0,1]. Then such values need to be rejected, transformed or replaced with zeroes and ones. The LGD distribution often has a high peak at zero, since there are many customers who default but finally pay in full. This peak can be partly due to 'cures', i.e. defaulters who get back on track before the bank takes any action against them (Thomas, 2009a). There is sometimes another peak at one when many customers pay nothing. The spike at zero is typical for in-house collection, whereas the spike at one is typical for third party collection that normally deals with the debt which is harder to collect (Thomas *et al.*, 2012). In consequence, LGD is generally found difficult to model.

LGD is typically modelled for recovery periods that are longer than typical outcome periods in PD models. Under the IRB approach, the observation period for retail LGD must cover at least five years. LGD models for unsecured retail loans can be classified

as either one-stage or multi-stage approaches. As far as the former are concerned, a number of regression models were suggested: Ordinary Least Squares (OLS) regression (e.g. Querci, 2005; Bellotti and Crook, 2008 and 2012; Loterman *et al.*, 2009), Least Absolute Value (LAV) regression (Bellotti and Crook, 2008 and 2012), robust and ridge regression (Loterman *et al.*, 2009), beta regression (Loterman *et al.*, 2009; Arsova *et al.*, 2011) and fractional regression (Arsova *et al.*, 2011). Other one-stage models include tobit (Bellotti and Crook, 2008) and two-tailed tobit (Bellotti and Crook, 2012). Moreover, Zhang and Thomas (2012) used survival analysis, whereas Loterman *et al.* (2009) applied such techniques as CART, NNs, Multivariate Adaptive Regression Splines (MARS) and Least Squares Support Vector Machines (LSSVMs).

As far as the multi-stage approach is concerned, there are two and sometimes three stages, in which separate models are estimated. The first model usually discriminates positives from zeroes (and negatives, if any). In the two-stage approach, the second model allows for the estimation of the positive values. In the three-stage approach, the second model separates ones-or-greater from the rest, whereas the third model is built for the estimation of the remaining values, i.e. those from the interval (0,1).

In the first two stages, logistic regression and decision trees can serve as the discrimination models (e.g. Bellotti and Crook, 2008 and 2012; Matuszyk *et al.*, 2010; Zhang and Thomas, 2012). One can also combine two discrimination tasks into one using ordinal logistic regression (Arsova *et al.*, 2011). In the last stage, the following models were tried out: OLS (Bellotti and Crook, 2008 and 2012), LAV (Bellotti and Crook, 2008), robust, ridge and beta regression, CART, NNs, MARS and LSSVMs (Loterman *et al.*, 2009) as well as survival analysis (Zhang and Thomas, 2012). Another multi-stage approach was presented by Loterman *et al.* (2009): one can estimate a linear regression in the first stage and correct it using a non-linear model in the second stage. The nonlinear (e.g. CART, NN, MARS or LSSVM) model is applied to estimate the error of the linear regression.

It is not clear which LGD models are best. Linear regression is usually better than survival analysis (Zhang and Thomas, 2012), tobit models and simple decision trees (Bellotti and Crook, 2008), but it tends to be outperformed by nonlinear models such as

NNs and MARS (Loterman *et al.*, 2009). However, one should bear in mind that such findings may depend on the performance measures used. For example, in one research, OLS was better than LAV for MSE, while for MAE the opposite was true (Bellotti and Crook, 2008).

Apart from Mean Square Error (MSE) and Mean Absolute Error (MAE), the following performance measures are used for LGD models: Root Mean Square Error (RMSE), coefficient of determination (R-squared), Pearson's, Spearman's and Kendall's correlation coefficients as well as AUC and area over the Regression Error Characteristic (REC) curve (Loterman *et al.*, 2009). The correlation coefficients measure correlation between the observed and predicted LGD. Since the REC curve estimates the CDF of the squared or absolute residual, the area over the curve (AOC) estimates the expected regression error (Bi and Bennett, 2003). The AUC requires a binary variable such as the observed LGD classified into two groups, e.g. below-the-mean and over-the-mean. Thus, the AUC measures how well the model separates lower and higher values of LGD. However, Somers' D would be more suitable for this purpose, since it does not need any arbitrary classification of the dependent variable. In the multi-stage approach, the performance of each model should also be assessed separately, using appropriate measures, provided that the models are estimated independently. Regardless of the measure used, most LGD models perform rather weakly.

In order to improve model performance and/or produce a more normal-shaped distribution, some transformations of the original LGD are introduced. Since the beta distribution seems especially promising for such variables as LGD, a beta transformation is often applied (e.g. Loterman *et al.*, 2009; Matuszyk *et al.*, 2010). It was also used in the famous LossCalc model developed by Moody's KMV (Gupton and Stein, 2005). Other possible transformations include: log, fractional logit and probit (Bellotti and Crook, 2008) as well as the Box-Cox transformation (Loterman *et al.*, 2009; Matuszyk *et al.*, 2010). However, transformations do not necessarily lead to a better model performance. For example, Loterman *et al.* (2009) found that transformations do not improve the performance of OLS regression models.

The covariates of LGD models can be classified into five groups: socio-demographic variables (e.g. customer's age, residential status), customer's financial situation (e.g. income, number of credit cards, credit bureau score), account details (e.g. age, loan amount), payment history (e.g. outstanding balance, number of months with arrears within the last year or within the loan life) and macroeconomic variables such as interest rate or unemployment rate. A similar, yet not identical, classification was suggested by Bellotti and Crook (2008). Socio-demographic variables are normally collected at application. Information on the customer's financial situation can be updated on the basis of credit bureau reports. Account details should reflect the situation at default, and payment history should cover the period until default (provided that the model is developed using only data on defaulted loans). Macroeconomic variables can be collected at default or at an earlier date, since their impact on the customer's ability to pay may be delayed. Using macroeconomic variables is one way to assess the downturn LGD (Caselli *et al.*, 2008; Bellotti and Crook, 2012).

3.2.2 Bayesian statistics

3.2.2.1 Basics

So far, LGD modelling has been based on frequentist (classical) statistics, in which inference is made using sample data as the only source of information. Bayesian statistics, in turn, allows for the incorporation of other sources of information (e.g. expert knowledge). This extra knowledge is called the 'prior information', and is described with the prior probability distributions of the model parameters. The prior distributions are then updated using data, which yields the posterior distributions of the parameters, conditional on the observations. Providing a full distributional profile of the parameters is one of the advantages of Bayesian statistics. Other advantages include a coherent description of uncertainty in the model and direct interpretation of confidence ('credible') intervals. Bayesian statistics also enables an integrated estimation of complex and multilevel models (Lynch, 2007).

Since sample data and the prior information can to some extent compensate for each other, Bayesian methods can be successfully applied even if there is little data or no additional knowledge. In the former case, the extra knowledge plays a major role and

thus, the so-called informative priors are used. In the latter case, non-informative priors are chosen.

The relationship between the prior and posterior distributions of the model parameters $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$ can be described using Bayes' theorem:

$$p(\boldsymbol{\theta}|\boldsymbol{x}) = \frac{p(\boldsymbol{x}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\boldsymbol{x})} = \frac{\prod_{i=1}^{n} p(x_i|\boldsymbol{\theta})p(\boldsymbol{\theta})}{\int \prod_{i=1}^{n} p(x_i|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}}$$

where $p(\boldsymbol{\theta})$ denotes the prior probability density and $p(\boldsymbol{\theta}|\boldsymbol{x})$ denotes the posterior density, given the data $\boldsymbol{x} = (x_1, \dots, x_n)$ (Bernardo and Smith, 2003, p. 243). Besides Bernardo and Smith (2003), comprehensive publications on Bayesian statistics include ones by Congdon (2004) and Gelman *et al.* (2004).

3.2.2.2 MCMC methods

It is often difficult to derive the posterior distributions analytically. In order to generate samples from the posterior distributions, stochastic simulation methods are usually employed, with Markov chain Monte Carlo (MCMC) being the most popular ones. MCMC methods are based on the construction of a Markov chain that converges to the posterior distribution. Thus, let $\{\boldsymbol{\theta}^{(t)}\}$ be a Markov chain: $p\left(\boldsymbol{\theta}^{(t+1)}\middle|\boldsymbol{\theta}^{(t)}, \dots, \boldsymbol{\theta}^{(1)}\right) = p\left(\boldsymbol{\theta}^{(t+1)}\middle|\boldsymbol{\theta}^{(t)}\right)$. Under some assumptions, as $t \to \infty$, the distribution of $\boldsymbol{\theta}^{(t)}$ converges to its equilibrium that does not depend on the initial state of the Markov chain $\boldsymbol{\theta}^{(0)}$. This equilibrium is the posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{x})$ (Ntzoufras, 2009).

In general, an MCMC algorithm consists of the following steps (Ntzoufras, 2009):

1) Selection of the initial (starting) values $\boldsymbol{\theta}^{(0)}$;
2) Generating *T* values until the equilibrium is reached;
3) Convergence monitoring;
4) Discarding the first *B* values ('burn-in period');
5) Treating $\{\boldsymbol{\theta}^{(B+1)}, \dots, \boldsymbol{\theta}^{(T)}\}$ as the sample ('MCMC output');
6) Analysis of the posterior distributions: calculating posterior summary statistics, plotting densities etc.

The initial values can be randomly generated from the prior distributions. In case of informative priors, their means or modes can serve as the starting values. Some researchers use the maximum likelihood estimates. For certain problems, multiple Markov chains with different initial values are preferred. A number of methods for selecting the starting values, including those based on the simulated annealing algorithm, are mentioned by Brooks (1998). In order to eliminate the influence of the initial values on the final results, the first *B* iterations are assumed to constitute a burn-in period (such *B* is selected that the equilibrium is reached by the *B*th iteration). The values which are generated in this period are discarded and the remaining values are treated as the sample that is called the 'MCMC output'. Thus, the starting values should have no impact on the results but can still affect the speed of convergence (Brooks, 1998).

The MCMC output sample is not independent, since there are autocorrelations of lag l (l = 1, 2, 3, …), i.e. correlations between $\boldsymbol{\theta}^{(t)}$ and $\boldsymbol{\theta}^{(t+l)}$ that result from the Markov property of the chain. In consequence, the variances of the parameters are underestimated, just as standard errors are underestimated in a classical model in which observations are not independent (Lynch, 2007). A common solution to this problem is referred to as 'thinning the chain'. In this solution, one selects such a sampling lag (thinning interval) $L > 1$ that the autocorrelations of lag $l \geq L$ are low. Then one takes the first value from each sequence of L iterations to obtain an independent sample. Alternatively, one can calculate the means of the parameters for each sequence and treat them as the sample (Lynch, 2007).

The convergence monitoring can cover autocorrelations, ('trace') plots of the generated values, quantile and ergodic mean plots, the Monte Carlo (MC) errors as well as some statistical tests. An ergodic mean is a mean until the current iteration (Ntzoufras, 2009). The MC error is a measure of variability of the parameter estimate due to the simulation. It is typically estimated using either the batch mean method or the window estimator method (Ntzoufras, 2009). The batch mean method consists in dividing the MCMC output sample into a number of batches (e.g. 30 or 50) and calculating both the batch means and the sample mean. The MCMC error is computed as the standard deviation of the batch means (i.e. their variation from the sample mean). The window

estimator method is based on the variance formula for autocorrelated samples. The formula is limited to the autocorrelations from a certain window (i.e. those of lag $l \leq w$), since the autocorrelations of higher lags are low, have little impact on the variance and thus can be ignored. The MC error which is low in comparison to the posterior standard deviation of the parameter demonstrates that the posterior mean of this parameter has been estimated with high precision (Ntzoufras, 2009).

3.2.2.3 Metropolis-Hastings algorithm and Gibbs sampler

Two of the most popular MCMC methods are the Metropolis-Hastings algorithm and its special case, the Gibbs sampler. In each iteration $t = 1, \dots, T$, the Metropolis-Hastings algorithm runs as follows (Lynch, 2007, p. 108):

1) The candidate values of the parameters $\boldsymbol{\theta}^c$ are randomly drawn from a proposal distribution $q\left(\boldsymbol{\theta}^c \middle| \boldsymbol{\theta}^{(t-1)}\right)$;
2) The ratio R is calculated as:

$$R = \frac{p(\boldsymbol{\theta}^c|\boldsymbol{x})q\left(\boldsymbol{\theta}^{(t-1)}\middle|\boldsymbol{\theta}^c\right)}{p(\boldsymbol{\theta}^{(t-1)}|\boldsymbol{x})q(\boldsymbol{\theta}^c|\boldsymbol{\theta}^{(t-1)})} = \frac{p(\boldsymbol{x}|\boldsymbol{\theta}^c)p(\boldsymbol{\theta}^c)q\left(\boldsymbol{\theta}^{(t-1)}\middle|\boldsymbol{\theta}^c\right)}{p(\boldsymbol{x}|\boldsymbol{\theta}^{(t-1)})p(\boldsymbol{\theta}^{(t-1)})q(\boldsymbol{\theta}^c|\boldsymbol{\theta}^{(t-1)})}$$

3) A number u is randomly drawn from a standard uniform distribution $U(0,1)$;
4) The values of the parameters $\boldsymbol{\theta}^{(t)}$ are set as:

$$\boldsymbol{\theta}^{(t)} = \begin{cases} \boldsymbol{\theta}^c & \text{if } R > u \\ \boldsymbol{\theta}^{(t-1)} & \text{otherwise} \end{cases}$$

Comparing the ratio R to the randomly drawn number u is equivalent to updating the parameters, i.e. accepting the candidate values, with probability $\alpha = min(1, R)$ (Ntzoufras, 2009). The formula for R ensures that better candidates have a greater chance to be accepted. The candidates are generated from a proposal distribution. Theoretically, any distribution can serve as a proposal. Similarly to the initial values, the proposal distribution should have no effect on the final results but can influence the speed of convergence (Ntzoufras, 2009). Proposal densities can be classified as either symmetric or asymmetric, depending on whether $q\left(\boldsymbol{\theta}^c \middle| \boldsymbol{\theta}^{(t-1)}\right) = q\left(\boldsymbol{\theta}^{(t-1)} \middle| \boldsymbol{\theta}^c\right)$ or not.

In case of asymmetric proposals, some values might be selected as candidates more often than others, but R is adjusted for this (Lynch, 2007).

In the Gibbs sampler, the full conditional posterior distributions are used as proposals. In this algorithm, each iteration $t = 1, \dots, T$ consists of the following steps (Lynch, 2007, p. 89):

1) The parameter $\theta_1^{(t)}$ is randomly drawn from $p\left(\theta_1 \middle| \theta_2^{(t-1)}, \theta_3^{(t-1)}, \dots, \theta_k^{(t-1)}, \boldsymbol{x}\right)$;
2) The parameter $\theta_2^{(t)}$ is randomly drawn from $p\left(\theta_2 \middle| \theta_1^{(t)}, \theta_3^{(t-1)}, \dots, \theta_k^{(t-1)}, \boldsymbol{x}\right)$;
3) The parameter $\theta_3^{(t)}$ is randomly drawn from $p\left(\theta_3 \middle| \theta_1^{(t)}, \theta_2^{(t)}, \dots, \theta_k^{(t-1)}, \boldsymbol{x}\right)$;

…

j) The parameter $\theta_j^{(t)}$ is randomly drawn from
$p\left(\theta_j \middle| \theta_1^{(t)}, \dots, \theta_{j-1}^{(t)}, \theta_{j+1}^{(t-1)}, \dots, \theta_k^{(t-1)}, \boldsymbol{x}\right)$;

…

k) The parameter $\theta_k^{(t)}$ is randomly drawn from $p\left(\theta_k \middle| \theta_1^{(t)}, \theta_2^{(t)}, \dots, \theta_{k-1}^{(t)}, \boldsymbol{x}\right)$.

The parameters are updated in each iteration (the acceptance probability α is equal to one). Each parameter is sampled from the (univariate) conditional distribution given the current values of other parameters. This enables treating other parameters as fixed. As a result, a complex joint distribution is replaced with relatively simple univariate distributions, which makes the Gibbs sampler easy to apply and thus commonly used (Lynch, 2007). Among popular MCMC methods are also the random-walk Metropolis, componentwise Metropolis-Hastings and Metropolis within Gibbs algorithms as well as the independence sampler and the slice Gibbs sampler (Ntzoufras, 2009). Some other sampling methods are discussed by Brooks (1998).

3.2.3 Review of Bayesian methods in credit risk modelling

For at least 10 years, Bayesian methods have been successfully applied in credit risk modelling in general and in credit scoring in particular. A rich source of useful information on the Bayesian approach to financial risk management, including credit

risk management, is a book edited by Böcker (2010). This section provides a review of selected applications of Bayesian methods in credit risk modelling.

Since Bayesian statistics can effectively deal with data scarcity, it is found a useful tool for low default portfolios (LDPs). The advantage of this approach is that it allows for the formal incorporation of expert opinion, which is especially valuable when there is little data. Kiefer (2009) used Bayesian methods to estimate PD for an LDP of loans to highly rated, large, international banks. He used elicitation of prior distributions to quantify expert opinion on the unknown PD. A beta distribution was fitted to the assessments provided by the expert, and the method of moments was employed to estimate its parameters.

Fernandes and Rocha (2011) applied Bayesian and frequentist logistic regression, among other techniques, to estimate PD for a corporate LDP. In their research, the posterior means of the parameters which had been produced using the Bayesian approach were very similar to the estimates obtained using the frequentist approach. The authors performed the bootstrap analysis of the discriminatory power measures of both models. This is unconventional, since in Bayesian statistics, the information about parameter uncertainty is already embodied in the posterior distributions and thus, no additional bootstrapping is needed. Nevertheless, the generated distributions of the Gini coefficient were alike in both approaches. What is surprising is that the distributions of the KS statistic are different despite almost the same results for the parameter estimates. The observed default rates generally followed the desired, increasing trend in the successive ranges of the Bayesian score.

Mira and Tenconi (2003) also built both Bayesian and classical logistic regression models. They developed a Bayesian hierarchical logistic regression to assess PD of companies from different sectors. In most sectors there were few defaults and in one sector there were no defaults at all. In order to improve the MCMC method performance, the Metropolis-Hastings algorithm with delayed rejection was applied. The modification consists in delaying the update of the parameters in case of the candidate rejection: if the first candidate is rejected, the second one is proposed. In that application, the Bayesian approach outperformed the classical model.

Dwyer (2007) used Bayesian methods to assess PD for a corporate LDP. These methods allow for the determination of the posterior distribution of PD even if no defaults are observed at all. Another Bayesian approach presented in that paper produces the posterior distribution of the aggregate shock in the macroeconomic environment, given the observed default rate. Comparing this distribution with the actual stage of the cycle can help validate the PD model. This shows one way in which Bayesian statistics enable taking into account the macroeconomic conditions in the model validation process.

Another way is to employ Bayesian methods in the stress testing process. Park *et al.* (2010) applied a dynamic hierarchical Bayesian model to estimate PD of large companies. They used the small area estimation method to deal with missing data. They also added latent variables to allow for correlations amongst customers and time-series correlations in the model. Moreover, they proposed a stress testing methodology, in which the coefficients are stressed instead of the corresponding financial variables used in the model. As an example, the authors chose the 75th percentiles of the posterior distributions of the coefficients.

Bayesian statistics allows for the incorporation of expert knowledge or some extra information into a model. It also offers a way to update an old model with new data that are not sufficient to build a new tool. Ziemba (2005) showed how to update a generic scorecard with new information using Bayesian methods. In the presented case study, the existing logistic regression model was enriched with additional variables that had been collected after the introduction of a new application procedure. The prior information came from the old scorecard. The updated model outperformed both the old scorecard and the model that was developed using only new data, especially when the amount of new data was scarce.

Along similar lines, Konstantinos *et al.* (2003) applied Bayesian methods in risk-based pricing (RBP). Since many banks offer new credit card holders special conditions for the initial period, additional data on payments and card usage can be collected in this period. In the Bayesian framework, this extra information was used to revise such parameters as the probability of non-payment and the estimate of unpaid balance,

which, in turn, were used to update the risk-based APR. A scoring model employed at application was a source of the prior information.

Finally, Bayesian methods can be applied as an alternative to the frequentist ones. Giudici (2001) used Bayesian discrete graphical models and Bayesian model selection methods to investigate links between variables and find the best scorecard ('Bayesian data mining'). The graphs described conditional independencies. Bayesian model averaging was employed to estimate probability that the data support the existence of links between variables, regardless of the model selected. The developed Bayesian model was more parsimonious than the frequentist one built using the backward selection of variables.

Miguéis *et al*. (2012) modelled PD using binary quantile regression in the Bayesian framework. They argued that – while such methods as logistic regression focus on the relationship between regressors and an average value of the dependent variable – the extreme quantiles of the dependent variable distribution may sometimes be more important (indeed, they play a major role e.g. in stress testing, which is not mentioned in that paper). Uncertainty related to the estimation of PD for a given applicant was measured as the difference between the 0.95th and the 0.05th quantiles. Subsequently, a matrix of PD and the associated uncertainty was produced and a segmentation of applicants was suggested on its basis.

Chen and Åstebro (2003) suggested a Bayesian reject inference technique based on the Bound and Collapse (BC) method under the assumption that the data are Missing Not at Random (MNAR). The BC method produces conditional probabilities from incomplete data by bounding the intervals for parameter estimates using the available information, and collapsing the bounds to point estimates for missing values. In the presented example, using the proposed reject inference technique improved the discriminatory power of a scoring model under the MNAR assumption. That application can be viewed as yet another example of employing Bayesian methods when there is some lack of data (in this case, the lack of data on performance of the rejected applicants).

3.3 Methodology

3.3.1 Frequentist approach

In this research, Bayesian methods are compared and contrasted with the frequentist approach. The latter is similar to the two-step approach presented by Matuszyk *et al.* (2010). Let y_i denote LGD of the ith loan ($i = 1, \ldots, N$). The first of the two models separates positives from zeroes and negatives. It takes the form of a logistic regression (see section 2.3.1):

$$P(y_i > 0) = \frac{1}{1 + e^{-\boldsymbol{\beta}_1 \boldsymbol{x}_i}}$$

where $\boldsymbol{\beta}_1$ are the parameters and $\boldsymbol{x}_i$ are the covariates. The second model allows for the estimation of the positive values. It is a linear regression with parameters $\boldsymbol{\beta}_2$ and covariates $\boldsymbol{z}_i$:

$$E(y_i | y_i > 0) = \boldsymbol{\beta}_2 \boldsymbol{z}_i$$

The logistic regression predicts, whether there will be a (positive) loss or not. Here, its result will be referred to as the 'probability of loss'. The linear model yields the estimated LGD, provided that there is a loss. In this application, the estimation has been performed using SAS. The models have been developed on the training sample and tested on the validation sample. Based on the findings of Loterman *et al.* (2009), no transformations have been applied to the original LGD. The covariates of both regressions have been chosen using the stepwise selection (they have been selected because of their statistically significant relationship with the dependent variables and not because of their role in the recovery process).

There are two problems inherent in this approach. Firstly, the two models are estimated independently, although the use of the second model is conditional on the outcome of the first one. In this situation, their independent estimation can be considered problematic from the methodological point of view: the approach is incoherent in terms of handling uncertainty. Since there is no joint probability framework, uncertainty is not

propagated from the first to the second model and then into the output. Thus, a part of uncertainty about the LGD estimates is ignored. In particular, this may lead to confidence intervals that are too narrow and give a false impression of accuracy.

Secondly, it is not clear how to use the frequentist approach, once the models have been built, i.e. which value should be taken as the predicted LGD for a given loan. One option is to set a cut-off for the first model. Then zero is taken, if the probability of loss is less than the cut-off, and the estimated LGD is taken otherwise. Here, this will be called the 'cut-off approach'. It raises another question, though, which is how to set the cut-off. One idea is to choose the percentage of loans with positive values of LGD in the training sample.

Alternatively, it is possible to randomly decide, whether there will be a loss or not. One can draw a number from a Bernoulli distribution with parameter equal to the probability of loss. If the result is zero, zero is taken as the predicted LGD. If the result is one, the estimated LGD is taken. Equivalently, one can draw a cut-off from a standard uniform distribution $U(0,1)$ for each loan separately and compare the probability of loss to it ('random cut-off approach').

Yet another option is to calculate the predicted LGD as a product of the probability of loss and the estimated LGD. This product can be viewed as a mean of the discrete distribution, in which a random variable takes a value of the estimated LGD with the probability of loss, and zero with the complement probability. This will be referred to as the 'probability times value approach'. Regardless of the approach chosen, the result is a point estimate of LGD for each loan. Instead, one can take the above-mentioned simple distribution with only two possible values.

3.3.2 Bayesian approach

In this research, Bayesian methods have been used, since they allow for an integrated estimation of hierarchical models. In consequence, the Bayesian approach is free from the problems that are discussed in the previous section. In this approach, there is a single, hierarchical model instead of two separate ones. The structure of the model, which resembles the random cut-off approach, is illustrated in Figure 3.1. As shown in

this figure (graph), both variables $\boldsymbol{x}_i$ and parameters $\boldsymbol{\beta}_1$ affect p_i, which influences b_i, which, in turn, along with variables $\boldsymbol{z}_i$ and parameters $\boldsymbol{\beta}_2$, τ_1 and τ_2, has an effect on the dependent variable y_i, where $i = 1, \ldots, N$. Implementing the same hierarchical structure, including the same covariates, in both the Bayesian and frequentist approaches makes these approaches comparable. It must be stressed that the aim of this research is not to estimate the frequentist models using Bayesian methods but to see them as a Bayesian hierarchical model that is free from their drawbacks. Thus, the exact correspondence between the Bayesian and frequentist approaches is not of interest here.

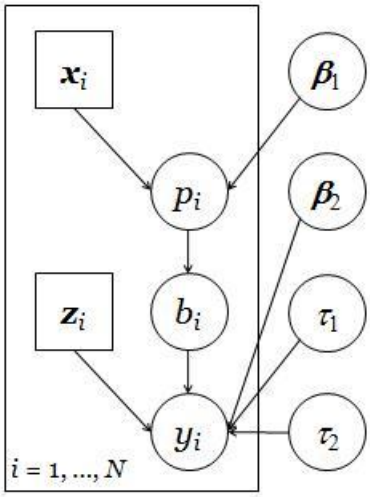

Figure 3.1. Bayesian hierarchical model

Obviously, it is possible to implement the two-step approach in the Bayesian framework as two separate models. However, there would be the same problems as in the frequentist framework: the models would be estimated independently, uncertainty would not be propagated properly, and it would be unclear how to use that approach to make predictions. This would not allow for the full utilisation of the advantages offered by Bayesian statistics.

Contrary to Bayesian methods, the frequentist framework does not enable performing an integrated estimation of the models in a straightforward way. Thus, in this framework it would not be so easy to develop a single hierarchical model parallel to the proposed Bayesian approach. Naturally, it is possible to build one-step models as mentioned in section 3.2.1. Because of the LGD distribution shape, it may be difficult to fit such models to the data, though (this motivated the development of the two-step models,

which, in turn, have disadvantages that are difficult to overcome in the frequentist framework). If Bayesian methods were also applied to estimate the one-step models, they would share the drawback of poor fit. The integrated estimation, which is their important advantage, would not be needed any more. Nevertheless, they would still produce individual predictive distributions of LGD.

The proposed Bayesian model works as follows. For each loan from the training sample, the probability of loss p_i is calculated using the logistic regression formula with parameters $\boldsymbol{\beta}_1$ and variables $\boldsymbol{x}_i$. Subsequently, a number b_i is drawn from a Bernoulli distribution with parameter p_i. If b_i equals zero, then y_i follows a normal distribution with zero mean and precision τ_1 (the precision, which is the reciprocal of the variance, is commonly used in Bayesian statistics). If b_i equals one, then y_i follows a normal distribution with mean computed using the linear regression formula with parameters $\boldsymbol{\beta}_2$ and variables $\boldsymbol{z}_i$, and precision τ_2. Then the observed value of y_i is used to update the parameters $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$, τ_1 and τ_2. This is the only place where it is fed into the model. The upper part of the model is not provided with additional information, whether there was a loss or not.

For each loan from the validation sample, the same operations are performed as described above, except for disclosing the observed value of y_i and updating the parameters. As a result, for each loan there is an individual predictive distribution of LGD that is a mixture of the two normal distributions mentioned above: $N(0, \tau_1^{-1})$ and $N(\boldsymbol{\beta}_2 \boldsymbol{z}_i, \tau_2^{-1})$. The resulting predictive distributions are bimodal. The adopted approach is similar to the non-Bayesian model suggested by Hlawatsch and Ostrowski (2011), who employed a mixture of two beta distributions to account for the bimodality of LGD for corporate loans. Hlawatsch and Ostrowski (2011) assumed that the first beta distribution is right-skewed, whereas the second one is left-skewed, and estimated the parameters of the mixture distribution using the Expectation Maximisation (EM) algorithm.

As far as the prior distributions are concerned, weakly informative priors are adopted for the parameters $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$, whereas slightly more informative priors are used for τ_1 and τ_2 to take into account the knowledge of the shape of the empirical LGD

distribution. Informative priors are not necessary in this application, since there is a large training sample. For each element of $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$, the prior is a normal distribution with zero mean and small precision (large variance): $N(0, 100^2)$ for intercepts and $N(0, 10^2)$ for others. The parameter τ_1 is assumed to follow a gamma distribution with shape parameter 10 and inverse scale parameter 0.00001. Thus, τ_1 has a very large expected value (10^6) and an even larger variance (10^{11}). In the model, τ_1 serves as precision of the normal distribution with zero mean, so the larger the τ_1, the smaller the variance of this distribution. This is designed to model the peak of the LGD distribution at zero.

The parameter τ_2 is assumed to follow a gamma distribution with parameters 0.01 and 0.01. Hence, the expected value of τ_2 is one and its variance equals 100, which gives relatively small precision (large variance) of the normal distribution with mean based on the linear regression formula. This aims to model the rest of the LGD distribution. The initial values of all model parameters are set to be equal to the expected values of their prior distributions.

The model has been fitted using OpenBUGS. OpenBUGS is a popular, programming language based software for performing Bayesian inference (Lunn *et al*., 2009). OpenBUGS generates samples from the posterior distributions using MCMC methods based on the Gibbs sampler. The code which has been developed for this research is demonstrated in Appendix B. In this application, the first 10000 iterations have been discarded as the burn-in period, and the next 100000 iterations have provided the MCMC output. Since relatively high autocorrelations up to lag four have been observed, a sampling lag $L = 5$ has been used to obtain an independent sample.

3.4 Data

The methods presented above have been applied to the data on personal loans that were granted by a large UK bank between 1987 and 1998 and defaulted between 1988 and 1999 (see Table 3.1). The data cover the recovery periods until 2004, when some loans were still being paid. The original loan amounts started from £500, whereas the loan terms varied from 12 to 60 months. There have been ca. 50000 records in the dataset. After the removal of outliers and missing values of LGD, a total of ca. 48000 records

have remained. Subsequently, the training and validation samples of 10000 loans each have been randomly selected from the dataset. Since the period covered by the data is long enough to include the whole economic cycle, 'out of time' validation does not seem necessary here.

Characteristics	***Values***
Original dataset size	49943
Dataset size w/o outliers and missing values	47853
Training sample size	10000
Validation sample size	10000
Loan open dates	1987-1998
Default dates	1988-1999
Recovery periods	Until 2004
Loan amounts at opening (in £)	500-16000
Loan terms (in months)	12-60
LGD	–0.04-1.23

Table 3.1. Data characteristics

The empirical distribution of LGD is demonstrated in Figure 3.2. Since ca. 30% of the loans were paid in full, it has a high peak at zero. There is no information on which customers were 'cures'. Less than 10% of the loans were not repaid at all. There are many cases of LGD greater than one and few cases of LGD less than zero. They have been kept unchanged, since the models which are used in this application can cope with such values. The mean and median are equal to 0.5 and 0.59, respectively. The standard deviation equals 0.39.

In the dataset, there are variables from four out of five groups mentioned in section 3.2.1. Macroeconomic variables have not been used here. Socio-demographic variables have been collected at application. Some account details reflect the situation at opening and some at default. The payment histories cover the period until default. Thus, the life of the loan means the time from opening to default, whereas the last 12 months mean the last year before default etc. The variables have been standardised.

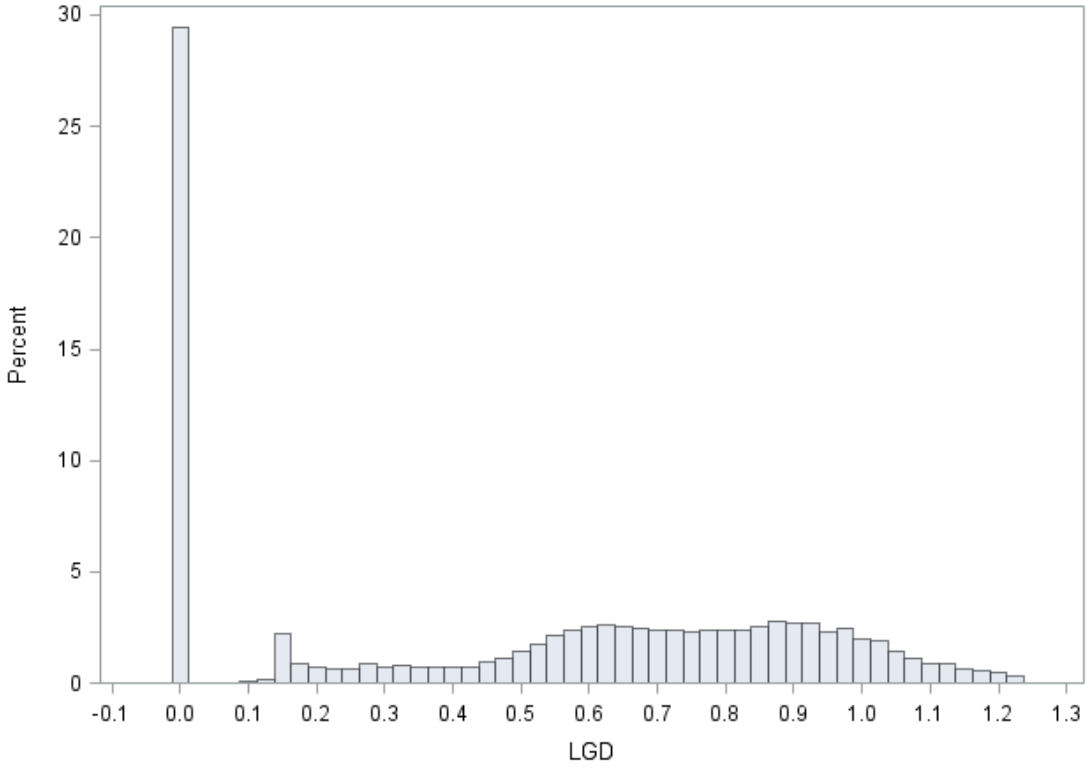

Figure 3.2. Empirical distribution of LGD

3.5 Results

3.5.1 Model convergence and performance

In the frequentist approach, the quality of each of the two models has been assessed separately before measuring the performance of the entire LGD model. The logistic regression discriminatory power has been measured with the Gini coefficient and the KS statistic, whereas the linear regression goodness of fit has been assessed using the R-squared. In the training sample, the Gini coefficient and the KS statistic equal 0.42 and 0.31, respectively. Almost the same values of these measures have been obtained on the validation sample, which means that the discriminatory power of the first model is good and stable. The R-squared of the linear regression is equal to 0.16 on both the training and validation samples. Thus, the goodness of fit of the second model is rather poor but stable. This is in line with the findings of Matuszyk *et al.* (2010).

In the Bayesian approach, monitoring of the MCMC algorithm convergence has been based on autocorrelations, quantiles and trace plots of the generated values as well as the MC errors. The autocorrelations are low due to the use of a sampling lag. In the successive iterations, the quantiles and generated values of each parameter have been remaining within their zones with no visible tendencies, which demonstrates that the algorithm has converged. The MC errors are relatively low, since they do not exceed 1.6% of the posterior standard deviations of the parameters (see Table 3.2). This shows that the posterior means of the parameters have been estimated with high precision.

Parameter	***Frequentist***	***Bayesian***			
	Estimate	***Posterior mean***	***Posterior std. dev.***	***MC error***	***MC%***
$\boldsymbol{\beta}_1$					
Intercept	1.084	1.087	0.026	$1.19 \cdot 10^{-4}$	0.45
Age of exposure (months)	–0.545	–0.545	0.062	$8.93 \cdot 10^{-4}$	1.45
Amount of loan at opening	0.338	0.339	0.025	$9.93 \cdot 10^{-5}$	0.39
Total number of advances/ arrears within the whole life of the loan	–1.478	–1.481	0.062	$5.25 \cdot 10^{-4}$	0.84
Number of months with arrears >0 within the life of the loan	0.073	0.076	0.078	$1.23 \cdot 10^{-3}$	1.57
Number of months with arrears >1 within the last 12 months	–0.529	–0.531	0.040	$3.08 \cdot 10^{-4}$	0.76
$\boldsymbol{\beta}_2$					
Intercept	0.719	0.718	0.003	$9.14 \cdot 10^{-6}$	0.32
Joint applicant present	–0.012	–0.012	0.003	$8.53 \cdot 10^{-6}$	0.29
Total number of advances/ arrears within the whole life of the loan	–0.143	–0.146	0.015	$1.89 \cdot 10^{-4}$	1.23

Parameter	***Frequentist***	***Bayesian***			
	Estimate	***Posterior mean***	***Posterior std. dev.***	***MC error***	***MC%***
Term of loan (months)	–0.037	–0.037	0.003	$1.01 \cdot 10^{-5}$	0.32
Worst arrears within the life of the loan	0.178	0.180	0.016	$1.91 \cdot 10^{-4}$	1.22
Number of months with arrears >2 within the last 12 months	–0.053	–0.053	0.004	$1.36 \cdot 10^{-5}$	0.31
τ_1	–	$1.46 \cdot 10^{8}$	$3.83 \cdot 10^{6}$	$1.26 \cdot 10^{4}$	0.33
τ_2	–	17.580	0.294	$9.37 \cdot 10^{-4}$	0.32

Table 3.2. Estimation results

In both the frequentist and Bayesian approaches, the LGD model performance has been measured with MSE and MAE as well as Pearson's, Spearman's and Kendall's correlation coefficients. All these measures have already been used for LGD models (see section 3.2.1) and it seems that none of them are considerably better or more important than the other. As mentioned earlier, it is not clear how to use the frequentist LGD model. Therefore, its performance has been assessed using three approaches (cut-off, random cut-off and probability times value).

In the cut-off approach, the performance measures of the entire LGD model have been calculated for a number of cut-offs set for the logistic regression. Figure 3.3 shows that the model performance strongly depends on the cut-off level. The higher the cut-off, the more loans are assigned zeroes and the fewer loans are assigned the estimated LGD values. The model performs best for the cut-offs around 0.5 but not higher than 0.7 (the percentage of loans with positive values of LGD in the sample). The optimal cut-offs vary amongst the performance measures used.

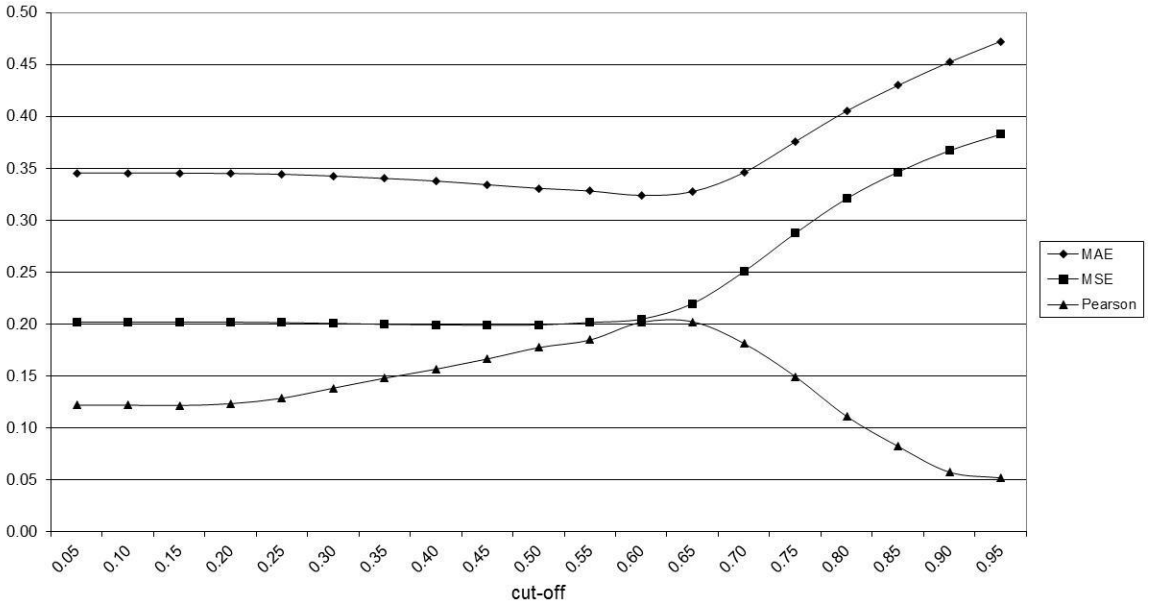

Figure 3.3. Performance of the frequentist LGD model
(cut-off approach, validation sample)

The random cut-off approach has been implemented in the Bayesian framework. Thus, there are the posterior distributions of the performance measures applied. The posterior means and standard deviations are presented in Table 3.3. The results of the frequentist random cut-off approach vary from one use to another, since there is random drawing involved. Therefore, the bootstrap has been performed in order to produce the distributions of MSE, MAE and the correlation coefficients. In the bootstrap algorithm, a new sample of the same size has been chosen from the original one, using sampling with replacement (i.e. with repetition allowed). The frequentist random cut-off approach has been applied to the new sample, and then the performance measures have been calculated. This has been repeated 10000 times on the training and validation samples separately. The bootstrap estimates of means and standard deviations of the performance measures are almost identical as those produced in the Bayesian approach (the differences are only in the fourth decimal place). The model performance is stable. Similar values of the errors and the correlation coefficients were obtained on some datasets by Loterman *et al.* (2009).

Performance measure	*Training sample*		*Validation sample*	
	Mean	*Std. dev.*	*Mean*	*Std. dev.*
MSE	0.244	0.003	0.245	0.003
MAE	0.364	0.003	0.365	0.003
Pearson's correlation	0.081	0.010	0.085	0.010
Spearman's correlation	0.107	0.010	0.115	0.010
Kendall's correlation	0.084	0.007	0.090	0.007

Table 3.3. Model performance measures (random cut-off approach)

In addition, the probability times value approach has been applied. It has also been implemented in the Bayesian framework, where it produces the predictive distributions of LGD calculated as LGD* = $p_i \boldsymbol{\beta}_2 \boldsymbol{z}_i$. The values of the performance measures which have been calculated in the frequentist probability times value approach are almost exactly the same as the corresponding posterior means presented in Table 3.4. Again, the differences are in the fourth decimal place. The posterior standard deviations of the performance measures are not shown here since they are very small. The results are stable and slightly better than those yielded in the random cut-off approach. The individual predictive distributions of LGD* are unimodal and extremely concentrated.

Performance measure	*Training sample*	*Validation sample*
MSE	0.142	0.143
MAE	0.328	0.329
Pearson's correlation	0.256	0.268
Spearman's correlation	0.241	0.255
Kendall's correlation	0.169	0.179

Table 3.4. Model performance measures (probability times value approach)

3.5.2 Parameter estimates

As expected, the posterior means of the parameters which have been produced in the Bayesian framework are very similar to the estimates obtained in the frequentist approach (see Table 3.2). The similarity of the posterior means and the corresponding frequentist estimates was also observed e.g. by Fernandes and Rocha (2011). These

similarities are likely to result from the large sample sizes (e.g. Courgeau (2012) noted that as the number of cases increases, the Bayesian and frequentist estimates converge to each other). The observed similarities may also be related to using non-informative (as in Fernandes and Rocha, 2011) or weakly informative priors (as in this application): when informative priors are not used, data practically remain the only source of information for inference, as in frequentist statistics.

In this research, the following interpretation of the posterior means (or the frequentist estimates) of the parameters $\boldsymbol{\beta}_1$ is suggested. The newer the exposure and the larger the loan amount, the higher is the probability that there will be a (positive) loss. However, the larger the number of arrears within the loan life and the larger the number of months with arrears >1 within the last year, the lower is the probability that there will be a loss. Matuszyk *et al.* (2010) explained similarly surprising findings using the metaphor of 'falling off a cliff'. The customers who tend to be in arrears ('to keep their heads above water') are more likely to succeed than those who have no delinquencies prior to default ('going underwater'). The explanation is that the latter default because of some sudden changes in their lives ('falling off a cliff') which may affect their ability to pay forever.

The posterior means (or the frequentist estimates) of the parameters $\boldsymbol{\beta}_2$ can be interpreted as follows. The longer the term of a loan, the lower is the LGD. The presence of a joint applicant has a negative impact on LGD. Moreover, the larger the number of arrears within the loan life and the larger the number of months with arrears >2 within the last year, the lower is the LGD. The posterior means of τ_1 and τ_2 are larger than their prior means. Thus, the variances of the normal distributions are smaller than initially assumed. This is especially true of the distribution that is designed to model the peak at zero.

The posterior distributions of the model parameters are presented in Figures 3.4-3.9. Figure 3.4 illustrates that the most accurately estimated element of $\boldsymbol{\beta}_1$ is the second parameter (the one for the amount of loan at opening). Figure 3.7 demonstrates that the first, third and fifth elements of $\boldsymbol{\beta}_2$ (for the joint applicant present, the term of loan and the number of months with arrears >2 within the last 12 months) are considerably more accurately estimated than the second and fourth parameters. Figures 3.6 and 3.9 show

how the distributions of τ_1 and τ_2 have changed from priors to posteriors, which results in higher precision of both normal distributions whose mixture is the predictive distribution of LGD. The prior distributions of other model parameters have not been plotted here, since they are less informative and thus not worth presenting.

3.5.3 Predictive distributions of LGD

In the Bayesian approach, there is an individual predictive distribution of LGD for each loan, rather than just a point estimate as in the frequentist approach. Examples of such distributions for three selected loans from the validation sample are shown in Figures 3.10-3.12. Each of them is a mixture of two normal distributions that are mixed in various proportions. Thus, the predictive distributions are bimodal. In fact, they have much narrower peaks at zero, but a smoothing method (kernel density estimation with a Gaussian kernel) has been used here for visualisation purposes. The dashed lines mark the observed values of LGD.

Having the predictive distributions, one can use their characteristics such as means and quantiles. If the predictive mean of LGD is treated as a point estimate for each loan, then the performance measures take the same values as presented in Table 3.4. Using the predictive median or other quantiles instead of the mean does not considerably improve the model performance. For the median, only MAE is slightly lower than for the mean, with values of 0.316 and 0.319 on the training and validation samples, respectively.

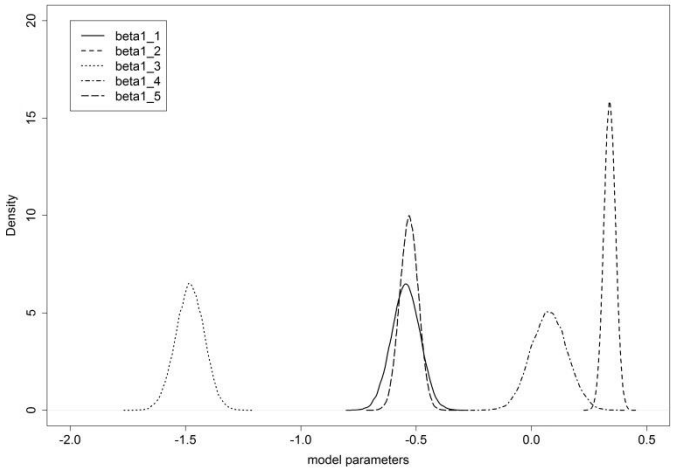

Figure 3.4. Posterior distributions of the parameters $\boldsymbol{\beta}_1$ (without the intercept)

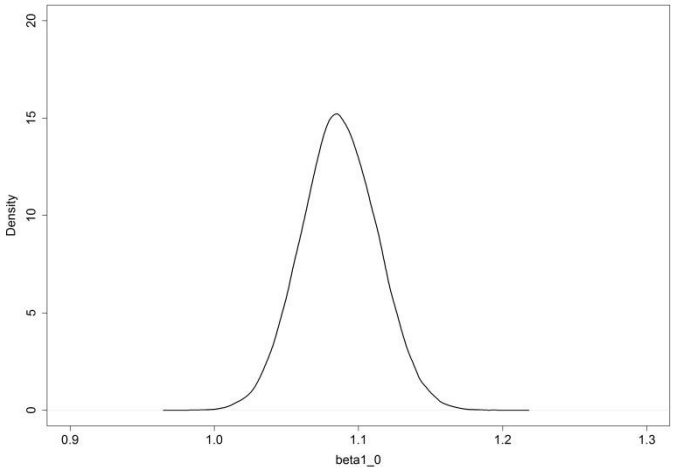

Figure 3.5. Posterior distribution of the intercept of $\boldsymbol{\beta}_1$

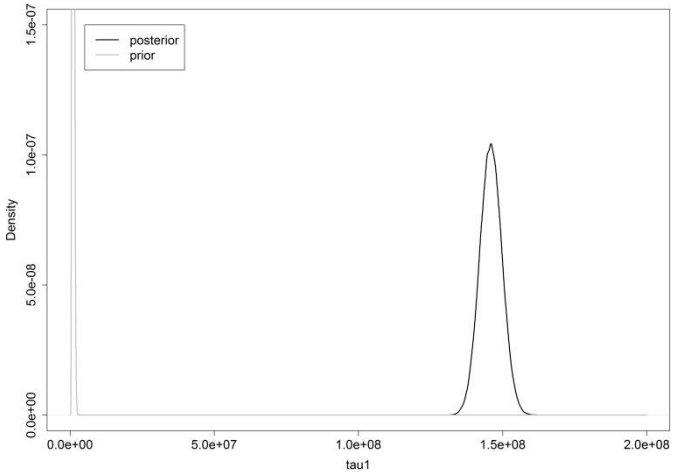

Figure 3.6. Posterior and prior distributions of τ_1

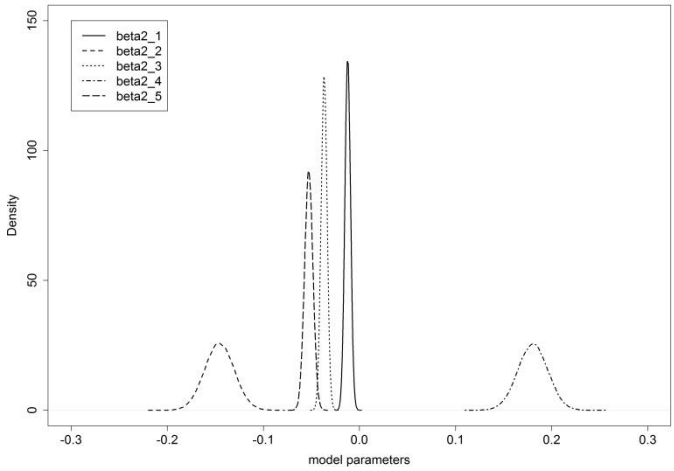

Figure 3.7. Posterior distributions of the parameters $\boldsymbol{\beta}_2$ (without the intercept)

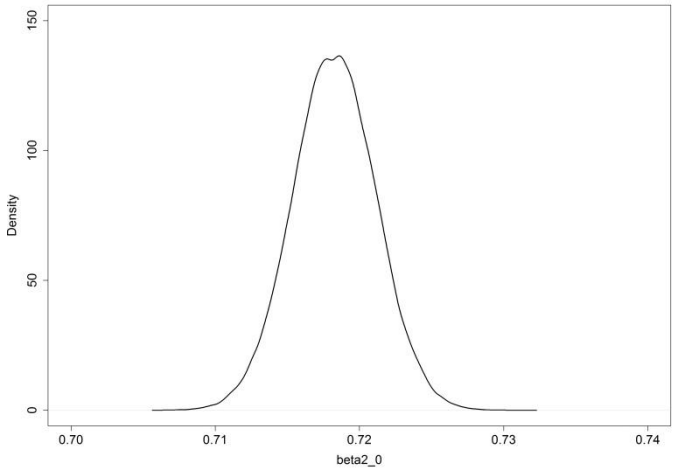

Figure 3.8. Posterior distribution of the intercept of $\boldsymbol{\beta}_2$

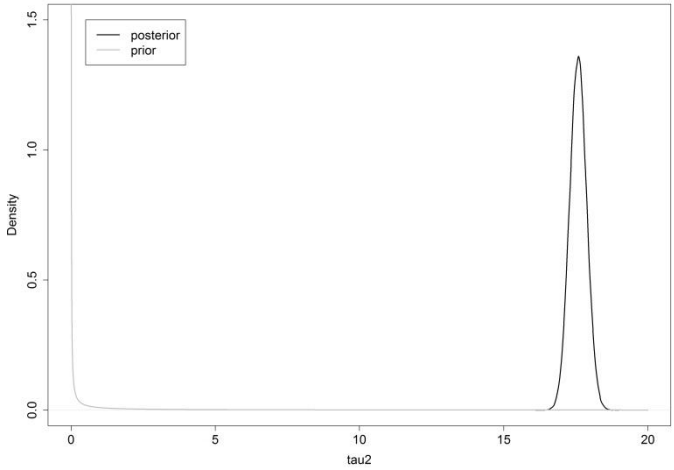

Figure 3.9. Posterior and prior distributions of τ_2

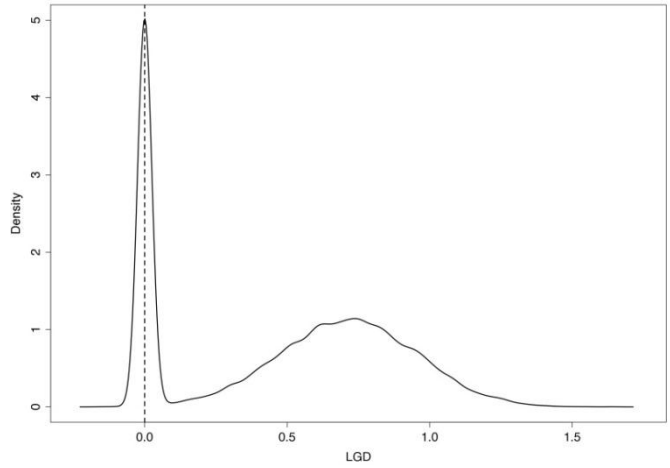

Figure 3.10. Predictive distribution of LGD for the loan (1)

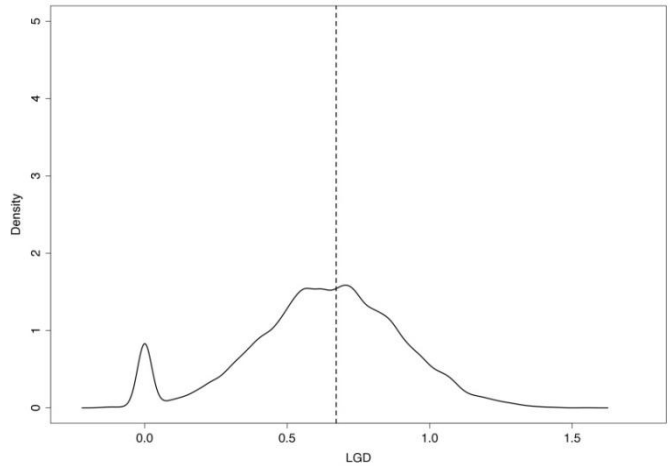

Figure 3.11. Predictive distribution of LGD for the loan (2)

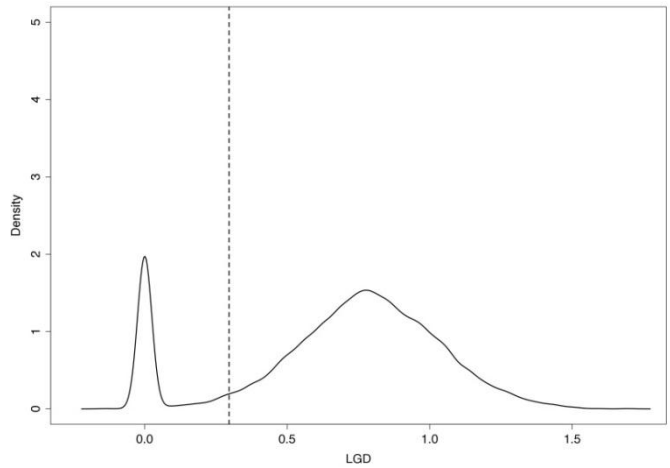

Figure 3.12. Predictive distribution of LGD for the loan (3)

3.6 Discussion

The individual predictive distributions provide much more information and offer more possibilities than the point estimates of LGD. This section suggests how they could be used in banking practice.

Kim (2006) proposed using theoretical distributions to produce various LGD estimates, including the downturn LGD, for corporate exposures in a non-Bayesian framework. In the Bayesian framework, one could approximate the downturn LGD with a certain quantile of the predictive distribution for each loan. The posterior distributions of the parameters reflect all reasonable sources of uncertainty in a Bayesian model (Gelman *et al.*, 2004); what is usually not reflected is the model uncertainty, although there are some Bayesian methods that allow dealing with it (e.g. Draper, 1995). Thus, all reasonable sources of uncertainty are handled and – explicitly or implicitly – incorporated in the model, including uncertainty arising from inability to capture each and every influence on the dependent variable in the model (e.g. uncertainty related to such omitted factors as the changing macroeconomic conditions or systematic risk). Kim (2006) defined the economic downturn as "the state that the systematic risk factor takes on value at the 99.9% quantile". From the equivariance of quantiles under monotonic transformations (e.g. Hao and Naiman, 2007), it follows that if LGD is assumed to be a monotonic function of the systematic risk factor, then the selected quantile of the LGD distribution will correspond to the quantile of the same order of the underlying systematic risk factor distribution. Hence, e.g. the 0.999th quantiles will reflect both the downturn conditions and the downturn LGD. According to Kim (2006), the choice of the quantile depends on the user's perception of the severity of downturns and the 0.999th quantile can be used for extremely severe downturns. In the validation sample, choosing the 0.9th and 0.95th quantiles results in the average predicted downturn LGD of 0.97 and 1.06, respectively (while the average observed LGD of these loans was equal to 0.5 in the changing economic conditions of over a decade). Choosing the 0.75th quantile leads to the average predicted downturn LGD of 0.8, which means that such a quantile may reflect moderate downturn conditions.

According to the Basel II document, the banks which use the AIRB approach "must estimate an LGD for each facility that aims to reflect economic downturn conditions

where necessary to capture the relevant risks. […] For this purpose, banks may use averages of loss severities observed during periods of high credit losses, forecasts based on appropriately conservative assumptions, or other similar methods" (BCBS, 2006, paragraph 468). The approximation of the downturn LGD suggested in this section could be classified as a 'forecast based on appropriately conservative assumptions'. It could be useful when downturn data are lacking, which is often the case in banking practice. Otherwise, one should opt for methods based on historical downturn data.

In addition, selected quantiles of the predictive distributions can be used as the stressed LGD. One can also apply the methodology proposed by Park *et al.* (2010), who stressed the coefficients instead of the corresponding financial variables in the PD model where PD was a symmetric function of the variables and their coefficients. They used the 75th percentiles of the posterior distributions of the coefficients as reflecting a stress situation. Within the approach suggested in this research, one can stress the model parameters instead of such variables as the number of months with arrears >2 within the last 12 months. Then the appropriate quantiles of the posterior distributions of these parameters can be used to generate the stressed LGD.

Moreover, the predictive distributions of LGD can be a useful tool in the collection process. For example, a bank may wish to identify and try to recover only those loans that are likely to be paid at least partially, if not in full. Based on the predictive distributions, the bank can select the loans, for which 90% credible intervals do not include one: $P(LGD < 1) \geq 0.9$. In this application, such loans make up ca. 60% of the validation sample (in fact, 96% of them were paid at least partially). Another bank may be able to try to recover e.g. only 25% of the defaulted loans. The bank can order the loans by $P(LGD < 1)$ and take actions against the one-fourth with the highest probabilities. Yet another bank refrains from punitive actions once half of the debt has been recovered. Thus, that bank may wish to know which loans are likely to be paid in more than 50%, e.g. $P(LGD < 0.5) \geq 0.9$. Generally, the predictive distributions can be used to diversify collection strategies in order to improve the work-out process. Understandably, changing the collection process will generate the need to update the LGD model (there is a clear analogy to the Lucas critique, see section 4.3.2). In order to

test effectiveness of the new model based strategies, a champion/challenger approach can be used (Thomas *et al.*, 2002).

Furthermore, the predictive distributions of LGD can help set a cut-off for the score used to accept and reject applicants. This should be based on a sample of similar loans that have already been granted. The loans need to be ranked according to the scores at application. Having the estimates of PD, LGD and EAD, one can compute the expected loss for each loan from the sample (this 12-month estimate would need to be adjusted for the loan lifetime expected loss to take a long term perspective). One can also calculate the expected profit made with the complement probability (1 – PD). Then the probability-weighted sum of the expected profit and loss can be computed for each loan. As a result, there can be an estimate of profit/loss on the entire portfolio for each level of the cut-off. The above calculations can involve the LGD quantile which reflects possible worsening of the economic situation (in particular, the downturn LGD can be used along with the downturn PD). Then a cut-off can be chosen that corresponds to the break-even point, i.e. neither profit nor loss on the portfolio. With such a cut-off, normally there should be a profit, but even in adverse economic conditions, loss is unlikely.

Finally, the individual predictive distributions, and credible intervals in particular, offer the benchmarks which can help confirm that the selected LGD estimates are sufficiently conservative.

3.7 Conclusions

In this research, Bayesian methods are compared and contrasted with the frequentist two-step approach to modelling LGD for unsecured retail loans. Three practical suggestions on the use of the latter are presented and called ‘cut-off’, ‘random cut-off’ and ‘probability times value’. Two of them (random cut-off and probability times value) have been implemented in the Bayesian framework. Then both the Bayesian and frequentist approaches have been applied to the data on personal loans granted by a large UK bank.

As expected, the posterior means of the parameters which have been produced in the Bayesian framework are very similar to the frequentist estimates. The posterior means and standard deviations of the model performance measures are also almost identical as the corresponding bootstrap estimates that have been generated in the frequentist random cut-off approach. In comparison with the random cut-off approach, the probability times value approach has yielded slightly better posterior means of the performance measures. Again, the posterior means are almost the same as the results obtained in the frequentist probability times value approach.

In spite of the similar performance, the Bayesian model is free from the drawbacks of the frequentist approach. It is more coherent and allows for a much better description of uncertainty. The most important advantage of the Bayesian model is that it generates an individual predictive distribution of LGD for each loan, whereas the frequentist approach only produces a point estimate. The predictive distributions provide a lot of information (including benchmarks for LGD estimates) and can be used, among other purposes, for stress testing and approximating the downturn LGD in case of downturn data scarcity.

Obviously, it is possible to generate some distributions of LGD within the frequentist framework. One way is taking into account the standard error of the predicted LGD from the second model (linear regression). This allows for the determination of confidence intervals after the adoption of the normality assumption (e.g. Maddala, 2001). If the error term is assumed to follow a normal distribution, then the predicted LGD follows a normal distribution, too. That approach has serious drawbacks. It assumes normality of the error term and – in consequence – also of the LGD distribution, whereas empirical LGD distributions are known for being far from normal-shaped. Furthermore, it ignores uncertainty from the first model (logistic regression), which may lead to confidence intervals being too narrow.

Another way to generate LGD distributions is using bootstrap methods. If the sample is large, the results may be numerically similar to those obtained in the Bayesian framework. However, if the sample is small, the Bayesian approach offers the advantage of utilising the prior information, which can be useful e.g. in case of LDPs. It

is also worth remembering that Bayesian methods yield distributions of the model parameters, whereas the bootstrap only produces distributions of their estimators (Rubin, 1981). As a result, Bayesian credible intervals have much more natural and straightforward interpretation than bootstrap-based confidence intervals (Jaynes, 1976). Differences between the two approaches are both technical and philosophical, and the choice is up to the potential user. Nevertheless, there are some connections between these approaches (Efron, 2003), including the so-called Bayesian bootstrap (Rubin, 1981).

Yet another way to obtain LGD distributions is using survival analysis (Zhang and Thomas, 2012). In survival analysis, the time until an event occurs is usually modelled. Zhang and Thomas (2012) applied the Cox proportional hazards model, but instead of the time, they estimated how much is recovered until the end of the collection process (or – in case of censored observations – the end of the period covered by data). In consequence, they obtained a probability of being in the collection process for each value of the Recovery Rate (RR = 1 – LGD), which gives the RR distribution for each loan. However, the distributions derived from the Cox proportional hazards model have a major drawback. Since hazard function lines of different loans never cross one another, the ranking of loans is the same for each quantile of the distributions. The Bayesian approach which has been proposed in this research is free from such limitations and thus much more flexible.

It seems that a similar Bayesian hierarchical model could be applied to model mortgage LGD. It could replace separate repossession and haircut models (see section 1.2.3).

Further modifications of this approach could include using more informative priors, which might be beneficial in case of smaller samples than in this application. If no extra information is available, one could collect expert opinion and employ the elicitation methods to transform it into the prior distributions (O'Hagan *et al*., 2006). The expert opinion could be provided by some industry representatives on the condition of anonymity, whereas in banking practice it could be obtained from internal sources.

Moreover, one could apply Bayesian model selection to find the best covariates of the logistic and linear regressions (when the model structure is fixed, variable selection is equivalent to model selection). In Bayesian model selection, the posterior model probabilities are compared (Wasserman, 2000). The best model is identified as the one with the highest posterior probability amongst the analysed models. Alternatively, Bayesian model averaging could be performed instead of model selection (Wasserman, 2000).

Finally, one could try to improve the model performance by changing the approach. In the Bayesian framework, one could use more sophisticated models than the regressions and employ some transformations of LGD (as it could be done in the frequentist framework). One could also apply more complex Bayesian graphical models (Madigan and York, 1995), although this might be computationally expensive.

Chapter 4

Dynamic affordability assessment: predicting an applicant's ability to repay over the life of the loan[3]

4.1 Introduction

The concepts of an affordable loan and affordability assessment are introduced in section 1.2.5. There is little literature on statistical models and methods for affordability assessment. Finlay (2006) proposed a linear regression model to estimate expenditure to income ratio for such purposes and a logistic regression model to estimate probability of overindebtedness, both based on application data and credit reports. However, in the conclusions to his paper, the dynamic nature of income and expenditure is mentioned as a possible argument against the use of those static models. On the other hand, Thomas (2009a) presented a rough idea of structural models based on affordability where default is a result of cash flow problems. Although it was suggested with a view to modelling the credit risk of portfolios of consumer loans, it could also be applied for assessing affordability. In that approach, the asset process is modelled. Each month, the consumer's realizable assets are increased by his or her income and reduced by both expenditure and loan repayment. Once realizable assets become negative or fall below a percentage of the total debts, the consumer defaults. The dynamics of the asset process could be modelled by treating income and expenditure as functions of economic conditions. In this research, the above-mentioned general idea, with some modifications, has been developed into a complete theoretical framework.

Introducing dynamics into consumer risk models is one of the current challenges in credit scoring (Crook and Bellotti, 2008; Thomas, 2011). Suggested approaches include Markov chains and survival analysis (e.g. Thomas *et al.*, 2001), panel data models

[3] This chapter is based on the following manuscript: Bijak, K., Thomas, L.C. and Mues, C. (2013) Dynamic affordability assessment: predicting an applicant's ability to repay over the life of the loan, under review in *The Journal of Credit Risk*.

(Crook and Bellotti, 2008; Crook, 2012), Kalman filtering (Whittaker *et al.*, 2007; Bijak, 2011) and using macroeconomic variables (Bellotti and Crook, 2007; Thomas, 2011). As far as affordability assessment is concerned, both the Office of Fair Trading (OFT) and the Financial Services Authority (FSA)[4] recommend taking a long term perspective and considering future changes in the applicant's income and expenditure. In practice, though, a static approach is often used, based on current income and estimated current consumption as well as existing debts reported by credit bureaus. Such an approach assumes that the customer's financial situation will stay the same in the future. As a result, it is likely to underestimate possible increase in consumption, which may lead to granting too much credit, overindebtedness and default. On the other hand, if possible increase in income is underestimated, the customer may be offered less credit than he or she would be able to repay and thus, the lender will lose potential profits. Contrary to that static approach, dynamic affordability assessment is proposed in this research.

In this research, affordability is defined as a function that assigns to each possible instalment amount a probability of the applicant defaulting over the loan repayment period. Consequently, affordability assessment means estimation of this function. It is assumed that the customer's income and consumption vary over time. Changes in income and consumption are modelled with random effects models for panel data (time-series cross-sections). Panel data analysis is suggested, since cross-sectional analysis does not allow for the introduction of dynamics, whereas time series analysis requires long observation periods and generally seems more suitable for modelling aggregate quantities. The model formulas are derived from the economic literature. Consumption is described with a log-linearized version of the Euler equation. The estimated models are then applied in a simulation that is run for the applicant. In each iteration, the predicted income and consumption time series are generated, and the customer's ability to repay is assessed over the life of the loan, for all possible instalment amounts. In consequence, each amount can be assigned with a probability of the event of interest (be it default or just failure to pay) over any time period. In particular, affordability can be

[4] On 1 April 2013 the FSA ceased to exist and most of its responsibilities were transferred to two new authorities: the Prudential Regulation Authority (PRA), which is a part of the Bank of England, and the Financial Conduct Authority (FCA).

assessed and the maximum affordable instalment can be identified. The design of this approach is such that a loan is affordable if the applicant is able to repay it while also meeting consumption costs and repayments of all other debts month after month until the loan is paid in full, which is in line with the guidelines of the OFT (2011) and the suggestions of the FSA (2010). The proposed approach is illustrated with an example based on artificial data.

This chapter is structured as follows. Section 4.2 is on the research background that covers overindebtedness, codes of practice and guidelines on responsible lending as well as affordability assessment solutions used in banking practice. In section 4.3, the methodology is presented (income and consumption change models, simulation design and affordability assessment). In section 4.4, artificial data are described. In section 4.5, an example based on the artificial data is demonstrated. Section 4.6 is a discussion on what data would be needed to apply this theoretical framework in practice. Section 4.7 includes a summary and conclusions.

4.2 Background

4.2.1 Overindebtedness

Affordability assessment is inextricably linked to the concepts of consumer overindebtedness and responsible lending. Irresponsible lending practices are blamed for exacerbating overindebtedness (Kempson, 2002). In particular, increasing the credit limit or granting credit without reasonable affordability assessment may lead to the customer being overindebted, which often ends in default. The financial crisis has raised interest in overindebtedness across Europe (Fondeville *et al.*, 2010). In the UK, the scale and drivers of this phenomenon have been intensively studied for over 10 years (e.g. Kempson, 2002; Oxera, 2004; Disney *et al.*, 2008; Bryan *et al.*, 2010).

There are many definitions of overindebtedness, e.g. “the circumstance where the household’s credit-financed spending plans are inconsistent with its potential income stream” (Disney *et al.*, 2008). According to Betti *et al.* (2001), there are three models (types of definitions) of overindebtedness: administrative, subjective and objective (quantitative). Under the administrative model, overindebtedness occurs when it is

declared before the court or registered by an official authority. The subjective model assumes that overindebted are those who self-define themselves as overindebted. Under the objective model, overindebtedness is assessed using such measures as debt service to income ratio. Using a mix of the latter two models, the Department of Trade and Industry (2005) listed the following indicators of overindebtedness: spending more than 25 per cent of gross income on repayments of unsecured loans, spending more than 50 per cent of gross income on repayments of both secured and unsecured loans, having four or more credit commitments, being in arrears for more than three months and considering repayments a 'heavy burden'.

However, Betti *et al.* (2001) criticised applying the same overindebtedness thresholds to all customers no matter what stage of life they are in. For example, young persons, whose incomes are likely to increase over time, can cope with higher debt to income ratios than older persons. Therefore, Betti *et al.* (2001) suggested taking into account not only the customer's current income but also their permanent income, i.e. expected income over a long period of time as defined by Friedman (1957). According to the Permanent Income Hypothesis (PIH), current consumption depends on permanent rather than current income and is sensitive to permanent but not transitory income shocks (Snowdon and Vane, 2005). Apart from the PIH, Betti *et al.* (2001) proposed applying the Life-Cycle Theory (Modigliani and Brumberg, 1954). According to this theory, consumers smooth their consumption over time, e.g. young persons may borrow against their expected future incomes. These suggestions, repeated by Disney *et al.* (2008), are in favour of a dynamic approach to affordability assessment.

4.2.2 Responsible lending

Affordability assessment is considered the main component of responsible lending, i.e. "acceptable practices that ensure borrowers can afford the repayments and know the consequences, and still try to accommodate as many people as possible" (Anderson, 2007, p. 627). Consequently, disregarding the significance of affordability assessment is one of the features of irresponsible (reckless) lending. The Consumer Credit Directive states that "it is important that creditors should not engage in irresponsible lending" (Council Directive 2008/48/EC, point 26).

Irresponsible lending is a worldwide problem and there are some legislative attempts to tackle it in many countries. For example, in the US, mortgage lenders must make a reasonable determination that "the consumer has a reasonable ability to repay the loan, according to its terms, and all applicable taxes, insurance (including mortgage guarantee insurance), and assessments" (Dodd-Frank Wall Street Reform and Consumer Protection Act, 2010, section 1411(a)(2)). Such a determination must include "the consumer's credit history, current income, expected income the consumer is reasonably assured of receiving, current obligations, debt-to-income ratio or the residual income the consumer will have after paying non-mortgage debt and mortgage-related obligations, employment status, and other financial resources" (section 1411(a)(2)). In Australia, lenders must assess "whether the credit contract will be unsuitable for the consumer if the contract is entered or the credit limit is increased in that period" (National Consumer Credit Protection Act 2009, paragraph 129(1)(b)). The contract will be unsuitable if it is likely that "the consumer will be unable to comply with the consumer's financial obligations under the contract, or could only comply with substantial hardship" (paragraph 131(2)(a)). In South Africa, lenders must take reasonable steps to assess "the proposed consumer's existing financial means, prospects and obligations" and, before increasing a credit limit, they "must complete a fresh assessment of the consumer's ability to meet the obligations that could arise under that credit facility" (National Credit Act 2005, sections 81(2)(a)(iii) and 119(3)). Lenders "must not enter into a reckless credit agreement with a prospective consumer", e.g. a credit agreement that "would make the consumer over-indebted", i.e. "unable to satisfy in a timely manner all the obligations under all the credit agreements to which the consumer is a party" (sections 81(3), 80(1)(b)(ii) and 79(1), respectively).

Examples of irresponsible lending practices include (among other things): lack of policies and procedures for reasonable affordability assessment, lack of affordability assessment in individual cases, failure to assess whether an applicant is likely to be able to repay in a sustainable manner, granting credit without having assessed affordability and granting credit when the affordability assessment results suggest that it is likely to be unsustainable (OFT, 2011). In the UK, such practices may even lead to revoking a consumer credit licence, since the Consumer Credit Act 2006 states that the practices

which look to the OFT as involving irresponsible lending are taken into account when considering the creditor's fitness to hold the licence (section 29, subsection (2)).

According to the best practice set out in the Guide to Credit Scoring, banks should assure applicants that "as responsible lenders, we take into account your personal circumstances to establish the appropriate level of credit to grant to you" (Association for Payment Clearing Services *et al.*, 2000, Appendix 2). In line with the Lending Code, which sets more standards of good practice for UK banks, "before lending any money, granting or increasing an overdraft or other borrowing, subscribers should assess whether the customer will be able to repay it in a sustainable manner" (British Bankers' Association *et al.*, 2011, Section 4, paragraph 50) and "before giving a customer a credit limit, or increasing an existing limit, subscribers should assess whether they feel the customer will be able to repay it" (Section 6, paragraph 115). Moreover, "issuers should undertake appropriate checks to assess a customer's ability to repay […] before increasing a credit limit" (The UK Cards Association, 2011, Section 2.4).

The OFT suggested that lenders use various sources of information to assess affordability, e.g. evidence of income and expenditure and/or credit reports provided by credit bureaus. If income or expenditure is used, one should take into account not only the applicant's current situation but also the expected future changes over the life of the loan. Generally, lenders are encouraged to view credit sustainability in a long term perspective: they can accept occasional missing of a payment on a due date or – in some circumstances – even temporary (initial) inability to repay (OFT, 2011, paragraphs 4.7 and 4.9).

As far as mortgages are concerned, the FSA proposed that lenders take into account the applicant's income, expenditure and debts, and calculate his or her free disposable income in order to assess affordability. They should use statistical data to estimate expenditure. Furthermore, they should assess the applicant's ability to repay over the loan repayment period, considering variability of income over time (FSA, 2010). However, these FSA suggestions were only put in a consultation paper and thus are not binding for banks.

To sum up, there are some codes of practice and guidelines on responsible lending, including affordability assessment, but they are rather general and do not advocate any specific statistical models or methods. Nevertheless, it is worth noticing that both the OFT and FSA recommend taking a long term perspective and considering future changes in the applicant's income and expenditure.

4.2.3 Banking practice

In the industry, there are concerns that if responsible lending criteria are too strict, the existing business model may not be sustainable any more (Wilkinson, 2007). There are also concerns that such criteria may limit consumer access to bank credit and, as a result, banks may lose their customers to non-banking financial companies that are not subject to any regulations on responsible lending. However, lending to those who can be reasonably identified as unlikely to repay is neither ethical nor profitable (although it can be part of a generally profitable, yet still unethical, business model). Therefore, accurate affordability assessment is important.

The affordability measure which is widely used in banking is debt service to income ratio, the same that can be used to assess overindebtedness. The debt service to income ratio can also be computed using application data, information on the applicant's credit commitments from credit bureaus as well as his or her expenditure estimate where expenditure is modelled on public data (Lucas, 2005). After taking into account the new instalment, this ratio can be compared to a threshold (cut-off) in order to assess affordability. Generally, approaches to affordability assessment are often based on information from the above-mentioned three sources: application data (including income), credit reports and estimation of expenditure. This allows for calculating disposable income (Dell, 2007; Maydon, 2011). The result can be then compared to the new instalment in the credit decision making process.

There are two approaches to affordability assessment for mortgages: income multiples and affordability models (FSA, 2009). The former are fixed and can only vary between groups of applicants; this is a 'one-fits-all' approach. The latter use estimates of the applicant's income and expenditure to calculate the maximum affordable loan amount for this customer. Various methods are applied and the models differ in their complexity

level. Large lenders use this approach more often than small lenders (FSA, 2009). Affordability models have a clear advantage over income multiples as they are based not only on income but also on expenditure. Thus, it is not surprising that they become more and more popular (FSA, 2009).

As far as credit cards are concerned, an affordability model can be applied to assess the impact of changes in credit limit on the customer's risk profile. For example, Somers (2009) built a model for Lloyds Banking Group that estimates the probability of the customer being bad (i.e. defaulting). This stepwise regression model takes into account the forecasted limit that is estimated using another model with a risk score as the only variable. In the affordability model, the following variables are used: a risk score, the log ratio of the actual limit and the forecasted limit as well as a number of characteristics multiplied by this log ratio. The latter are added to adjust the model outcome for those customers where the forecasted limit differs from the actual one. This is part of a solution designed to determine new credit limits.

Since it is impossible to assess affordability without information on the applicant's debts, credit bureaus seem a natural place to develop solutions that are dedicated to affordability assessment. An example of such a solution is Experian's Affordability Index (Experian, 2011). It is a multi-scorecard model where the customer's status definition is based not only on their delinquencies but also on the Consumer Indebtedness Index. Among factors which indicate a high indebtedness level are excessive credit activity and high utilisation of credit cards. When assessing affordability, Experian takes into account (among other things) the applicant's socio-demographic characteristics, income and credit commitments as well as his or her expenditure estimated using the Office for National Statistics (ONS) Expenditure and Food Survey (EFS) data (Russell, 2005; Brooksby, 2009). Another example is Callcredit's Affordability Suite which includes such tools as indicators based on debt to income ratios and a score to assess probability of default as a result of overindebtedness (Callcredit, n.d.).

It is difficult to conclude much about the affordability models observed by the FSA (2009), since their details are not publicly available. Most of the other above-mentioned

solutions, which are applied in banking, use at least some of the sources of information suggested by the OFT. Nevertheless, these approaches are static and, as far as it can be ascertained, none of them directly implement the OFT and FSA recommendations to assess the ability to repay over the life of the loan, taking into account variability of income and expenditure over time.

4.3 Methodology

4.3.1 Income change model

In this research, the proposed approach to affordability assessment is based on income and consumption models. There is much economic literature on modelling these quantities at the individual or household level; to mention just one example, Miles (1997) estimated income and consumption regressions. This and the next sections focus only on those models that are designed for panel data. Such models are less commonly used than models for cross-sectional data because of their higher complexity and lower availability of suitable datasets. However, panel data models have the advantage of not ignoring the fact that things change over time.

As far as income is concerned, net labour income is usually modelled. Similar models are built both at the household level (e.g. Guiso *et al.*, 1992; Jappelli and Pistaferri, 2006) and at the individual level (e.g. Auten and Carroll, 1999; Koskinen *et al.*, 2007). If household income is modelled, characteristics of the head of household are taken into account as well as family size or number of earners. Regardless of the modelling level (individual/household), similar regressors are included both where income is the dependent variable (e.g. Lillard and Willis, 1978; Guiso *et al.*, 1992; Etienne, 2006) and where income change is the dependent variable (e.g. Lusardi, 1992; Auten and Carroll, 1999; Jappelli and Pistaferri, 2006). Either the individual's characteristics or their changes can be used as regressors to model income change. Using the characteristics reflects the belief that the relationship between them and income may vary over time, e.g. earnings of more educated workers are likely to grow faster than earnings of less educated workers (Auten and Carroll, 1999). No matter how income is modelled, since its distribution tends to be right-skewed, the log transformation is often performed to

eliminate the skewness (e.g. Lusardi, 1992; Etienne, 2006; Jappelli and Pistaferri, 2006).

In income models, the following characteristics are most frequently used: education level, occupation, region, age and sex (see Table 4.1). Among other income determinants which are included in the models are sector of occupation (e.g. Guiso *et al.*, 1992) and year of birth (e.g. Etienne, 2006). The latter is used to control for the cohort effect. If one believes that younger generations are always better off than older ones and this relationship is linear, year of birth can be directly implemented in the model. Otherwise, one can consider its polynomial like Etienne (2006) or a set of dummy variables e.g. to indicate those cohorts who entered the labour market in recessions, since this might negatively affect their income for a long time. Obviously, the other above-mentioned variables, except for age, are coded as sets of dummies. Instead of age, one can use its polynomial (e.g. Etienne, 2006; Jappelli and Pistaferri, 2006).

Income model	***Education level***	***Occupation***	***Region***	***Age***	***Sex***
Auten and Carroll (1999)		✓	✓	✓	
Etienne (2006)	✓	✓		✓	
Guiso *et al.* (1992)	✓	✓	✓	✓	✓
Jappelli and Pistaferri (2006)	✓		✓	✓	✓
Lillard and Willis (1978)	✓		✓		
Lusardi (1992)	✓	✓			✓

Table 4.1. Income determinants in selected models

Although macroeconomic variables such as Gross Domestic Product (GDP) can be explicitly included in income models (e.g. Koskinen *et al.*, 2007), macroeconomic conditions (referred to as ‘aggregate shocks’ in the economic literature) are often taken

into account by using fixed time effects in the form of time dummies (e.g. Lillard and Willis, 1978; Lusardi, 1992; Jappelli and Pistaferri, 2006). Time dummies can capture the combined effect of macroeconomic variables that are not used as regressors in the model (Lillard and Willis, 1978). Thus, time effects describe the macroeconomic environment as a whole and not only its selected elements such as production or unemployment. In panel data models, one can also implement random time effects but this requires data covering long time periods and thus is rarely used. This is normally operationalized by using time dummies.

In income models for panel data, individual effects are components that are specific to households or individuals and are constant over time. In fixed effects (FE) models, individual effects are estimated along with the other parameters (e.g. Etienne, 2006), whereas in random effects (RE) models, individual effects are part of the error term (e.g. Lillard and Willis, 1978). The original formula for a RE model, which was developed by Balestra and Nerlove (1966), included also a random time-specific component but usually this component is either omitted or replaced. Since fixed effects control for all permanent characteristics, only time-variant regressors can be included in FE models unless a more complicated estimator, such as the Hausman-Taylor, is used. The Hausman-Taylor estimator enables estimating the effect of time-invariant regressors by using instrumental variables that are based on the time averages of the time-variant regressors (Verbeek, 2004). Another disadvantage of FE models is that they cannot be applied to predict for individuals or households outside the training sample because there are no estimates of their individual effects. If individual effects are assumed to be related to income and not to its change, they can be removed from the model by first differencing and, as a result, they can be absent in income change models (Auten and Carroll, 1999). Such a transformation potentially allows for using the first difference (FD) estimator that is a more convenient estimation method (Wooldridge, 2010). However, first differencing eliminates also time-invariant regressors from the model and in practice it would rule out most characteristics. Therefore, a RE model is more suitable to predict income change.

When assessing affordability, the applicant's current income is known. Starting with this initial value, their income in consecutive months can be predicted using an income

change model. Taking into account all the above considerations, the following RE model is proposed for the purposes of this research:

$$\Delta lnY_{it+1} = \alpha_0 + \alpha_1 age_{it} + \alpha_2 cohort_i + \alpha_3 sex_i + \alpha_4 education_i + \alpha_5 occupation_i + \alpha_6 sector_i + \alpha_7 region_i + \mu_i + \eta_{t+1} + \varepsilon_{it+1}$$

which, after operationalizing random time effects η_{t+1} as time dummies $D_{t+1}^{(s)}$, gives:

$$\Delta lnY_{it+1} = \sum_{s=1}^{T-1} \gamma_s D_{t+1}^{(s)} + \alpha_0 + \alpha_1 age_{it} + \alpha_2 cohort_i + \alpha_3 sex_i + \alpha_4 education_i + \alpha_5 occupation_i + \alpha_6 sector_i + \alpha_7 region_i + \mu_i + \varepsilon_{it+1}$$

where Y_{it+1} is the ith customer's income in month $t+1$ and age_{it} represents their age in month t. The other characteristics are assumed to be constant as they typically remain relatively time-invariant. Sex is included since, "after discussion with industry experts", Finlay (2006) came to a conclusion that it may be allowed in affordability models, although it is debatable. If T denotes the number of months in the training sample, there are $T-1$ time dummies $D_{t+1}^{(1)}, \ldots, D_{t+1}^{(T-1)}$ such that:

$$D_{t+1}^{(s)} = \begin{cases} 1 & \text{if } s = t+1 \\ 0 & \text{otherwise} \end{cases}$$

In this model, as in most RE models, the error term is the sum of the random individual effect μ_i (specific to the customer and time-invariant) and an idiosyncratic component ε_{it+1} (also customer-specific but varying over time). Using a RE model requires adopting some assumptions on the error term elements (Greene, 2000): μ_i and ε_{it} are orthogonal and both of them are white noise, i.e. they have zero means and are spherical (homoscedastic and not serially correlated):

$$E(\mu_i \varepsilon_{jt}) = 0 \quad \text{for all } i \text{ and } j \text{ and } t$$
$$E(\mu_i) = E(\varepsilon_{it}) = 0 \quad \text{for all } i \text{ and } t$$

$$E(\mu_i \mu_j) = \begin{cases} \sigma_\mu^2 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

$$E(\varepsilon_{it} \varepsilon_{js}) = \begin{cases} \sigma_\varepsilon^2 & \text{if } i = j \text{ and } t = s \\ 0 & \text{otherwise} \end{cases}$$

Due to the presence of the random individual effect, there is autocorrelation of the error term and thus, the Generalised Least Squares (GLS) estimator is recommended. For more details on RE models and their estimation, one can refer to the appropriate chapters in general econometrics textbooks (e.g. Greene, 2000; Maddala, 2001; Verbeek, 2004) or the panel data literature such as Wooldridge (2010). RE models can be estimated using popular statistical software packages such as SAS or Stata.

4.3.2 Consumption change model

Consumption is usually modelled using Euler equations. The Euler equation of consumption was first used by Hall (1978), who proposed a random walk model inspired by the Lucas critique. In his seminal paper, Lucas (1976) pointed out that the then-used macroeconometric models were sensitive to changes in policy rules and once the rules were changed, the models were no longer valid, even though they had been developed for the policy makers. He argued that, under the rational expectations hypothesis, economic agents (including consumers) rationally adjust their expectations to changing policy rules, which should be taken into account in the models. Therefore, in response to the Lucas critique, microfoundations were incorporated into macroeconometric models according to the belief that analysis of individual economic agents' expectations and behaviour may help understand the economy (Snowdon and Vane, 2005). These ideas led to the formulation of the Euler equation of consumption, i.e. an intertemporal first-order condition for the consumer's optimisation problem (there is no direct equivalent for income, since most people have only limited impact on their income and cannot choose its level in the same way in which they can make their consumption choices).

The Euler equation and its log-linearized version which are presented below have been partly motivated by those applied by Zeldes (1989), Runkle (1991) and Lusardi (1992). In this approach, the ith consumer has a constant relative risk aversion utility function:

$$U(C_{it}, \theta_{it}) = \frac{C_{it}^{1-a}}{1-a} exp(\theta_{it})$$

where C_{it} is their consumption at time t, a is the Arrow-Pratt measure of relative risk aversion and θ_{it} represents factors that shift the consumer's tastes. The absolute risk aversion measures the curvature of a consumer's utility function. The Arrow-Pratt measure of relative risk aversion is the absolute risk aversion calculated for the given consumption and multiplied by this consumption (Huang and Litzenberger, 1988). If it is assumed to be constant, it can be represented by the risk aversion coefficient a.
The taste-shifting factors include age (expressed in the same units as t) and other characteristics represented here by the generic variable X_{it}. They also contain a consumer-specific component b_i, a time-specific component b_t and an idiosyncratic component b_{it} that is orthogonal to both b_i and b_t (it is also assumed that all components have zero means):

$$\theta_{it} = b_0 age_{it} + b_1 age_{it}^2 + b_3 X_{it} + b_i + b_t + b_{it}$$

The Euler equation is an equilibrium condition. If the consumer makes optimal consumption choices, then their current marginal utility is equal to the present value of the expected future marginal utility corrected for their time preference rate:

$$U'(C_{it}, \theta_{it}) = E_t \left[\frac{U'(C_{it+1}, \theta_{it})(1 + r_i)}{(1 + d_i)} \right]$$

where U' is the derivative of U with respect to the consumption; r_i and d_i are the interest rate and the time preference (discount) rate, respectively (in this version of the Euler equation, the interest rate is constant over time but may vary between consumers). The ratio of the marginal utilities corrected for r_i and d_i is equal to one plus the expectation error e_{it+1}:

$$\frac{U'(C_{it+1}, \theta_{it+1})(1 + r_i)}{U'(C_{it}, \theta_{it})(1 + d_i)} = 1 + e_{it+1}$$

The expectation error has zero mean and variance σ_e^2. The relationship between the interest rate and the time preference rate shapes the individual's consumption path over time. If r_i and d_i are assumed to be equal, they eliminate each other from the equation (e.g. Lusardi, 1992). In this research, a more general assumption is adopted. Since both r_i and d_i are consumer-specific, their relationship is also specific to the consumer:

$$\frac{1+r_i}{1+d_i} = 1 + \frac{r_i - d_i}{1+d_i} = 1 + z_i$$

The mean of z_i equals zero and its variance is σ_z^2. Moreover, z_i and b_{it} are independent and so are z_i and e_{it}. The formula for the marginal utilities ratio is linearized by taking logs. The second-order Taylor approximation of a function $ln(1+x)$ is given by $ln(1+x) \cong x - \frac{x^2}{2}$. Using such approximations of $ln(1+z_i)$ and $ln(1+e_{it+1})$ results in the following consumption change model:

$$\Delta lnC_{it+1} = \beta_0 + \beta_1 age_{it} + \beta_2 \Delta X_{it+1} + \nu_i + \lambda_{t+1} + \zeta_{it+1}$$

where:

$$\beta_0 = \frac{b_0 + b_1 - \frac{\sigma_z^2}{2} + \frac{\sigma_e^2}{2}}{a}$$

$$\beta_1 = \frac{2b_1}{a}$$

$$\beta_2 = \frac{b_2}{a}$$

$$\nu_i = \frac{\left(z_i - \frac{z_i^2}{2}\right) + \frac{\sigma_z^2}{2}}{a}$$

$$\lambda_{t+1} = \frac{b_{t+1} - b_t}{a}$$

$$\zeta_{it+1} = \frac{(b_{it+1} - b_{it}) - \left(e_{it+1} - \frac{e_{it+1}^2}{2}\right) - \frac{\sigma_e^2}{2}}{a}$$

In this model, the error term is the sum of the individual effect ν_i, the time effect λ_{t+1} and an idiosyncratic component ζ_{it+1} (the original consumer-specific component b_i has been ruled out from the model by taking differences). In order to make the means of the error term elements equal zero, $\frac{\sigma_z^2}{2a}$ and $-\frac{\sigma_e^2}{2a}$ have been added to ν_i and ζ_{it+1}, respectively, and then subtracted from the intercept β_0.

In such models, nondurable consumption change is usually modelled. However, nondurable consumption is often limited to food expenditure because of data availability (e.g. Hall and Mishkin, 1982). Although the Euler equation was originally formulated at the individual level, most models are developed at the household level, since household surveys are the main source of panel data on consumption. The model built by Finlay (2006) to estimate expenditure to income ratio is also at the household level. Nevertheless, a loan application (including affordability) is usually assessed at the individual level (unless it is a joint application). Therefore, the models proposed in this research are at the individual level as well.

There are just a few characteristics that are typically used in consumption change models: age of the head of household as well as change in the number of children and change in the number of adults or in the family size (e.g. Hall and Mishkin, 1982; Zeldes, 1989; Lusardi, 1992; Jappelli and Pistaferri, 2000). Instead of age, its polynomial can be implemented (e.g. Hall and Mishkin, 1982). Instead of the number of all children, one could consider the number of children in different age groups, since they have different consumption needs. Other variables, such as income change, are added to test economic hypotheses. However, income variables turn out to be insignificant in some consumption change models, which suggests that current consumption does not depend on current income and thus supports the PIH (e.g. Runkle, 1991). As in income models, aggregate shocks are often taken into account by using fixed time effects in the form of time dummies (e.g. Zeldes, 1989; Lusardi, 1992). Individual effects are sometimes also included: for example, Zeldes (1989) incorporated household-specific components as fixed effects into a consumption change model.

Since FE models cannot be applied to predict outside the training sample, a RE model of consumption change is proposed for the purposes of this research:

$$\Delta lnC_{it+1} = \beta_0 + \beta_1 age_{it} + \beta_2 \Delta children_{it+1}^{(0-3)} + \beta_3 \Delta children_{it+1}^{(4-15)} + \beta_4 \Delta children_{it+1}^{(16-19)} + \nu_i + \lambda_{t+1} + \zeta_{it+1}$$

which can be operationalized by replacing random time effects λ_{t+1} with time dummies:

$$\Delta lnC_{it+1} = \sum_{s=1}^{T-1} \delta_s D_{t+1}^{(s)} + \beta_0 + \beta_1 age_{it} + \beta_2 \Delta children_{it+1}^{(0-3)} + \beta_3 \Delta children_{it+1}^{(4-15)} + \beta_4 \Delta children_{it+1}^{(16-19)} + \nu_i + \zeta_{it+1}$$

where C_{it+1} is the ith customer's consumption in month $t+1$ and e.g. $children_{it+1}^{(0-3)}$ represents the number of children aged zero to three years old. As far as the error term elements are concerned, the same assumptions are adopted on them as on μ_i and ε_{it} in the income change model: ν_i and ζ_{it} are orthogonal and both of them are white noise.

Although the above model is designed for panel data and includes time-variant regressors, it is not a dynamic model *sensu stricto* since it is not autoregressive (Maddala, 2001). This caveat also relates to the income change model. Nevertheless, the proposed approach to affordability assessment is dynamic in nature.

4.3.3 Multi-equation models (optional)

If one believes that, contrary to the PIH, current income may affect consumption, then income change should be added as a regressor to the consumption change model. In such case, the two models cannot be estimated independently any more. Instead, they should be treated as a system of equations or – since they describe casual relationships – a structural equation model. Subsequently, one could estimate a two-equation recursive model (i.e. a special case of a simultaneous equations model):

$$\Delta lnY_{it+1} = \alpha_0 + \alpha_1 age_{it} + \alpha_2 cohort_i + \alpha_3 sex_i + \alpha_4 education_i + \alpha_5 occupation_i + \alpha_6 sector_i + \alpha_7 region_i + \mu_i + \eta_{t+1} + \varepsilon_{it+1}$$

$$\Delta lnC_{it+1} = \beta_0 + \beta_1 age_{it} + \beta_2 \Delta children_{it+1}^{(0-3)} + \beta_3 \Delta children_{it+1}^{(4-15)} + \beta_4 \Delta children_{it+1}^{(16-19)} + \beta_5 \Delta lnY_{it+1} + \nu_i + \lambda_{t+1} + \zeta_{it+1}$$

If one believes that credit card limit, which depends on income, may also have an impact on consumption, then a three-equation recursive model seems appropriate:

$$\Delta lnY_{it+1} = \alpha_0 + \alpha_1 age_{it} + \alpha_2 cohort_i + \alpha_3 sex_i + \alpha_4 education_i + \alpha_5 occupation_i + \alpha_6 sector_i + \alpha_7 region_i + \mu_i + \eta_{t+1} + \varepsilon_{it+1}$$

$$\Delta lnC_{it+1} = \beta_0 + \beta_1 age_{it} + \beta_2 \Delta children_{it+1}^{(0-3)} + \beta_3 \Delta children_{it+1}^{(4-15)} + \beta_4 \Delta children_{it+1}^{(16-19)} + \beta_5 \Delta lnY_{it+1} + \beta_6 \Delta lnL_{it+1} + \nu_i + \lambda_{t+1} + \zeta_{it+1}$$

$$\Delta lnL_{it+1} = \psi_0 + \psi_1 account\_age_{it} + \psi_2 \Delta lnY_{it+1} + \omega_i + \chi_{t+1} + \xi_{it+1}$$

where L_{it+1} is the ith customer's credit card limit in month $t+1$ and $account\_age_{it}$ represents age of the credit card account in month t. As in the other equations, ω_i, χ_{t+1} and ξ_{it+1} are the random individual effect, the random time effect and an idiosyncratic component, respectively. In order to estimate the three-equation model, one would need to additionally obtain data on customers' credit card accounts from credit bureaus. Obviously, estimating multi-equation models requires another, more sophisticated econometric apparatus. To find out more on estimating systems of equations and simultaneous equations models for panel data, it is recommended to refer to Wooldridge (2010). The rest of this chapter focuses on the approach with separate income and consumption change models that can be estimated independently, since none of the dependent variables double as regressors in this approach.

4.3.4 Simulation

The proposed models should be estimated on a training sample and tested/validated on a hold-out sample. The results will contain estimates of the model parameters ($\widehat{\boldsymbol{\alpha}}$, $\widehat{\boldsymbol{\beta}}$, $\widehat{\boldsymbol{\gamma}}$, $\widehat{\boldsymbol{\delta}}$) as well as variances of the individual effects ($\hat{\sigma}_\mu^2$, $\hat{\sigma}_\nu^2$) and the idiosyncratic components ($\hat{\sigma}_\varepsilon^2$, $\hat{\sigma}_\zeta^2$). The final models may slightly differ from the proposed ones, since any variables which turn out to be insignificant should be removed from the equations.

Once the models are estimated, future income and consumption can be predicted for any applicant whose current income and expenditure are known (e.g. stated in the loan application). A simulation can be run to take into account the random components (individual effects and idiosyncratic components) as well as unknown future

macroeconomic conditions (time effects). It is assumed that the individual effects and idiosyncratic components follow normal distributions.

As far as the macroeconomic environment is concerned, it is assumed that each future month is similar to one of the months in the training sample. This is especially true if the training sample covers a sufficiently long time period. Then the set of dummy variables control for the macroeconomic conditions that occur over the whole economic cycle. Thus, in the simulation, for each future month a time dummy is randomly selected (this would be replaced with drawing numbers from another normal distribution, if random time effects were used). The random selection of time dummies is a conservative approach that stems from the lack of knowledge of future macroeconomic conditions. If one has reliable macroeconomic forecasts, the randomly selected dummy variables can be replaced with a sequence of dummies that best describe the forecasted development of the macroeconomic situation.

Apart from the time dummies, the only time-variant regressors are age and changes in the number of children in different age groups. The latter can be calculated if the children's age is known at the time of application. It is assumed that the customer will not have more children in the loan repayment period. In each iteration of the simulation, the applicant's income and consumption are predicted over the repayment period that starts in month $A + 1$, i.e. the next month after the loan application is made, and lasts for P months.

Each iteration comprises of the following steps:

1) The initial values are set as: $\hat{Y}_A = Y_A$ and $\hat{C}_A = C_A$ (at the time of application);
2) M is randomly drawn from $N\left(0, \hat{\sigma}_\mu^2\right)$;
3) N is randomly drawn from $N(0, \hat{\sigma}_\nu^2)$;
4) For each month $t + 1 = A + 1, \dots, A + P$:
   a) E_{t+1} is randomly drawn from $N(0, \hat{\sigma}_\varepsilon^2)$;
   b) Z_{t+1} is randomly drawn from $N\left(0, \hat{\sigma}_\zeta^2\right)$;
   c) S is randomly selected from among $1, \dots, T$ (since T is a reference category for the time dummies, it is assumed that $\hat{\gamma}_T = \hat{\delta}_T = 0$);

d) The changes $\Delta ln\hat{Y}_{t+1}$ and $\Delta ln\hat{C}_{t+1}$ are predicted using the estimated models (the subscript i is omitted, since the simulation is run for a given applicant):

$$\Delta ln\hat{Y}_{t+1} = \hat{\gamma}_S + \hat{\alpha}_0 + \hat{\alpha}_1 age_t + \hat{\alpha}_2 cohort + \hat{\alpha}_3 sex + \hat{\alpha}_4 education + \hat{\alpha}_5 occupation + \hat{\alpha}_6 sector + \hat{\alpha}_7 region + \mathrm{M} + \mathrm{E}_{t+1}$$

$$\Delta ln\hat{C}_{t+1} = \hat{\delta}_S + \hat{\beta}_0 + \hat{\beta}_1 age_t + \hat{\beta}_2 \Delta children_{t+1}^{(0-3)} + \hat{\beta}_3 \Delta children_{t+1}^{(4-15)} + \hat{\beta}_4 \Delta children_{t+1}^{(16-19)} + \mathrm{N} + \mathrm{Z}_{t+1}$$

e) The predicted income and consumption $\hat{Y}_{t+1}$ and $\hat{C}_{t+1}$ are calculated as:

$$\hat{Y}_{t+1} = exp\left(ln\hat{Y}_t + \Delta ln\hat{Y}_{t+1}\right)$$
$$\hat{C}_{t+1} = exp\left(ln\hat{C}_t + \Delta ln\hat{C}_{t+1}\right)$$

The above steps are repeated many times: there are e.g. 1000 or even better 10000 iterations in the simulation. As a result, a large number of pairs of the predicted income and consumption time series are generated. They represent various possible paths of development of the customer's financial situation. For each of them, the ability to repay can be assessed.

4.3.5 Affordability check

In this section, it is explained how the applicant's ability to repay is assessed for a given pair of the predicted income and consumption time series and a given instalment amount of the new loan. Since the final result of this assessment is binary (default or no default), it is referred to as an 'affordability check'. It is described with an example where a customer has a credit card and is applying for an instalment loan. However, it can be adapted to any portfolio of credit cards and loans (including mortgages). The information on the applicant's debts can be obtained from credit bureaus. Understandably, it is assumed that no other loans or credit cards will be granted to the customer in the loan repayment period.

4.3.5.1 Order of payments

For simplicity's sake, it is assumed that all transactions are made once a month: the customer gets his or her income, meets expenditure and makes other payments. He or she behaves rationally and makes optimal consumption choices. If there is enough money to meet all commitments, order of payments does not matter. Otherwise, the order is important. Consumption costs are always covered first. Loan payments are made before credit card payments. Furthermore, loan arrears are settled before on-time instalment payments, which is how lenders usually allocate money that comes into their account. Finally, the customer pays as much towards their credit card balance as they can after all other commitments are met (although this is not an obligatory payment).

To sum up, the following order of payments is assumed:

1) Consumption;
2) Loan arrears;
3) Loan instalment;
4) Credit card minimum payment;
5) Credit card balance.

Alternatively, the customer may prefer to make the credit card minimum payment before loan payments. Nevertheless, consumption, minimum payment, loan arrears and instalment will be referred to as 'obligatory payments'. Obviously, there may be no arrears to pay, and if a credit card is not used, there is no minimum payment, either. If the full credit card balance is paid, minimum payment is not required any more.

4.3.5.2 Making payments

Each month the customer tries to meet all commitments (including the full credit card balance) out of income only. If this is not possible, he or she uses both income and savings, and the latter are reduced afterwards. If income and savings are not enough to make all obligatory payments and the allocated limit is at least partly available, the customer also uses a credit card. Naturally, this makes the credit card balance rise. Since all transactions are made once a month, the next month's initial balance is the final balance from a given month.

The customer makes as many payments as they can, according to the order of payments. If they cannot meet all commitments, the last one is likely to be only partly met (e.g. a part of the instalment may be paid). The unpaid instalment or its part increases the loan arrears. The unpaid (part of) credit card interest increases the credit card balance. If minimum payments are missed or not fully made in three consecutive months, the credit card is suspended. Arrears as well as savings roll from month to month and can cumulate over time.

4.3.5.3 Saving

If there is any money left after all commitments are met (including the full credit card balance being paid), it can be saved and used later when needed. Repaying a loan out of savings is still considered by the OFT as meeting repayments 'in a sustainable manner' (OFT, 2011, paragraph 4.3). It is assumed that a fraction p of the money left is saved and $1-p$ is spent e.g. on durable consumption. If $p=1$, then all the money is saved. If $p=0$, saving is not allowed.

4.3.5.4 Reducing consumption

If the customer cannot make all obligatory payments in a given month $t+1$, they may reduce their consumption by a small fraction q so that $\hat{C}^*_{t+1} = (1-q)\hat{C}_{t+1}$ (if $q=0$, reducing consumption is not allowed). If an even smaller reduction is enough, then $(1-q)\hat{C}_{t+1} < \hat{C}^*_{t+1} < \hat{C}_{t+1}$. Since $\hat{C}^*_{t+1}$ is not the consumer's optimal choice, it is not used to calculate the estimated consumption in the next month. Although limiting expenditure in an attempt to avoid missing payments seems a very likely scenario, one can ask whether the loan is still affordable when a customer is forced to reduce their consumption to meet other commitments. For example, if a consumer has to give up 5% of their expenditure, can they still afford 'normal/reasonable outgoings' as the OFT expects (OFT, 2011, paragraph 4.4)? The answer is up to a potential user of this approach.

4.3.5.5 Failing to pay and defaulting

It is assumed that the customer fails to pay in a given month if they cannot make all obligatory payments even after a consumption reduction. If they fail to pay in three consecutive months, they default. This definition is similar to those used by credit bureaus in that that it does not matter on which loan/credit the default occurs. In this

respect, it is in line with the OFT recommendations which state that a customer should be able to make other debt repayments as well (OFT, 2011, paragraph 4.4). However, it is also possible to analyse only failures and defaults on the new loan.

4.3.5.6 Miscellaneous

In order to avoid modelling inflation rate, it is assumed that income and consumption are in the application time's pounds. A similar assumption was adopted e.g. by Lillard and Willis (1978). It is also assumed that the customer can neither lend nor invest their money, and cannot realize assets, such as properties, to make payments. According to the OFT, having to realize assets means that the loan is not repaid 'in a sustainable manner' (OFT, 2011, paragraph 4.3).

4.3.6 Affordability assessment

The dynamic affordability assessment is based on affordability checks for all pairs of the predicted income and consumption time series that have been generated in the simulation. As a result of the affordability checks, for each pair of time series there is a prediction whether and in which month(s) the customer will fail to pay or default. Since there are a large number of such pairs, a proportion of those, where defaults are predicted to occur, can be an estimate of probability of default over the loan repayment period. Probability of failure can be estimated in a similar way. Probabilities of default and failure can be calculated not only for the whole repayment period, but also for shorter periods such as the first year of repayment. However, all these probabilities are only for a given instalment amount.

As far as affordability is concerned, Thomas (2009a) suggested that the probability of the customer being good/bad may be a function of the interest rate charged on the loan. In this research, affordability is defined as a function $A(x)$ that assigns to each possible instalment amount $x \in X$ a probability of the applicant defaulting over the loan repayment period; A is continuous but can be approximated with a discrete function. In order to estimate this function, affordability checks for all pairs of time series need to be repeated for all possible instalment amounts (e.g. £500, £501, £502, …, £1000). Similarly, one can estimate a function that assigns a probability of failure instead of a probability of default. Nevertheless, in this research, affordability is linked to the latter,

since according to the OFT, a loan can be considered as being repaid 'in a sustainable manner' even despite occasional failing to pay (OFT, 2011, paragraph 4.7).

Once affordability is assessed, one can find the maximum affordable instalment MAI that corresponds to the maximum affordable loan for the applicant. One can take the last amount that is associated with acceptable probability of default. Therefore, MAI can be identified as the highest possible instalment amount $x \in X$ for which affordability is less or equal to the cut-off (5% for the sake of the example or any other value that is deemed appropriate):

$$MAI = max\{x \in X: A(x) \leq 0.05\}$$

Alternatively, one can take the last amount before a sharp increase in probability of default. Thus, MAI can be determined as the highest reasonable instalment amount $x \in X_R$ for which marginal affordability does not exceed the threshold (e.g. 0.1%):

$$MAI = max\{x \in X_R: A'(x) \leq 0.001\}$$

where A' is the derivative of A with respect to x and $X_R \subset X$ (the estimated function is likely to be S-shaped and, after the sharp increase, marginal affordability can become low again but for high, unreasonable amounts).

If one is interested in identifying the maximum affordable instalment rather than assessing affordability, it is possible to use the bisection method to reduce the computation time. When, for example, the cut-off is set to 5%, the algorithm works as follows:

1) The initial values are set as: $x_L = min\{X\}$ and $x_U = max\{X\}$;
2) The following steps are repeated until convergence is reached:
   a) The midpoint is calculated as $x_M = E\left(\frac{x_L + x_U}{2}\right)$;
   b) If $A(x_M) \leq 0.05$, then $x_L = x_M$;
   c) If $A(x_M) \geq 0.05$, then $x_U = x_M$;
3) The maximum affordable instalment is determined as $MAI = x_M$.

It has been assumed that, as long as all obligatory payments are made, the loan is still repaid 'in a sustainable manner' and thus affordable, even if the customer occasionally needs to use a credit card to cover part of their consumption costs. However, the maximum affordable instalment could be redefined in a more conservative way. For example, one could take the highest amount such that the customer will avoid default without the need to use a credit card in the loan repayment period with at least 95% probability. Obviously, the resulting instalment amounts will be generally lower.

In this research, it is argued that the dynamic affordability assessment is in line with recommendations of the OFT and FSA. Firstly, the applicant's ability to repay is assessed over the life of the loan. Secondly, possible future changes in their income and expenditure are taken into account. Finally, in this approach, a loan is affordable if the applicant is able to repay it while also meeting consumption costs and repayments of all other debts month after month until the loan is paid in full.

The proposed methodology is suggested with a view to assessing affordability and determining the maximum affordable instalment in the credit decision making process (at the time of application). Understandably, the same income and consumption change models should be used for all applicants. The simulation needs to be run for each applicant separately, since it is applicant-specific. Nevertheless, with modern computer technology this should not pose a problem in practice.

4.4 Artificial data

The dynamic affordability assessment is illustrated with an example based on artificial data. In this example, a hypothetical forty-five-year-old childless man is applying for an instalment loan with a two-year (twenty-four-months) repayment period. What needs to be determined is the maximum affordable instalment. At the time of application, the customer's net income and expenditure are equal to £2300 and £1500, respectively. He has a credit card with a limit of £1000. The minimum payment is the greater of interest plus 1% of the credit card balance and £5 (or the full balance if it is less than £5). The monthly interest rate is fixed at 1.5%. There are no default fees/charges if an instalment is missed or the minimum payment is not paid on time (such fees and charges can be easily introduced, though). In the first month of the loan repayment period, the customer

has no savings which could help him meet commitments but the full credit card limit is available. The latter assumptions can be modified according to the lender's knowledge by adopting some initial values of savings and/or the credit card balance.

It is assumed that the income and consumption change models have been built on a five-year (sixty-month) training sample so that there are 59 time dummies. As a result, there are some estimates of the model parameters as well as variances of the individual effects and the idiosyncratic components (see Table 4.2). In the absence of available data, their values have been chosen arbitrarily here and for illustration purposes only. On the basis of these estimates, the simulation has been run for the above-mentioned hypothetical applicant. The simulation has consisted of 10000 iterations.

Estimates	***Values***
$\hat{\gamma}_1, \dots, \hat{\gamma}_{59}$	From $-1.5 \cdot 10^{-3}$ to $1.5 \cdot 10^{-3}$
$\hat{\alpha}_0$	10^{-3}
$\hat{\alpha}_1$	$5 \cdot 10^{-6}$
$\hat{\alpha}_2 cohort + \hat{\alpha}_3 sex + \hat{\alpha}_4 education + \hat{\alpha}_5 occupation + \hat{\alpha}_6 sector + \hat{\alpha}_7 region$	10^{-4} (a value of the whole expression for the applicant)
$\hat{\sigma}_\mu^2$	$2 \cdot 10^{-3}$
$\hat{\sigma}_\varepsilon^2$	$3 \cdot 10^{-3}$
$\hat{\delta}_1, \dots, \hat{\delta}_{59}$	From $-3 \cdot 10^{-3}$ to $3 \cdot 10^{-3}$
$\hat{\beta}_0$	$2 \cdot 10^{-3}$
$\hat{\beta}_1$	$5 \cdot 10^{-6}$
$\hat{\beta}_2$	$5 \cdot 10^{-3}$
$\hat{\beta}_3$	$3 \cdot 10^{-3}$
$\hat{\beta}_4$	$4 \cdot 10^{-3}$
$\hat{\sigma}_v^2$	$2.5 \cdot 10^{-3}$
$\hat{\sigma}_\zeta^2$	$3.5 \cdot 10^{-3}$

Table 4.2. Assumed estimates

The dynamic affordability assessment has been performed using Microsoft Visual Basic for Applications (VBA). When assessing affordability, several variants of assumptions have been considered. In the basic variant, all the money left is saved ($p = 1$), reducing consumption is not allowed ($q = 0$) and the customer meets commitments according to the order of payments. In the other variants:

1) Only half of the money left can be saved ($p = 0.5$);
2) Saving is not allowed ($p = 0$);
3) Consumption can be reduced by up to 5% ($q = 0.05$);
4) Consumption can be reduced by up to 10% ($q = 0.1$);
5) The alternative order of payments is assumed (the credit card minimum payment is made before the loan payments).

The results obtained for different variants of assumptions have then been compared.

4.5 Results

The simulation has generated 10000 pairs of the predicted income and consumption time series that cover the two-year repayment period. In the last month of this period, the average predicted income is equal to ca. £2519, which corresponds to an annual increase of 4.65%. In the same month, the average predicted consumption is equal to ca. £1684, which corresponds to an annual increase of 5.95%. On average, the applicant's consumption is predicted to grow a bit faster than his income.

4.5.1 Probabilities of default

At the time of application, the customer's disposable income equals £2300 – £1500 = £800. For illustration purposes, possible instalment amounts ranging from £300 to £1300 (i.e. £800 ± £500) have been analysed. In practice, though, a narrower range would be sufficient. The assessed affordability for a range of reasonable amounts is illustrated in Figure 4.1. Probabilities of default for selected amounts are also presented in Table 4.3. They can be interpreted as follows. For example, in the basic variant, if the new instalment is equal to £750, the probability that the applicant will default in the repayment period is 0.0739 (ca. 7%). Unsurprisingly, if there are limits on saving, probabilities of default are higher (but only for amounts that do not exceed £817, since

being able to repay higher amounts hardly depends on savings). If consumption can be reduced, the probabilities are lower for all amounts. Allowing a reduction of up to 10% results in probabilities of default that are much lower than in the basic variant. However, this assumption may be considered a step too far.

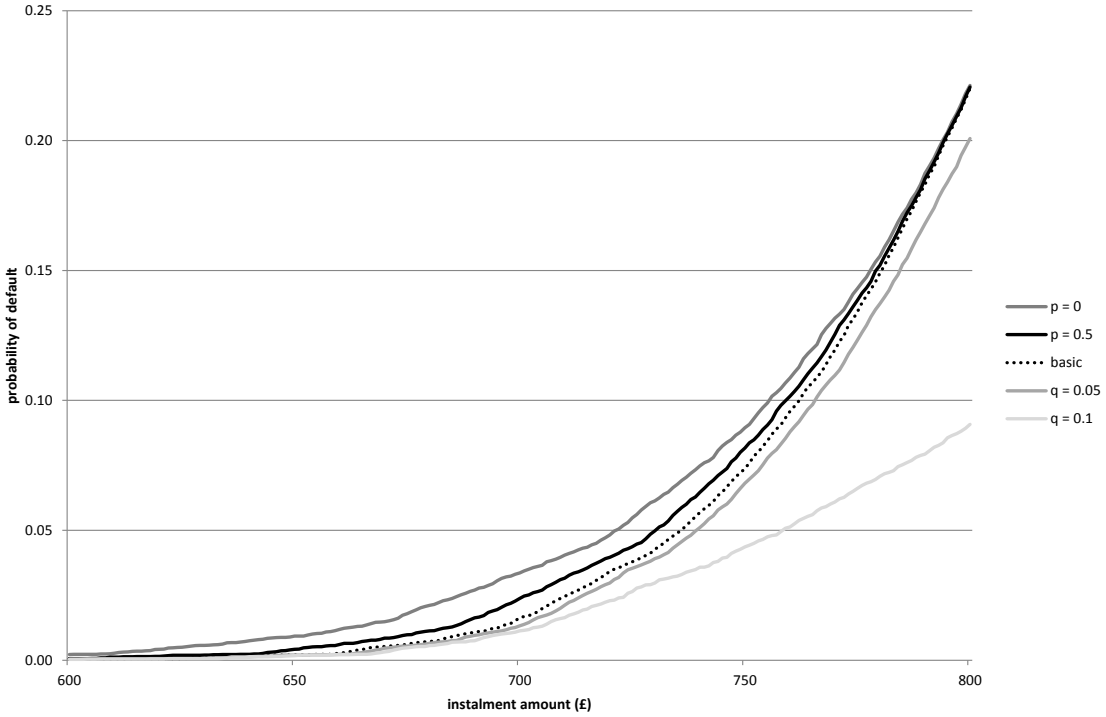

Figure 4.1. Affordability for different variants of assumptions

Instalment amount	***Basic variant***	$p = 0.5$	$p = 0$	$q = 0.05$	$q = 0.1$
£400	0.0000	0.0000	0.0000	0.0000	0.0000
£450	0.0000	0.0000	0.0001	0.0000	0.0000
£500	0.0000	0.0001	0.0001	0.0000	0.0000
£550	0.0001	0.0001	0.0006	0.0001	0.0001
£600	0.0005	0.0006	0.0022	0.0004	0.0003
£650	0.0020	0.0042	0.0093	0.0020	0.0018
£700	0.0164	0.0238	0.0336	0.0132	0.0113
£750	0.0739	0.0817	0.0895	0.0681	0.0436
£800	0.2201	0.2206	0.2212	0.2008	0.0908

Instalment amount	***Basic variant***	$\boldsymbol{p = 0.5}$	$\boldsymbol{p = 0}$	$\boldsymbol{q = 0.05}$	$\boldsymbol{q = 0.1}$
£850	0.4604	0.4604	0.4604	0.3692	0.1697
£900	0.7278	0.7278	0.7278	0.5714	0.2876
£950	0.9314	0.9314	0.9314	0.8295	0.4914
£1000	0.9959	0.9959	0.9959	0.9743	0.8022
£1050	0.9999	0.9999	0.9999	0.9992	0.9762
£1100	1.0000	1.0000	1.0000	0.9999	0.9996
£1150	1.0000	1.0000	1.0000	1.0000	0.9999
£1200	1.0000	1.0000	1.0000	1.0000	1.0000

Table 4.3. Affordability for different variants of assumptions

The results which have been obtained for the alternative order of payments are almost identical to those for the basic variant and thus are not reported here. As far as defaults are concerned, order of payments hardly makes any difference. The alternative order of payments has led to some additional defaults but only in ca. 650 out of 10000000 affordability checks (10000 simulation iterations times 1000 possible instalment amounts).

The maximum affordable instalments for several reasonable cut-offs are demonstrated in Table 4.4. In the basic variant, the maximum new instalment for which probability of default does not exceed 5% is equal to £735. When the cut-off is set to 10%, MAI equals £762. As expected, if only half or none of the money left can be saved, the amounts are lower, and if reducing consumption is allowed, they are higher. Nevertheless, for each reasonable cut-off, the results are quite similar except for those for the variant where consumption can be reduced by up to 10%. This shows that the proposed approach may be relatively robust to the assumptions.

Instead of using such cut-offs as in Table 4.4, one can take the last amount before a relatively sharp rise in probability of default (see Figure 4.1). When the threshold is 0.1% in the basic variant, MAI equals £732: within the range of reasonable amounts, every pound above £732 increases probability of default by more than 0.001 (i.e. 0.1 percentage point). One can think of this increase as marginal affordability.

Cut-off (probability of default)	*Basic variant*	$p = 0.5$	$p = 0$	$q = 0.05$	$q = 0.1$
0.05	£735	£730	£721	£739	£758
0.06	£742	£737	£728	£745	£768
0.07	£747	£743	£736	£751	£779
0.08	£752	£749	£743	£756	£790
0.09	£757	£754	£750	£761	£799
0.10	£762	£759	£755	£765	£808

Table 4.4. Maximum affordable instalments for different variants of assumptions

4.5.2 Probabilities of failure

The above analysis has linked affordability to probability of default. For comparison purposes, a similar analysis has been performed for failures instead of defaults (a failure is defined here as inability to make all obligatory payments in one or more months). Obviously, the obtained probabilities are higher and the maximum instalment amounts are lower (see Figure 4.2 as well as Tables 4.5 and 4.6). In the basic variant, if the new instalment is equal to £750, probability of failure in the repayment period is 0.0969 (ca. 10% compared to ca. 7% probability of default). When the cut-off is set to 5%, the maximum new instalment equals £722 (compared to £735). When the cut-off is 10%, the amount is equal to £751 (compared to £762). The results of this analysis for the alternative order of payments are exactly the same as those for the basic variant, since order of payments does not matter until the customer is going to fail to pay.

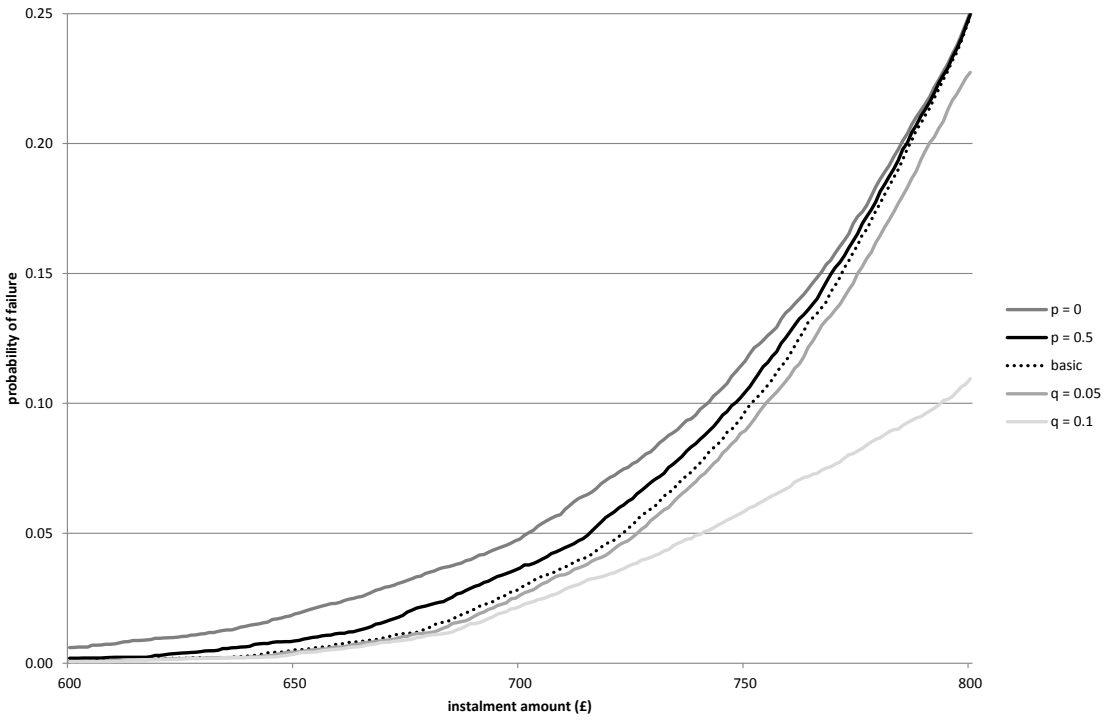

Figure 4.2. Probabilities of failure for selected amounts and different variants of assumptions

Instalment amount	***Basic variant***	$p = 0.5$	$p = 0$	$q = 0.05$	$q = 0.1$
£400	0.0000	0.0000	0.0001	0.0000	0.0000
£450	0.0000	0.0000	0.0001	0.0000	0.0000
£500	0.0000	0.0001	0.0006	0.0000	0.0000
£550	0.0001	0.0005	0.0018	0.0001	0.0001
£600	0.0006	0.0019	0.0060	0.0006	0.0006
£650	0.0051	0.0086	0.0190	0.0042	0.0035
£700	0.0288	0.0366	0.0478	0.0261	0.0218
£750	0.0969	0.1043	0.1163	0.0895	0.0587
£800	0.2487	0.2496	0.2505	0.2274	0.1095
£850	0.4779	0.4779	0.4779	0.3917	0.1924
£900	0.7311	0.7311	0.7311	0.5998	0.3090
£950	0.9341	0.9341	0.9341	0.8554	0.5294
£1000	0.9962	0.9962	0.9962	0.9851	0.8520

Instalment amount	*Basic variant*	$p = 0.5$	$p = 0$	$q = 0.05$	$q = 0.1$
£1050	0.9999	0.9999	0.9999	0.9998	0.9913
£1100	1.0000	1.0000	1.0000	1.0000	0.9998
£1150	1.0000	1.0000	1.0000	1.0000	1.0000
£1200	1.0000	1.0000	1.0000	1.0000	1.0000

Table 4.5. Probabilities of failure for selected amounts and different variants of assumptions

Cut-off (probability of failure)	*Basic variant*	$p = 0.5$	$p = 0$	$q = 0.05$	$q = 0.1$
0.05	£722	£715	£701	£726	£740
0.06	£729	£722	£710	£733	£751
0.07	£735	£729	£718	£739	£761
0.08	£741	£736	£727	£744	£772
0.09	£747	£742	£735	£750	£784
0.10	£751	£747	£741	£754	£793

Table 4.6. Maximum instalment amounts for different variants of assumptions

4.6 Discussion

In order to apply this theoretical framework in practice, one would need monthly panel data on income and expenditure at least for a few thousand consumers whose characteristics such as age etc. are known. The data, which are needed to develop the proposed models, should cover several years (ideally ca. seven years, i.e. the whole economic cycle). There may be two sources of such data: surveys and current account transactions (Thomas, 2009a; Maydon, 2011). As far as the latter is concerned, Thomas (2009a) suggested that the total value of credits can be an estimate of the consumer's income, whereas the total value of debits can be an estimate of his or her expenditure in a given month. If this is applied, the models could be said to be at the account level. Suitable panel surveys may be difficult to obtain but transaction data are available in

each bank and also for some credit bureaus such as Experian that uses Current Account Turnover (CATO) data provided by UK banks (Experian, 2011). There is no need to use data on the consumers' debts to build the models but such data from a credit report are needed later to perform affordability checks in order to assess affordability.

Understandably, there may be some challenges in applying the proposed methodology to real-life data. The total values of credits and debits may turn out to be rather biased estimates of the consumer's income and expenditure (this is more likely if the estimates are based on the bank's own transaction data, since the consumer may have two or more current accounts with different banks). Some of the characteristics which are used in the models may be unavailable (e.g. children's age) or not allowed (e.g. sex). The suggested models may not fit the data very well etc.

The models should be separately tested/validated on a hold-out sample. The whole approach could be validated by analysing only probabilities of default assigned to those instalment amounts that were actually agreed. It seems that there should be no reject inference problem, since the models are going to be developed on a sample of all customers whose income and consumption history is known (and not only those who applied for a loan and were accepted by the lender). Thus, there should be no sample bias as the models are not going to be used to make predictions about customers who are substantially different from those in the training sample. In order to assess the performance of the whole approach, the analysed probabilities can be matched with the customers' performance over the loan repayment periods (default on any loan/credit card or no default). Subsequently, calibration of the whole approach can be validated using appropriate tests and its discriminatory power can be measured e.g. with the Gini coefficient and/or the KS statistic as normally in credit scoring (Thomas, 2009a). Obviously, such performance assessment is not possible for the customers who were rejected or did not apply for any loan at all.

4.7 Conclusions

The OFT and FSA recommend taking into account dynamic aspects of affordability. In this research, a theoretical framework for dynamic affordability assessment is proposed. Income and consumption change models are suggested on the basis of the economic

literature. The models are used in a simulation to estimate affordability which is defined as a function that assigns to each possible instalment amount a probability of the applicant defaulting over the loan repayment period. This allows for the identification of the maximum affordable instalment (two identification methods are suggested). The dynamic affordability assessment is demonstrated on an example based on artificial data. The results suggest that it may be relatively robust to the assumptions on saving and reducing consumption. Interestingly, almost identical results have been obtained for different orders of payments. It is argued that the dynamic affordability assessment is in line with recommendations of the OFT and FSA. It also offers significant advantages over the static approach, such as taking a long term perspective and considering the dynamic nature of the customer's financial situation.

The proposed approach could help determine the maximum affordable loan for the applicant. In the simulation and when assessing affordability, lots of other results are produced though. One can analyse probabilities of default and failure over any time period (e.g. the first 6 or 12 months) and construct some sort of 'survival functions'. It is possible to analyse different repayment periods for different instalment amounts (so that the loan amount is constant). For a given instalment amount, one can derive distributions of loan arrears, credit card arrears, credit card balance, savings etc. in any month (there are as many predicted time series of each of these quantities as there are affordability checks for the analysed instalment amount). It is also possible to compute a distribution of EAD and even a very crude approximation of the distribution of LGD for the new loan (without taking into account the collection process and its impact on the customer's behaviour).

With appropriate samples, one could apply and test the proposed approach on real-life data. Since the output of this research is a theoretical framework, there is also plenty of room for further modifications. For example, one could use a more sophisticated version of the Euler equation by including liquidity constraints or precautionary saving. Under the assumption of liquidity constraints, optimal consumption might have been higher if the consumer had been able to borrow more. When precautionary saving is allowed, consumption can be reduced to set aside savings in the presence of uncertainty about the future.

Moreover, in the simulation, one could introduce permanent and transitory income shocks that may occur with very low probability. Income shocks can be both positive (e.g. promotion) and negative (e.g. unemployment) but, at least according to the PIH, consumption is sensitive only to permanent shocks. This could make the simulation even more realistic, although the OFT does not require taking into account the possibility of the applicant being made redundant in the future (OFT, 2011, paragraph 4.10). However, the OFT expects lenders to take into account future changes in the customer's personal circumstances such as retirement (OFT, 2011, paragraph 4.10). One could consider other changes which can affect income and/or consumption such as having (more) children. Probabilities of becoming a parent can be easily obtained for all sex and age groups, and could be incorporated into the simulation.

Finally, one could try to simultaneously estimate both the model parameters and the simulation parameters (at least p and q). Bayesian methods could be used to create a statistical emulator of such a complex model (e.g. Kennedy and O'Hagan, 2001). For this purpose, the training sample would need to be matched with data on agreed instalments, granted loans and customers' performance. The simultaneous estimation should then allow for the maximisation of discriminatory power of the whole approach. Nevertheless, the above suggestions do not exhaust the possibilities of modifying the theoretical framework for dynamic affordability assessment proposed in this research.

Chapter 5
Conclusions

5.1 Final remarks

In this chapter, the research conclusions are presented. The chapter is structured as follows. Section 5.2 is the research summary. Section 5.3 summarises the specific conclusions from the research. Section 5.4 contains recommendations for practitioners in the industry. In section 5.5, it is discussed how the research contributes to credit scoring, and in section 5.6, some suggestions of further research are presented.

5.2 Summary

This thesis addresses three selected modelling problems that are encountered in credit scoring. The research focuses on segmentation, modelling LGD (Loss Given Default) for unsecured loans and affordability assessment.

In order to assess the credit risk of bank customers, a single scoring model (scorecard) can be developed for the entire customer population, e.g. using logistic regression. However, it is often expected that segmentation, i.e. dividing the population into a number of groups and building separate scorecards for them, will improve the model performance. The most common statistical methods for segmentation are the two-step approaches, where logistic regression follows Classification and Regression Trees (CART) or Chi-square Automatic Interaction Detection (CHAID) trees. In this research, the two-step approaches are applied as well as a new, simultaneous method, in which both segmentation and scorecards are optimised at the same time: Logistic Trees with Unbiased Selection (LOTUS). For reference purposes, a single-scorecard model is used. The above-mentioned methods are applied to the data provided by two of the major UK banks and one of the European credit bureaus. The model performance measures are then compared to examine whether there is an improvement due to the segmentation methods used. It is also analysed when segmentation can bring the improvement. It is

found that segmentation does not always improve model performance in credit scoring: for none of the analysed real-world datasets do the multi-scorecard models perform considerably better than the single-scorecard ones. Besides, in this application, there is no difference in performance between the simultaneous and two-step approaches.

LGD is the loss borne by the bank when a customer defaults on a loan. For unsecured retail loans, LGD is often found difficult to model. In the frequentist (classical) two-step approach, the first model (logistic regression) is used to separate positive values from zeroes and the second model (e.g. linear regression) is applied to estimate the positive values. Those models are estimated independently, which can be considered problematic from the methodological point of view. The result is a point estimate of LGD for each loan. Alternatively, LGD can be modelled using Bayesian methods, since they are especially suitable for the estimation of hierarchical models. In the Bayesian framework, one can build a single, hierarchical model instead of two separate ones, which makes this a more coherent approach. In this research, Bayesian methods as well as the frequentist approach are applied to the data on personal loans provided by a large UK bank. As expected, the posterior means of parameters which are produced using Bayesian methods are very similar to the corresponding frequentist estimates. The most important advantage of the Bayesian model is that it generates an individual predictive distribution of LGD for each loan rather than just a point estimate. Potential applications of the predictive distributions include approximating the downturn LGD and stress testing LGD under Basel II.

Whereas credit scoring focuses mainly on creditworthiness (propensity to repay a loan), affordability (ability to repay) is often checked on the basis of current income and estimated current consumption as well as existing debts stated in a credit report. Contrary to that static approach, a theoretical framework for dynamic affordability assessment is proposed in this research. In this approach, both income and consumption are allowed to vary over time and their changes are described with random effects models for panel data. The models are derived from the economic literature, including the Euler equation of consumption. On their basis a simulation is run and predicted time series are generated for a given applicant. For each pair of the predicted income and consumption time series, the applicant's ability to repay is checked over the life of the

loan and for all possible instalment amounts. As a result, a probability of default is assigned to each amount, which can help find the maximum affordable instalment. This is illustrated with an example based on artificial data. Assessing affordability over the loan repayment period as well as taking into account variability of income and expenditure over time are in line with recommendations of the Office of Fair Trading (OFT) and the Financial Services Authority (FSA). In practice, the suggested approach could contribute to responsible lending.

5.3 Specific conclusions

The specific conclusions from the research are summarised below. The conclusions are presented in more detail in sections 2.7, 3.7 and 4.7.

As far as segmentation is concerned, the most important finding is that segmentation does not always improve model performance in credit scoring, since for none of the analysed datasets do the multi-scorecard models perform considerably better than the logistic regression. Moreover, no difference in performance has been observed between the two-step and simultaneous approaches. For a large sample with strong characteristics, all the models, including the logistic regression, have the same separating ability and are equally stable. It has been noticed that the segmentation contribution can be up to 20 percentage points, which means that segmentation itself can be a powerful tool, but it seems to leave little space for the scorecards to further discriminate customers. Finally, one can show an example of a situation in which segmentation improves the model performance and the simultaneous approach outperforms the two-step approaches on an artificial dataset. However, such a situation seems rather unusual in banking practice.

With regard to modelling LGD, the posterior means of the parameters which have been yielded in the Bayesian framework are very similar to the frequentist estimates. The posterior means and standard deviations of the model performance measures are also almost the same as the corresponding bootstrap estimates generated in the frequentist random cut-off approach. In comparison with the random cut-off approach, the probability times value approach has produced slightly better posterior means of the performance measures. These posterior means are almost identical as the results

obtained in the frequentist probability times value approach. Although the performance of both approaches is similar, the Bayesian model is more coherent than the frequentist one and allows for a better description of uncertainty. Besides, it generates an individual predictive distribution of LGD for each loan, whereas the frequentist approach only produces a point estimate. Such distributions provide a lot of information and can be used for various purposes.

As far as the dynamic affordability assessment is concerned, it is argued that the proposed approach is in line with recommendations of the OFT and FSA. Furthermore, it has been assumed that if there is any money left after all commitments are met, it can be saved and used later. It has also been assumed that if the customer cannot make all obligatory payments, they may reduce their consumption. In the example based on artificial data, several variants of the above-mentioned assumptions have been considered. The obtained results suggest that the dynamic affordability assessment may be relatively robust to these assumptions. Remarkably, almost the same results have been yielded for different orders of payments (the loan payments before the credit card minimum payment or the other way round). The proposed approach could help identify the maximum affordable loan. Nevertheless, in the simulation and when assessing affordability, lots of other results are also produced.

5.4 Recommendations for practitioners

On the basis of this research, the following recommendations can be formulated for practitioners in the industry.

When building a multi-scorecard model, it is advisable to develop a single-scorecard one for comparison purposes. In banking practice it is common not to make such comparisons, as there is a strong belief that segmentation allows for the model performance improvement. However, this research shows that the expected improvement may not occur. If there is no improvement, it is sensible to choose a single-scorecard model, since any additional costs generated by a multi-scorecard model should be compensated for by better risk assessment. Ultimately, when the segmentation goal is related to the model performance, maintaining a number of scorecards which perform like a single one is a waste of resources.

When modelling LGD, it is recommended to use Bayesian methods to produce an individual predictive distribution for each loan. The predictive distributions can be treated as the benchmarks for the LGD estimates. Their selected quantiles can be used to approximate the downturn LGD when downturn data are lacking, and can also serve as the stressed LGD. Furthermore, the predictive distributions can help diversify collection strategies in order to improve the work-out process. They can even be used to set a cut-off for the score used to accept and reject applicants. With so many possible applications, using Bayesian methods to model LGD seems to be worth the effort.

When assessing affordability, it is advisable to apply a dynamic approach such as the one proposed in this research. Employing panel data techniques is a well-known way of introducing dynamics into models. Random effects models seem an appropriate choice here, since they can be used outside the training sample. With regard to the variable selection, there is a rich economic literature on modelling income and consumption that can be a good source of inspiration. In order to take into account the random components and unknown future macroeconomic conditions, a simulation can be run. Importantly, taking a long term perspective and considering the dynamic nature of the customer's financial situation are also recommended by the OFT and FSA.

Generally, it is worth to challenge the established approaches, be it the segmentation-based, frequentist or static ones. Using new, more sophisticated methods may provide more information and a fuller picture of what is analysed. For example, Bayesian methods can give a better description of uncertainty associated with the estimation of LGD, whereas random effects models combined with a simulation can give an insight into an applicant's future ability to repay. Applying new methods may also facilitate meeting the regulator's recommendations and requirements, such as those related to affordability assessment or to the downturn and stressed LGD. Moreover, sometimes using simpler solutions than the established ones may help save resources without compromising on performance (after all, it is sensible to follow the Occam's razor principle). Examples of such solutions include single scorecards that perform like multi-scorecard models.

Finally, one should face the encountered modelling problems. Ignoring dynamics (as in the static approach) or downplaying uncertainty (as in the frequentist approach) is not the recommended solution. Nor is sticking to the established, segmentation-based approach, where the effects of heterogeneity on the model performance have not been analysed. Instead, it is advisable to look for methods that enable handling the encountered problems effectively. This relates not only to heterogeneity, uncertainty and dynamics but also to other modelling challenges that occur in credit scoring.

5.5 Contribution to credit scoring

As far as it can be ascertained, this research makes the following contribution to the credit scoring literature and knowledge. Understandably, the statements below relate to what is available in the public domain.

This research is the first one where different statistical methods of segmentation are compared and their contribution to the model performance is assessed. It is also one of the very few published works in which large credit bureau datasets are analysed in the context of segmentation. In this work, the LOTUS algorithm is used in credit scoring for the first time. Moreover, it is the first research where Bayesian methods are used to model LGD for retail loans. A number of possible applications of the resulting predictive distributions of LGD are suggested here. Furthermore, unlike previous studies, this research proposes a dynamic approach to affordability assessment and presents a complete theoretical framework for it. Panel data techniques and models derived from the economic literature (including the Euler equation of consumption) are proposed here to be used in affordability modelling. Finally, this work contributes to the sparse literature on segmentation methods and on statistical models for affordability assessment.

5.6 Further research suggestions

A number of specific suggestions of possible modifications and further analysis are presented in sections 2.7, 3.7 and 4.7. Among other ideas, they include employing other simultaneous approaches (e.g. LMT) to perform segmentation, applying Bayesian model selection to find the best covariates of the LGD model as well as using a more

sophisticated version of the Euler equation (e.g. the one with liquidity constraints) in affordability modelling.

This research addresses heterogeneity, uncertainty and dynamics. Another common modelling problem is combining micro- and macrolevel analysis. This problem is often encountered in social sciences and economics, where it is tackled in various ways, e.g. by introducing microfoundations into macroeconometric models (see section 4.3.2). In credit scoring, it occurs – among other areas – in portfolio PD modelling. Various approaches were proposed to estimate PD at the portfolio level (see section 1.2.2). Alternatively, a multilevel model could be used, where individual loans are at the first level and loan portfolios are at the second (top) level. Such a model could be applied to assess credit risk at both levels. This approach could enable taking into account all sorts of similarities and dependencies, including default correlations. The multilevel model could be developed using Bayesian methods since, as noted e.g. by Courgeau (2012), they allow for an effective multilevel analysis that is relatively easy to perform.

Appendices

Appendix A. Customer's characteristics

Dataset A_1	***Dataset A_2***
Age	**Age**[a]
Marital Status	Marital Status
Residential Status	Number of Children
MOSAIC Classification	Residential Status
Time at Current Address	**Time at Current Address**
Time at Previous Address	Home Phone
Home Phone	**Time with Current Employer**
Occupation	Gross Income
Time with Current Employer	**FiNPiN Classification**
Time with Previous Employer	Loan Type
Net Income	**Loan Amount**
Pension Scheme	**Loan Purpose**
Time With Bank	**Insurance**
Number of Credit Cards	**Payment Frequency**
Amex / Diners Card Holder	Number of Searches for Exact Name (Current Address)
Loan Amount	Time since Last CCJ for Exact Name (Current Address)
Loan Term	Number of Write-offs for Exact Name (Current Address)
Loan Purpose	Time since Last CCJ for Similar Name (Current Address)
Total Cost of Goods	Number of Write-offs for the Same Surname (Current Address)
Insurance	Number of Bad Events for the Same Surname (Current Address)
Payment Frequency	Number of Bad Events at the Postal Code (Current Address)

Appendix A

Dataset A_1	***Dataset A_2***
Payment Method	Number of Bad Events Which Have Turned Good at the Postal Code (Current Address)
Number of Searches in the Last 6 Months	**Percentage of Bad Events Which Have Turned Good at the Postal Code (Current Address)**
Value of CAIS (Bad Debts, Same Surname, Other Initial, Current and Previous Address)	**Number of Dormant Events at the Postal Code (Current Address)**
Value of CAIS (Bad Debts, Same Surname, Same Initial, Current and Previous Address)	Electoral Roll Status for the Same Surname (Current Address)
Value of CCJ (Same Surname, Other Initial, Current and Previous Address)	Time on Electoral Roll (Current Address)
Value of CCJ (Same Surname, Same Initial, Current and Previous Address)	Number of Searches for Exact Name (Previous Address)
Time since Most Recent CAIS (Bad Debt, Same Surname, Other Initial, Current and Previous Address)	Time since Last CCJ for Exact Name (Previous Address)
Time since Most Recent CAIS (Bad Debt, Same Surname, Same Initial, Current and Previous Address)	Number of Write-offs for Exact Name (Previous Address)
Time since Most Recent CCJ (Same Surname, Other Initial, Current and Previous Address)	Time since Last CCJ for Similar Name (Previous Address)
Time since Most Recent CCJ (Same Surname, Same Initial, Current and Previous Address)	Number of Write-offs for the Same Surname (Previous Address)
Number of CAIS (Bad Debts, Same Surname, Other Initial, Current and Previous Address)	Number of Bad Events for the Same Surname (Previous Address)

Dataset A_1	***Dataset A_2***
Number of CAIS (Bad Debts, Same Surname, Same Initial, Current and Previous Address)	Number of Bad Events at the Postal Code (Previous Address)
Number of CCJ (Same Surname, Other Initial, Current and Previous Address)	Number of Bad Events Which Turned Good at the Postal Code (Previous Address)
Number of CCJ (Same Surname, Same Initial, Current and Previous Address)	Percentage of Bad Events Which Have Turned Good at the Postal Code (Previous Address)
	Number of Dormant Events at the Postal Code (Previous Address)
	Electoral Roll Status for the Same Surname (Previous Address)
	Time on Electoral Roll (Previous Address)

[a] The characteristics which have been used in the reference logistic regression models are marked with a bold font.

Appendix B. OpenBUGS code

```
model {

# priors
for (k in 1:NK) { beta1[k] ~ dnorm(0, 0.01)}     # N(0, 10^2)
for (l in 1:NL) { beta2[l] ~ dnorm(0, 0.01)}     # N(0, 10^2)
c1 ~ dnorm(0, 0.0001)     # N(0, 100^2)
c2 ~ dnorm(0, 0.0001)     # N(0, 100^2)
tau1 ~ dgamma(10, 0.00001)     # E(tau1) = 10^6; Var(tau1) = 10^11
tau2 ~ dgamma(0.01, 0.01)      # E(tau2) = 1; Var(tau2) = 100

tau[1] <- tau1
tau[2] <- tau2

# model
mdy <- mean(dy[])
mvy <- mean(vy[])

for (i in 1:N) {

# training
     dp[i] <- 1/(1 + exp(-(c1+inprod(dx[i,],beta1[]))))

     db[i] ~ dbern(dp[i])
     d.index[i] <- db[i] + 1

     d.mu[i,1] <- 0
     d.mu[i,2] <- c2+inprod(dz[i,],beta2[])

     dy[i] ~ dnorm(d.mu[i,d.index[i]],tau[d.index[i]])

     db.new[i] ~ dbern(dp[i])
     d.index.new[i] <- db.new[i] + 1

     d.y[i] ~ dnorm(d.mu[i,d.index.new[i]],tau[d.index.new[i]])

     de[i] <- dy[i] - mdy
     d.esqr[i] <- de[i]*de[i]

     d.mu2[i] <- d.mu[i,d.index.new[i]]

     de2[i] <- dy[i] - d.mu2[i]

     d.eabs2[i] <- abs(de2[i])
     d.esqr2[i] <- de2[i]*de2[i]
     d.cov2[i] <- d.mu2[i]*de[i]

     d.mu3[i] <- dp[i]*d.mu[i,2]

     de3[i] <- dy[i] - d.mu3[i]

     d.eabs3[i] <- abs(de3[i])
     d.esqr3[i] <- de3[i]*de3[i]
     d.cov3[i] <- d.mu3[i]*de[i]
```

Appendix B

```
# validation
     vp[i] <- 1/(1 + exp(-(c1+inprod(vx[i,],beta1[]))))

     vb[i] ~ dbern(vp[i])
     v.index[i] <- vb[i] + 1

     v.mu[i,1] <- 0
     v.mu[i,2] <- c2+inprod(vz[i,],beta2[])

     v.y[i] ~ dnorm(v.mu[i,v.index[i]],tau[v.index[i]])

     ve[i] <- vy[i] - mvy
     v.esqr[i] <- ve[i]*ve[i]

     v.mu2[i] <- v.mu[i,v.index[i]]

     ve2[i] <- vy[i] - v.mu2[i]

     v.eabs2[i] <- abs(ve2[i])
     v.esqr2[i] <- ve2[i]*ve2[i]
     v.cov2[i] <- v.mu2[i]*ve[i]

     v.mu3[i] <- vp[i]*v.mu[i,2]

     ve3[i] <- vy[i] - v.mu3[i]

     v.eabs3[i] <- abs(ve3[i])
     v.esqr3[i] <- ve3[i]*ve3[i]
     v.cov3[i] <- v.mu3[i]*ve[i]

}

d.MAE <- mean(d.eabs2[])
d.MSE <- mean(d.esqr2[])
d.R2 <- 1 - sum(d.esqr2[])/sum(d.esqr[])
d.corr <- (mean(d.cov2[]) -
mean(d.mu2[])*mean(de[]))/(sd(d.mu2[])*sd(dy[]))

d.MAEx <- mean(d.eabs3[])
d.MSEx <- mean(d.esqr3[])
d.R2x <- 1 - sum(d.esqr3[])/sum(d.esqr[])
d.corrx <- (mean(d.cov3[]) -
mean(d.mu3[])*mean(de[]))/(sd(d.mu3[])*sd(dy[]))

v.MAE <- mean(v.eabs2[])
v.MSE <- mean(v.esqr2[])
v.R2 <- 1 - sum(v.esqr2[])/sum(v.esqr[])
v.corr <- (mean(v.cov2[]) -
mean(v.mu2[])*mean(ve[]))/(sd(v.mu2[])*sd(vy[]))

v.MAEx <- mean(v.eabs3[])
v.MSEx <- mean(v.esqr3[])
v.R2x <- 1 - sum(v.esqr3[])/sum(v.esqr[])
v.corrx <- (mean(v.cov3[]) -
mean(v.mu3[])*mean(ve[]))/(sd(v.mu3[])*sd(vy[]))

}
```

```
DATA
list( N=10000, NK=5, NL=5)
dy[] dx[,1] dx[,2] dx[,3] dx[,4] dx[,5] dz[,1] dz[,2] dz[,3] dz[,4]
dz[,5]
0.4777351986 0.3357063595 -0.160489597 -0.412955391 0.6865201433
0.9335050953 -0.594875968 -0.412955391 -0.774108861 -0.321010175
1.0018564243
1.1472599615 -1.120848278 0.5166218924 0.5832072493 -0.82241393 -
0.8227443 -0.594875968 0.5832072493 0.846297986 0.6285079765 -
0.85501503
0 -0.317453553 0.9680295521 0.5832072493 -0.536939916 -0.237327835
1.6808545892 0.5832072493 0.846297986 0.6285079765 -0.297953594

…

0.6444750167 2.7888721548 1.6451410417 -2.865048045 3.0518762588
0.9335050953 -0.594875968 -2.865048045 0.846297986 -2.658285625
1.0018564243
END


vy[] vx[,1] vx[,2] vx[,3] vx[,4] vx[,5] vz[,1] vz[,2] vz[,3] vz[,4]
vz[,5]
0.5479082402 0.1562996699 -0.844468369 -1.924481418 0.5663700777
0.932749908 1.6948980488 -1.924481418 0.8471034967 -1.757235919
1.0016627365
1.0857298738 -0.954832823 -1.526925913 0.839125949 -0.829965352 -
1.416154418 1.6948980488 0.839125949 -1.592710722 0.8737281249 -
1.233484787
0.8320755762 -0.777572023 -0.389496673 0.9180861594 -0.994240108 -
1.416154418 -0.589946989 0.9180861594 0.8471034967 0.9488985262 -
1.233484787

…

0.9667534889 -1.06216505 -1.071954217 0.7601657385 -0.788896663 -
0.828928337 -0.589946989 0.7601657385 0.0338320905 0.7985577237 -
1.233484787
END


INITS
list( beta1=c(0,0,0,0,0), beta2=c(0,0,0,0,0), c1=0, c2=0,
tau1=1000000, tau2=1)
```

References

Abid Ali, M., Hickman, P.J. and Clementson, A.T. (1975) The Application of Automatic Interaction Detection (AID) in Operational Research, *Operational Research Quarterly*, 26(2), Part 1, pp. 243-252.

Altman, E., Resti, A. and Sironi, A. (2005) *Recovery Risk*, London: Risk Books.

Anderson, R. (2007) *The Credit Scoring Toolkit*, New York: Oxford University Press.

Andreeva, G., Ansell, J. and Crook, J.N. (2005) Modelling the Purchase Propensity: Analysis of a Revolving Store Card, *Journal of the Operational Research Society*, 56(9), pp. 1041-1050.

Armitage, P. (1955) Tests for Linear Trends in Proportions and Frequencies, *Biometrics*, 11(3), pp. 375-386.

Arsova, A., Haralampieva, M. and Tsvetanova, T. (2011) *Comparison of regression models for LGD estimation*, Credit Scoring and Credit Control XII, Edinburgh.

Association for Payment Clearing Services, British Bankers' Association, Building Societies Association *et al*. (2000) *Guide to Credit Scoring*, London: Association for Payment Clearing Services, British Bankers' Association, Building Societies Association *et al*.

Association of Consumer Credit Information Suppliers and European Credit Research Institute (2011) *The European Credit Information Landscape: An analysis of a survey of credit bureaus in Europe*, Brussels: Association of Consumer Credit Information Suppliers and European Credit Research Institute.

Aurifeille, J.M. (2000) A bio-mimetic approach to marketing segmentation: Principles and comparative analysis, *European Journal of Economic and Social Systems*, 14(1), pp. 93-108.

References

Auten, G. and Carroll, R. (1999) The effect of income taxes on household income, *The Review of Economics and Statistics*, 81(4), pp. 681-693.

Baesens, B., van Gestel, T., Viaene, S., Stepanova, M., Suykens, J. and Vanthienen, J. (2003) Benchmarking state-of-the-art classification algorithms for credit scoring, *Journal of the Operational Research Society*, 54(6), pp. 627-635.

Balestra, P. and Nerlove, M. (1966) Pooling Cross Section and Time Series Data in the Estimation of a Dynamic Model: The Demand for Natural Gas, *Econometrica*, 34(3), pp. 585-612.

Banasik, J., Crook, J.N. and Thomas, L.C. (1996) Does scoring a subpopulation make a difference?, *The International Review of Retail, Distribution and Consumer Research*, 6(2), pp. 180-195.

Banasik, J., Crook, J.N. and Thomas, L.C. (1999) Not if but when will borrowers default, *Journal of the Operational Research Society*, 50(12), pp. 1185-1190.

Basel Committee on Banking Supervision (2005a) *Guidance on Paragraph 468 of the Framework Document*, Basel: Bank for International Settlements.

Basel Committee on Banking Supervision (2005b) *Studies on the Validation of Internal Rating Systems (revised)*, Basel Committee on Banking Supervision Working Paper No. 14.

Basel Committee on Banking Supervision (2006) *Basel II: International Convergence of Capital Measurement and Capital Standards. A Revised Framework*, Comprehensive version, Basel: Bank for International Settlements.

Basel Committee on Banking Supervision (2011) *Basel III: A global regulatory framework for more resilient banks and banking systems*, Revised version, Basel: Bank for International Settlements.

Bellotti, T. and Crook, J. (2007) *Incorporating macroeconomic variables into consumer credit analysis*, Symposium on Risk Management in the Retail Financial Services Sector, London.

Bellotti, T. and Crook, J. (2008) *Modelling and estimating Loss Given Default for credit cards*, University of Edinburgh Business School, Credit Research Centre Working Paper 08-1.

Bellotti, T. and Crook, J. (2012) Loss given default models incorporating macroeconomic variables for credit cards, *International Journal of Forecasting*, 28(1), pp. 171-182.

Benoit, D.F. and van den Poel, D. (2009) Benefits of quantile regression for the analysis of customer lifetime value in a contractual setting: An application in financial services, *Expert Systems with Applications*, 36(7), pp. 10475-10484.

Berk, R.A. (2008) *Statistical Learning from a Regression Perspective*, New York: Springer.

Bernardo, J.M. and Smith, A.F.M. (2003) *Bayesian Theory*, Chichester: Wiley.

Betti, G., Dourmashkin, N., Rossi, M.C., Verma, V. and Yin, Y. (2001) *Study of the problem of Consumer Indebtedness: Statistical Aspects*, Report to the Commission of the European Communities Directorate-General for Health and Consumer Protection.

Bi, J. and Bennett, K.P. (2003) Regression Error Characteristic Curves, In: Fawcett, T. and Mishra, N. (eds.) *Proceedings of the Twentieth International Conference on Machine Learning*, Menlo Park, CA: AAAI Press, pp. 43-50.

Bijak, K. (2011) Kalman filtering as a performance monitoring technique for a propensity scorecard, *Journal of the Operational Research Society*, 62(1), pp. 29-37.

Böcker, K. (ed.) (2010) *Rethinking Risk Measurement and Reporting*, Volumes I and II, London: Risk Books.

Bolton, R.J. and Hand, D.J. (2002) Statistical Fraud Detection: A Review, *Statistical Science*, 17(3), pp. 235-255.

Breeden, J.L., Thomas, L. and McDonald, J.W. (2008) Stress-testing retail loan portfolios with dual-time dynamics, *The Journal of Risk Model Validation*, 2(2), pp. 43-62.

References

Breiman, L., Friedman, J.H., Olshen, R.A. and Stone, C.J. (1998) *Classification and Regression Trees*, Boca Raton, FL: Chapman and Hall/CRC.

British Bankers' Association, Building Societies Association and The UK Cards Association (2011) *The Lending Code: Setting standards for banks, building societies and credit card providers*, London: British Bankers' Association, Building Societies Association and The UK Cards Association.

Brooks, S.P. (1998) Markov Chain Monte Carlo Method and Its Application, *Journal of the Royal Statistical Society: Series D (The Statistician)*, 47(1), pp. 69-100.

Brooksby, B. (2009) *Measuring Affordability at Origination*, Credit Scoring and Credit Control XI, Edinburgh.

Brown, I. (2011) *Regression Model Development for Credit Card Exposure at Default (EAD) using SAS/STAT and SAS Enterprise Miner 5.3*, SAS Global Forum 2011, Las Vegas, NV.

Bryan, M., Taylor, M. and Veliziotis, M. (2010) *Over-indebtedness in Great Britain: An analysis using the Wealth and Assets Survey and Household Annual Debtors Survey*, Report to the Department for Business, Innovation and Skills.

Burez, J. and van den Poel, D. (2008) Separating financial from commercial customer churn: A modeling step towards resolving the conflict between the sales and credit department, *Expert Systems with Applications*, 35(1-2), pp. 497-514.

Callcredit (n.d.) *The Affordability Suite – calculate your customer's Risk and Affordability*, Available at: http://www.callcredit.co.uk/products-and-services/credit-risk-and-affordability/the-affordability-suite [Accessed: 15 November 2012].

Caselli, S., Gatti, S. and Querci, F. (2008) The Sensitivity of the Loss Given Default Rate to Systematic Risk: New Empirical Evidence on Bank Loans, *Journal of Financial Services Research*, 34(1), pp. 1-34.

Chan, K.-Y. (2005) *LOTUS User Manual (version 2.2)*, Available at: http://www.stat.wisc.edu/~loh/lotus/Lotus/manual.pdf [Accessed: 30 April 2013].

Chan, K.-Y. and Loh, W.-Y. (2004) LOTUS: An algorithm for building accurate and comprehensible logistic regression trees, *Journal of Computational and Graphical Statistics*, 13(4), pp. 826-852.

Chen, G. and Åstebro, T. (2003) Bound and Collapse Bayesian Reject Inference When Data are Missing not at Random, In: Åstebro, T., Beling, P., Hand, D., Oliver, B. and Thomas, L.B. (eds.) *Mathematical Approaches to Credit Risk Management: Conference Proceedings*, Banff, Alberta: Banff International Research Station for Mathematical Innovation and Discovery.

Cochran, W.G. (1954) Some Methods for Strengthening the Common χ^2 Tests, *Biometrics*, 10(4), pp. 417-451.

Congdon, P. (2004) *Applied Bayesian Modelling*, Chichester: Wiley.

Consumer Credit Act 2006 (c. 14), London: TSO.

Council Directive 2006/48/EC of 14 June 2006 relating to the taking up and pursuit of the business of credit institutions (recast).

Council Directive 2006/49/EC of 14 June 2006 on the capital adequacy of investment firms and credit institutions (recast).

Council Directive 2008/48/EC of 23 April 2008 on credit agreements for consumers and repealing Council Directive 87/102/EEC.

Courgeau, D. (2012) *Probability and Social Science: Methodological Relationships between the two Approaches*, Dordrecht: Springer.

Crook, J. (2012) *Random Effects and Macroeconomic Variables in Credit Risk Models*, Model Risk in Retail Credit Scoring – Statistical Issues, London.

Crook, J. and Bellotti, T. (2008) *Dynamic Consumer Risk Models: An Overview*, Dynamic Consumer Risk Modelling and the Economy, Edinburgh.

Crowder, M., Hand, D.J. and Krzanowski, W.J. (2005) *On customer lifetime value*, Credit Scoring and Credit Control IX, Edinburgh.

References

Davenport, T.H. and Harris, J.G. (2007) *Competing on Analytics: The New Science of Winning*, Boston, MA: Harvard Business School Press.

De Andrade, F.W.M. and Thomas, L. (2007) Structural models in consumer credit, *European Journal of Operational Research*, 183(3), pp. 1569-1581.

Dell, D.D. (2007) *Making Affordability work for you!*, Credit Scoring and Credit Control X, Edinburgh.

Department of Trade and Industry (2005) *Over-indebtedness in Britain*, DTI report on the MORI Financial Services survey 2004, London: Department of Trade and Industry.

Desmet, P. (2001) Buying behavior study with basket analysis: pre-clustering with a Kohonen map, *European Journal of Economic and Social Systems*, 15(2), pp. 17-30.

Disney, R., Bridges, S. and Gathergood, J. (2008) *Drivers of Overindebtedness*, Report to the Department for Business, Enterprise and Regulatory Reform.

Dodd-Frank Wall Street Reform and Consumer Protection Act (2010), Washington, DC: GPO.

Donkers, B., Verhoef, P.C. and de Jong, M.G. (2007) Modeling CLV: A test of competing models in the insurance industry, *Quantitative Marketing and Economics*, 5(2), pp. 163-190.

Draper, D. (1995) Assessment and Propagation of Model Uncertainty, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 57(1), pp. 45-97.

Dwyer, D.W. (2007) The distribution of defaults and Bayesian model validation, *The Journal of Risk Model Validation*, 1(1), pp. 23-53.

Edelman, D. (2003) *Risk Based Pricing for Personal Loans: How it works and how to make it work*, Credit Scoring and Credit Control VIII, Edinburgh.

Efron, B. (2003) Second Thoughts on the Bootstrap, *Statistical Science*, 18(2), pp. 135-140.

Etienne, J.-M. (2006) *Health and Socio-Economic Status: Is it Measured Income or Permanent Income that Matters?*, XX Annual Conference of the European Society for Population Economics, Verona.

European Banking Authority (2010) *CEBS Guidelines on Stress Testing (GL32)*, Available at: http://www.eba.europa.eu/documents/Publications/Standards---Guidelines/2010/Stress-testing-guidelines/ST_Guidelines.aspx [Accessed: 8 March 2013].

European Banking Authority (n.d.) *Electronic Guidebook*, Available at: http://www.eba.europa.eu/Publications/Compendium-of-guidelines.aspx [Accessed: 9 October 2011].

European Commission's Expert Group on Credit Histories (2009) *Report of the Expert Group on Credit Histories*, Brussels: European Commission Directorate-General Internal Market and Services.

Experian (2011) *The affordability challenge*, Experian Briefing Paper.

Fernandes, G. and Rocha, C.A. (2011) *Low default modelling: a comparison of techniques based on a real Brazilian corporate portfolio*, Credit Scoring and Credit Control XII, Edinburgh.

Financial Services Authority (2009) *Mortgage Market Review*, Discussion Paper 09/3.

Financial Services Authority (2010) *Mortgage Market Review: Responsible Lending*, Consultation Paper 10/16.

Finlay, S.M. (2006) Predictive Models of Expenditure and Over-Indebtedness for Assessing the Affordability of New Consumer Credit Applications, *Journal of the Operational Research Society*, 57(6), pp. 655-669.

Fondeville, N., Őzdemir, E. and Ward, T. (2010) *Over-indebtedness: New evidence from the EU-SILC special module*, Research note 4/2010 for the European Commission Directorate-General for Employment, Social Affairs and Equal Opportunities.

References

Friedman, M. (1957) *A Theory of the Consumption Function*, Princeton, NJ: Princeton University Press.

Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B. (2004) *Bayesian Data Analysis*, Boca Raton, FL: Chapman and Hall/CRC.

Giudici, P. (2001) Bayesian data mining with application to benchmarking and credit scoring, *Applied Stochastic Models in Business and Industry*, 17(1), pp. 69-81.

Greene, W.H. (2000) *Econometric Analysis*, 4th ed., Upper Saddle River, NJ: Prentice Hall.

Guiso, L., Jappelli, T. and Terlizzese, D. (1992) Earnings uncertainty and precautionary saving, *Journal of Monetary Economics*, 30(2), pp. 307-337.

Gupton, G.M. and Stein, R.M. (2005) *LossCalc v2: Dynamic prediction of LGD*, Moody's KMV Research Paper.

Hall, R.E. (1978) Stochastic Implications of the Life Cycle-Permanent Income Hypothesis: Theory and Evidence, *Journal of Political Economy*, 86(6), pp. 971-987.

Hall, R.E. and Mishkin, F.S. (1982) The Sensitivity of Consumption to Transitory Income: Estimates from Panel Data on Households, *Econometrica*, 50(2), pp. 461-481.

Hand, D.J. (2009) Measuring classifier performance: a coherent alternative to the area under the ROC curve, *Machine Learning*, 77(1), pp. 103-123.

Hand, D.J. and Henley, W.E. (1993) Can reject inference ever work?, *IMA Journal of Mathematics Applied in Business and Industry*, 5(1), pp. 45-55.

Hand, D.J., Mannila, H. and Smyth, P. (2001) *Principles of Data Mining*, Cambridge, MA: The MIT Press.

Hand, D.J., Sohn, S.Y. and Kim, Y. (2005) Optimal bipartite scorecards, *Expert Systems with Applications*, 29(3), pp. 684-690.

Hao, L. and Naiman, D.Q. (2007) *Quantile Regression*, Thousand Oaks, CA: Sage Publications.

Hawkins, D.M. and Kass, G.V. (1982) Automatic Interaction Detection, In: Hawkins, D.M. (ed.) *Topics in Applied Multivariate Analysis*, Cambridge: Cambridge University Press, pp. 269-302.

Heckman, J.J. (1981) Heterogeneity and State Dependence, In: Rosen, S. (ed.) *Studies in Labor Markets*, Chicago, IL: University of Chicago Press, pp. 91-139.

Hlawatsch, S. and Ostrowski, S. (2011) Simulation and estimation of loss given default, *The Journal of Credit Risk*, 7(3), pp. 39-73.

Hosmer, D.W. and Lemeshow, S. (2000) *Applied Logistic Regression*, 2nd ed., New York: Wiley.

Huang, B. and Thomas, L.C. (2009) *Credit Card Pricing and Impact of Adverse Selection*, Credit Scoring and Credit Control XI, Edinburgh.

Huang, C. and Litzenberger, R.H. (1988) *Foundations for Financial Economics*, New York: North-Holland.

Izenman, A.J. (2008) *Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning*, New York: Springer.

Jappelli, T. and Pistaferri, L. (2000) Using subjective income expectations to test for excess sensitivity of consumption to predicted income growth, *European Economic Review*, 44(2), pp. 337-358.

Jappelli, T. and Pistaferri, L. (2006) Intertemporal Choice and Consumption Mobility, *Journal of the European Economic Association*, 4(1), pp. 75-115.

Jaynes, E.T. (1976) Confidence Intervals vs Bayesian Intervals, In: Harper, W.L. and Hooker, C.A. (eds.) *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science*, Dordrecht: D. Reidel, pp. 175-257.

Jiménez, G., Lopez, J.A. and Saurina, J. (2009) *EAD Calibration for Corporate Credit Lines*, Federal Reserve Bank of San Francisco Working Paper 2009-2.

References

Kass, G.V. (1980) An Exploratory Technique for Investigating Large Quantities of Categorical Data, *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 29(2), pp. 119-127.

Kempson, E. (2002) *Over-indebtedness in Britain*, Report to the Department of Trade and Industry.

Kennedy, M. and O'Hagan, T. (2001) Bayesian Calibration of Computer Models, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(3), pp. 425-464.

Kiefer, N.M. (2009) Default estimation for low-default portfolios, *Journal of Empirical Finance*, 16(1), pp. 164-173.

Kim, M.-J. (2006) Downturn LGD, Best Estimate of Expected Loss, and Potential LGD under Basel II, *Journal of Economic Research*, 11(2), pp. 203-223.

Konstantinos, S., Dimitrios, V. and Georgios, A. (2003) *Risk-Based Pricing (RBP) Using Bayesian Statistics: How to Market RBP in the Context of New Credit Card Customers*, Credit Scoring and Credit Control VIII, Edinburgh.

Koskinen, L., Nummi, T. and Salonen, J. (2007) *Modeling and Predicting Individual Salaries: A Study of Finland's Unique Dataset*, 2nd Colloquium of the Pensions, Benefits and Social Security Section of the International Actuarial Association, Helsinki.

Landwehr N., Hall M. and Frank E. (2005) Logistic Model Trees, *Machine Learning*, 59(1-2), pp. 161-205.

Larose, D.T. (2005) *Discovering Knowledge in Data: An Introduction to Data Mining*, Hoboken, NJ: Wiley.

Leow, M., Mues, C. and Thomas, L. (2009) *LGD Modelling for Mortgage Loans*, Credit Scoring and Credit Control XI, Edinburgh.

Leow, M., Mues, C. and Thomas, L. (2010) *Competing Risks Survival Model for Residential Mortgage Loans*, European Conference on Operational Research EURO XXIV, Lisbon.

Li, D.X. (2000) On Default Correlation: A copula Function Approach, *Journal of Fixed Income*, 9(4), pp. 43-54.

Lillard, L.A. and Willis, R.J. (1978) Dynamic aspects of earning mobility, *Econometrica*, 46(5), pp. 985-1012.

Loh, W.-Y. (2006) Logistic Regression Tree Analysis, In: Pham, H. (ed.) *Handbook of Engineering Statistics*, London: Springer, pp. 537-549.

Loterman, G., Brown, I., Martens, D., Mues, C. and Baesens, B. (2009) *Benchmarking State-Of-The-Art Regression Algorithms for Loss Given Default Modelling*, Credit Scoring and Credit Control XI, Edinburgh.

Lucas, R. (2005) *Improving Credit Offers Using Affordability Predictions*, Credit Scoring and Credit Control IX, Edinburgh.

Lucas, R.E. Jr. (1976) Econometric Policy Evaluation: A Critique, In: Brunner, K. and Meltzer, A.H. (eds.) *The Phillips Curve and Labor Markets*, Carnegie-Rochester Conference Series on Public Policy, Amsterdam: North-Holland, pp. 19-46.

Lunn, D., Spiegelhalter, D., Thomas, A. and Best, N. (2009) The BUGS project: Evolution, critique and future directions, *Statistics in Medicine*, 28(25), pp. 3049-3067.

Lusardi, A. (1992) *Permanent income, current income and consumption: Evidence from panel data*, Tilburg University, CentER for Economic Research Discussion Paper No. 9253.

Lynch, S.M. (2007) *Introduction to Applied Bayesian Statistics and Estimation for Social Scientists*, New York: Springer.

Maddala, G.S. (2001) *Introduction to Econometrics*, 3rd ed., Chichester: Wiley.

Madigan, D. and York, J. (1995) Bayesian Graphical Models for Discrete Data, *International Statistical Review*, 63(2), pp. 215-232.

References

Makuch, W.M. (2001) The Basics of a Better Application Score, In: Mays, E. (ed.) *Handbook of Credit Scoring*, Chicago: Glenlake Publishing Company, pp. 127-148.

Malik, M. and Thomas, L.C. (2010) Modelling credit risk of portfolio of consumer loans, *Journal of the Operational Research Society*, 61(3), pp. 411-420.

Matuszyk, A., Mues, C. and Thomas, L.C. (2010) Modelling LGD for unsecured personal loans: decision tree approach, *Journal of the Operational Research Society*, 61(3), pp. 393-398.

Maydon, T. (2011) *The evolution of customer affordability determination in the South African market*, Credit Scoring and Credit Control XII, Edinburgh.

Mays, E. (2004) *Credit Scoring for Risk Managers: The Handbook for Lenders*, Mason, OH: Thomson South-Western.

Miguéis, V.L., Benoit, D.F. and van den Poel, D. (2012) Enhanced decision support in credit scoring using Bayesian binary quantile regression, *Journal of the Operational Research Society*, doi:10.1057/jors.2012.116.

Miles, D. (1997) A Household Level Study of the Determinants of Incomes and Consumption, *The Economic Journal*, 107(440), pp. 1-25.

Mira, A. and Tenconi, P. (2004) Bayesian estimate of default probabilities via MCMC with delayed rejection, In: Dalang, R.C., Dozzi, M. and Russo, F. (eds.) *Seminar on Stochastic Analysis, Random Fields and Applications IV, Centro Stefano Franscini, Ascona, May 2002*, Basel: Birkhäuser Verlag, pp. 275-290.

Modigliani, F. and Brumberg, R. (1954) Utility Analysis and the Consumption Function: An Interpretation of Cross-Section Data, In: Kurihara, K.K. (ed.) *Post Keynesian Economics*, New Brunswick, NJ: Rutgers University Press, pp. 388-436.

Moral, G. (2006) EAD Estimates for Facilities with Explicit Limits, In: Engelmann, B. and Rauhmeier, R. (eds.) *The Basel II Risk Parameters: Estimation, Validation, and Stress Testing*, Berlin: Springer, pp. 197-242.

National Consumer Credit Protection Act 2009, Canberra: Office of Legislative Drafting and Publishing, Attorney-General's Department.

National Credit Act 2005, *Government Gazette*, 15 March 2006.

Ntzoufras, I. (2009) *Bayesian Modeling Using WinBUGS*, Hoboken, NJ: Wiley.

O'Hagan, A., Buck, C.E., Daneshkhah, A., Eiser, J.R., Garthwaite, P.H., Jenkinson, D.J., Oakley, J.E. and Rakow, T. (2006) *Uncertain Judgements: Eliciting Experts' Probabilities*, New York: Wiley.

Office of Fair Trading (2011) *Irresponsible lending – OFT guidance for creditors*, London: Office of Fair Trading.

Oxera (2004) *Are UK households over-indebted?*, Report prepared for: the Association for Payment Clearing Services, British Bankers' Association, Consumer Credit Association and Finance and Leasing Association.

Park, Y., Sirakaya, S. and Kim, T.Y. (2010) *A Dynamic Hierarchical Bayesian Model for the Probability of Default*, University of Washington, Center for Statistics and the Social Sciences Working Paper no. 98.

Pluto, K. and Tasche, D. (2006) Estimating Probabilities of Default for Low Default Portfolios, In: Engelmann, B. and Rauhmeier, R. (eds.) *The Basel II Risk Parameters: Estimation, Validation, and Stress Testing*, Berlin: Springer, pp. 79-103.

Qi, M. (2009) *Exposure at Default of Unsecured Credit Cards*, Office of the Comptroller of the Currency Economics Working Paper 2009-2.

Qi, M. and Yang, X. (2009) Loss given default of high loan-to-value residential mortgages, *Journal of Banking and Finance*, 33(5), pp. 788-799.

Qian, G., Rao, C.R., Wu, Y. and Shao, Q. (2008) Estimating the Number of Clusters in Logistic Regression Clustering by an Information Theoretic Criterion, In: Shalabh and Heumann, C. (eds.) *Recent Advances in Linear Models and Related Areas: Essays in Honour of Helge Toutenburg*, Heidelberg: Physica-Verlag, pp. 29-43.

Querci, F. (2005) *Loss Given Default on a medium-sized Italian bank's loans: an empirical exercise*, European Financial Management Association Annual Meetings, Milan.

References

Ralph, C. (2006) *Using Adaptive Random Trees (ART) for optimal scorecard segmentation*, Fair Isaac White Paper.

Rosch, D. and Scheule, H. (2003) Forecasting retail portfolio credit risk, *The Journal of Risk Finance*, 5(2), pp. 16-32.

Rubin, D.B. (1981) The Bayesian bootstrap, *The Annals of Statistics*, 9(1), pp. 130-134.

Runkle, D.E. (1991) Liquidity constraints and the permanent-income hypothesis: Evidence from panel data, *Journal of Monetary Economics*, 27(1), pp. 73-98.

Russell, P. (2005) *Over-indebtedness and responsible lending in the UK*, Credit Scoring and Credit Control IX, Edinburgh.

Siddiqi, N. (2005) *Credit Risk Scorecards: Developing and Implementing Intelligent Credit Scoring*, New York: Wiley.

Smith, W.R. (1956) Product Differentiation and Market Segmentation as Alternative Marketing Strategies, *Journal of Marketing*, 21(1), pp. 3-8.

Snowdon, B. and Vane, H.R. (2005) *Modern Macroeconomics: Its Origins, Development and Current State*, Cheltenham: Edward Elgar.

Somers, M. (2009) *Credit Card Initial Limits: How much is too much?*, Credit Scoring and Credit Control XI, Edinburgh.

Somers, M. and Whittaker, J. (2007) Quantile regression for modelling distributions of profit and loss, *European Journal of Operational Research*, 183(3), pp. 1477-1487.

Taplin, R., Minh To, H. and Hee, J. (2007) Modeling exposure at default, credit conversion factors and the Basel II Accord, *The Journal of Credit Risk*, 3(2), pp. 75-84.

The UK Cards Association (2011) *Credit card limit increases*, London: The UK Cards Association.

Thomas, L.C. (2009a) *Consumer Credit Models: Pricing, Profit, and Portfolios*, New York: Oxford University Press.

Thomas, L.C. (2009b) Modelling the credit risk for portfolios of consumer loans: Analogies with corporate loan models, *Mathematics and Computers in Simulation*, 79(8), pp. 2525-2534.

Thomas, L.C. (2010) Consumer finance: challenges for operational research, *Journal of the Operational Research Society*, 61(1), pp. 41-52.

Thomas, L.C. (2011) *Credit Scoring and the Edinburgh Conference: Fringe or International Festival*, Credit Scoring and Credit Control XII, Edinburgh.

Thomas, L.C., Edelman, D.B. and Crook, J.N. (2002) *Credit Scoring and Its Applications*, Philadelphia, PA: SIAM.

Thomas, L.C., Ho, J. and Scherer, W.T. (2001) Time will tell: Behavioural scoring and the dynamics of consumer credit assessment, *IMA Journal of Management Mathematics*, 12(1), pp. 89-103.

Thomas, L.C., Matuszyk, A. and Moore, A. (2012) Comparing debt characteristics and LGD models for different collections policies, *International Journal of Forecasting*, 28(1), pp. 196-203.

Thomas, L.C., Thomas S., Tang L. and Gwilym O.A. (2005) Impact of demographic and economic variables on financial policy purchase timing decisions, *Journal of the Operational Research Society*, 56(9), pp. 1051-1062.

Tong, E., Mues, C. and Thomas, L. (2011) *A zero-adjusted gamma model for estimating loss given default on residential mortgage loans*, Credit Scoring and Credit Control XII, Edinburgh.

Valvonis, V. (2008) Estimating EAD for retail exposures for Basel II purposes, *The Journal of Credit Risk*, 4(1), pp. 79-109.

Van Gestel, T. and Baesens, B. (2009) *Credit Risk Management. Basic concepts: financial risk components, rating analysis, models, economic and regulatory capital*, New York: Oxford University Press.

References

VantageScore (2006) *Segmentation for Credit Based Delinquency Models*, VantageScore White Paper.

Verbeek, M. (2004) *A Guide to Modern Econometrics*, 2nd ed., Chichester: Wiley.

Wagenmakers, E.-J., Lee, M., Lodewyckx, T. and Iverson, G.J. (2008) Bayesian Versus Frequentist Inference, In: Hoijtink, H., Klugkist, I. and Boelen, P. (eds.) *Bayesian Evaluation of Informative Hypotheses*, New York: Springer, pp. 181-207.

Wasserman, L. (2000) Bayesian Model Selection and Model Averaging, *Journal of Mathematical Psychology*, 44(1), pp. 92-107.

Wedel, M. and Kamakura, W.A. (2000) *Market Segmentation: Conceptual and Methodological Foundations*, New York: Springer.

Whittaker, J., Whitehead, C. and Somers, M. (2005) The Neglog Transformation and Quantile Regression for the Analysis of a Large Credit Scoring Database, *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 54(5), pp. 863-878.

Whittaker, J., Whitehead, C. and Somers, M. (2007) A dynamic scorecard for monitoring baseline performance with application to tracking a mortgage portfolio, *Journal of the Operational Research Society*, 58(7), pp. 911-921.

Wilkinson, G. (2007) *Responsible Lending*, Credit Scoring and Credit Control X, Edinburgh.

Wooldridge, J.M. (2010) *Econometric Analysis of Cross Section and Panel Data*, 2nd ed., Cambridge, MA: The MIT Press.

Yobas, M.B., Crook, J.N. and Ross, P. (2004) Credit scoring using neural and evolutionary techniques, In: Thomas, L.C., Edelman, D.B. and Crook, J.N. (eds.) *Readings in Credit Scoring: Foundations, Developments, and Aims*, New York: Oxford University Press, pp. 277-293.

Zeldes, S.P. (1989) Consumption and Liquidity Constraints: An Empirical Investigation, *Journal of Political Economy*, 97(2), pp. 305-346.

Zhang, J. and Thomas, L.C. (2012) Comparisons of linear regression and survival analysis using single and mixture distributions approaches in modelling LGD, *International Journal of Forecasting*, 28(1), pp. 204-215.

Zhang, Y., Ji, L. and Liu, F. (2010) *Local Housing Market Cycle and Loss Given Default: Evidence from Sub-Prime Residential Mortgages*, International Monetary Fund Working Paper WP/10/167.

Ziemba, A. (2005) *Bayesian updating of generic scoring models*, Credit Scoring and Credit Control IX, Edinburgh.